International Journal on Perceptive and Cognitive Computing

Volume 12, Issue 1, Year 2026

IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA
ISSN: 2462 – 229X
http://journals.iium.edu.my/aict/index.php/iJPCC

INTERNATIONAL JOURNAL ON PERCEPTIVE AND COGNITIVE COMPUTING (IJPCC)

Vol. 12 No. 1 (2026): January 2026
DOI: https://doi.org/10.31436/ijpcc.v12i1

COPYRIGHT TRANSFER AGREEMENT

1. Consent to publish: The Author(s) agree to publish the article named above with IIUM Press.
2. Declaration: The Author(s) declare that the article named above has not been published before in any Forman that it is not concurrently submitted to another publication, and also that it does not infringe on anyone's copyright. The Author(s) holds the IIUM Press and Editors of the journal harmless against all copyright claims.
3. Transfer of copyright: The Author(s) hereby agree to transfer the copyright of the article to IIUM Press, which shall have the exclusive and unlimited right to publish the article in any form, including in electronic media. However, the Author(s) will reserve the right to reproduce the article for educational and scientific purposes provided that written consent of the Publisher is obtained.

The International Journal on Perceptive and Cognitive Computing (IJPCC) journal follows the open access policy.

All articles published open access will be immediately and permanently free for everyone to read, download, copy and distribute for non-commercial purposes.

Committee Member	Sherzod Turaev	United Arab Emirates University, UAE
Committee Member	Zainab Senan Mahmod Attar Bashi	International Islamic University Malaysia
Committee Member	Messikh Azzeddine	Semiconductors Technology Research Center for Energetics, Algiers, Algeria
Committee Member	Untung Rahardja	Universitas Raharja, Indonesia

International Committee Board

Position	Name	Affiliation
Board Member	Ruhul A. Sarker	UNSW Canberra, Australia
Board Member	Iftikhar Sikder	Cleveland State University, USA
Board Member	Chehri Abdellah	University of Ottawa, Canada
Board Member	Muhammad Mostafa Monowar	King Abdul Aziz University, KSA
Board Member	Riadh Robbana	INSAT – Carthage University, Tunisia
Board Member	Mohammed Atiquzzaman	University of Oklahoma, USA
Board Member	AbdulRahman Alsamman	University of New Orleans, USA
Board Member	Mahfuz Aziz	University of South Australia, Australia
Board Member	Mostafa M. Fouda	Benha University, Egypt
Board Member	Md Mahbubur Rahim	Monash University, Australia
Board Member	Zubair Md. Fadlullah	Tohoku University, Japan
Board Member	Qurban A. Memon	UAE University, UAE
Board Member	Riaz Ahmed Shaikh	King Abdul Aziz University, KSA
Board Member	Mohammad Abdul Salam	Southern University and A&M College, USA
Board Member	Mohamed Essaaidi	Mohammed V University, Morocco
Board Member	Alaa Hussein Al-Hamami	Aman Arab University, Jordan
Board Member	Hilal M. Yousif Al-Bayatti	Applied Science University, Bahrain
Board Member	Siddeeq Y. Ameen	University of Mosul, Iraq
Board Member	Ismail Khalil	Institute of Telecooperation, Johannes Kepler University Linz, Austria

TABLE OF CONTENT

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.622

Integrating Edge Computing with Internet of Things Systems for Smart Homes

Marwan S. M. Al-Dabbagh
Department of Computer Science, College of Education for Pure Science, Mosul University, Mosul, IRAQ

*Corresponding author Marwan.aldabbagh@uomosul.edu.iq
(Received: 20th July 2025; Accepted: 22nd September, 2025; Published on-line: 30th January, 2026)

Abstract— The fast development of the Internet of Things (IoT) has dramatically transformed dwelling space, leading to the development of smart homes smart systems of connected devices that can perceive, compute and communicate. They enable automation, increase energy efficiency, safety and security, and improve user comfort etc. In the recent years, smart home systems were built upon centralized cloud computing architecture for data processing and decision-making. Even though the cloud brings in the ability of scalability and the centralized management of resources; but it also imposes some important challenges, including high end-to-end delay, high consumption, and an increase in vulnerability to data exposure since confidential information is moved to the other servers. Edge computing has been proposed as a promising paradigm to mitigate these issues by distributing computation and allowing data to be processed at, or near, the data source. This paper study the deployment of edge computing support for IoT systems in smart homes and perform a comparative study of two system architectures, the traditional cloud-based and the proposed edge-enabled architecture. The performance is evaluated through extensive simulations in OMNET++ using real smart home scenarios and traffic patterns. System efficiency is evaluated considering key performance parameters— end-to-end latency, bandwidth usage, CPU utilization and packet delivery ratio. It is shown that the edge-enabled architecture significantly reduces the latency, efficiently utilize the network resources, balance the processing load, and enhance the reliability of the packet delivery compared to cloud-only model.

Keywords— Internet of Things, Cloud computing, Edge computing, Smart homes, OMNET++ Simulator

I. Introduction

The exponential growth of digital and communication technologies during the last years has fuelled the development of the Internet of Things (IoT), a novel paradigm in which the real world is interconnected with the digital world [1]. IoT helps common objects such as home appliances, industrial machines, vehicles, and wearable devices to connect to each other and the internet, and interact on their own [2]. Such devices are equipped with the sensors, the actuators, microcontroller, and the communication interfaces that can enable them to perceive the surrounding, variety, and complexity, and to process their interactivity locally or remotely, and to communicate with other devices and/or cloud services. The IoT ecosystem enables real-time monitoring, automation, and intelligent decision making in several domains such as health, transportation, agriculture, smart cities etc. [3], [4].

An example of one of such application areas is smart home, a home environment facilitated by connected IoT devices [5]. These products are designed to make life and living in your home easy, comfortable, secure, efficient or just fun through intelligent automation and remote control. Examples are thermostats that automatically regulate for temperature and level of occupancy, light systems that are adapted based on user preferences, surveillance cameras with motion sensing, door locks that are remotely unlocked, and appliances that are controllable or monitorable trough smartphones [6]. In a smart home, different subdomain systems (energy management [7], environmental monitoring [8], and security cooperate with each other to maximize resource utilization and use experience [9].

The conventional smart home systems heavily depend on cloud computing [10]. In cloud-centric architectures as depicted in Fig. 1, data accumulated by IoT devices is transferred over internet to centralized cloud servers for storage, processing, and analysis. The cloud infrastructure offers powerful computing capabilities, elastic storage, and centralized data management, making it suitable for large-scale data analytics and complex decision-making [11]. However, cloud computing also introduces significant limitations, especially in the context of real-time and privacy-sensitive smart home applications. First, end-to-end latency increases due to the round-trip time required for data transmission between devices and remote cloud servers. This delay can be detrimental to latency-critical applications such as fire alarms, intrusion detection, or emergency response systems. Second, the bandwidth consumption required to transmit large volumes of sensor data can strain

https://doi.org/10.31436/ijpcc.v12i1.622

network resources, particularly in homes with multiple connected devices. Third, centralized data storage raises security and privacy concerns, as personal and sensitive information is exposed to potential breaches, unauthorized access, or misuse in remote data centers [12].

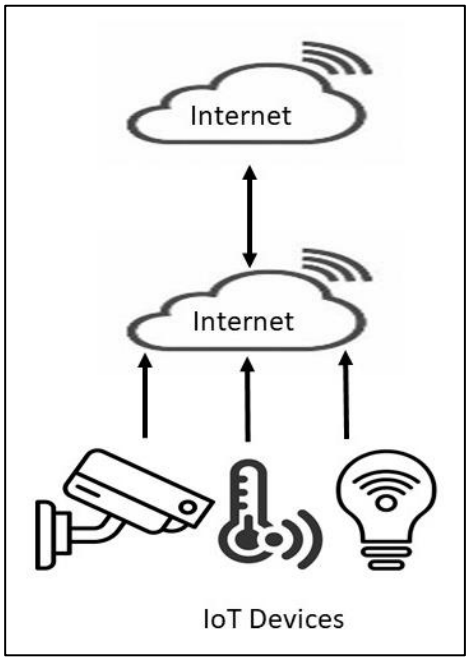


Fig. 1 Cloud-Based architecture

To address these challenges, the concept of edge computing has been introduced as a complementary or alternative computing paradigm. Edge computing, a strategy for computing on location where data is collected or used, allows IoT data to be gathered and processed at the edge, rather than sending the data back to a data center or cloud [13]. Fig. 2 illustrates the architecture of Edge computing in IoT, composed of three layers: IoT Devices Layer, which includes devices such as a security camera, a smart light bulb, and a motion sensor. These devices continuously collect environmental data and initiate communication within the smart home ecosystem. Edge Node Layer, which is represented by a wireless access point or gateway, the Edge Node is a localized processing unit—often a router, embedded system, or mini-server. It acts as a bridge between the IoT devices and the cloud for some data pre-processing, filtering, and sometimes real-time analytics. The edge node assists in moving computational burden from cloud, thereby speeding up the system and respond to high-priority events. Cloud Layer is still left as part of the architecture, where we perform deeper analytics, store huge amounts of data and train machine learning models. Data transmitted from edge to cloud is typically filtered or aggregated to avoid transmitting of redundant traffic and to offload the cloud onto only what is valuable [14]. In the smart home scenario, this pertains to taking advantage of intelligent edge nodes such as IoT gateways, micro servers, or embedded systems that can locally process, analyze and decide about the data. By removing the need for cloud, edge computing is the next level of innovation than can improve latency, reduce bandwidth requirement by pre-processing data before sending it and strengthen your data security by ensuring that your data is not "in-flight" or indeed stored in a remote easily accessible place. In addition, edge-enabled applications are more tolerant to network outage, since they are able to work even over a short time of occasional connections to the cloud [15].

This study aims to explore and evaluate the integration of edge computing within IoT-based smart home environments. Specifically, it presents a comparative analysis of two architectural models: a conventional cloud-based smart home system and a smart home system integrated with edge computing. These models are implemented and simulated using OMNET++, an event-driven network simulation platform that provides a flexible and modular environment for modelling complex network behaviors, IoT protocols, and distributed systems.

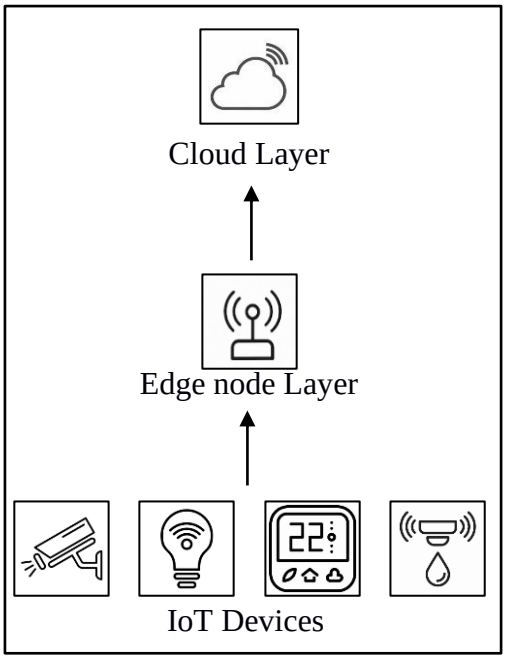


Fig. 2 Edge-Based architecture

II. Related Works

The proliferation of Internet of Things (IoT) in residential context has made possible the emergence of smart homes as intelligent environments, capable of providing comfort, safety, energy efficiency, and automation by means of interconnected devices. The devices collect and transmit data on an on-going basis; these data need to be processed in order to put them to good use. Historically, smart home

https://doi.org/10.31436/ijpcc.v12i1.622

systems have been based on cloud-centric approaches though recent work has shown interest in edge-based or hybrid computing to address the drawbacks of centralized processing. This section is a detailed survey for both cloud enabled and edge empowered smart home models.

Hsiao T. et al designed a cloud-based platform for improving residential energy efficiency and user comfort in the smart homes' environment. The system combines IoTs with computing and cloud for efficient power management. It offers real-time power consumption information, appliances control and statistical analysis of the power usage data collected [16].

Albataineh and Bollampall [17] introduced a hybrid cloud-edge computing solution to increase the productivity of smart homes with the processing of IoT data. The research utilized a profound machine learning engine, in which a decision tree was utilized, to determine how the communication between the edge layer, failover between edges and the Cloud layer can be made. The result shows that the developed model has more throughput and more power consumption [17].

Jun L. presented a study that delves into the integration of cloud computing technology into smart home systems of the Internet of Things (IoT). The study had analysed cloud computing and Internet of Things integration, improved service quality in the smart home industry, and explored cloud computing applications in smart home systems [18].

Ma S. proposed a cloud computing-edge cooperation framework in the context of IoT applications. The proposed model enhanced the performance of real-time and data processing. It effectively reduced data transmission delay and network latency. In addition, I enhance resource utilization, task processing success rate, and bandwidth usage is significantly reduced in various IoT scenarios [19].

Saxena M. et al. introduced a resource allocation strategy for IoT Networks which maximizes the use of edge devices. The proposed strategy minimizes the end-to-end latency through an edge-to-edge device offload strategy, and the Queue delay is reduced by optimizing CPU frequency [20].

Sithiyopasakul J. et al. reported a detailed performance analysis of cloud computing and IoT, specifically on three large systems: Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure. This study performed a comprehensive evaluation of the performance of cloud computing and IoT systems in terms of essential parameters, good performance with prospects to further improvement regarding response time, latency, and reliability were indicated in some cases. [21].

Kalra S., Mathur G, and Parashar A. presented a research on that about the wireless home automation system improvement using edge computing. The authors aimed to determine how introducing edge computing to home automation systems can overcome the main shortcomings of "classic" cloud-based architecture (high latency, security issues, ineffective real-time processing). It was based around creating a local edge-computing setup that consisted of hardware such as Raspberry Pi & OpenFaaS with a selection of sensors attached, and performing compare/contrasts between the only-cloud versus edge-configurations. Result depicted that edge computing is a more flexible, secure and power effective solution for the contemporary smart homes. But the experiments relied on a simple edge design based on Raspberry Pi and a limited number of sensors. Although this configuration is proof of concept, it is not scalable and does not resemble the real complexities of a bigger or mixed smart home ecosystem (dozens, hundreds or more than thousands of devices) [10].

Papcun P. et al. introduced an edge-enabled IoT gateway to lower data forwarding, cost, and underlined its capability. The role-edge within healthcare applications was underlined in this paper, which had defined four classes of IoT gateways: (1) normal gateway, (2) smart gateway, (3) intelligent gateway, and (4) edge gateway [22].

Pal T. et al. conducted a cooperative processing model based on cloud-edge computing and adopted the unified system deployment scheme in Kubernetes to implement cooperative processing. Experimental results show that the proposed approach provides better operational efficiency than systems firmly based on either the cloud or single-edge computing, so as to more adequately satisfy the real-time demands of smart homes [23].

III. METHODOLOGY

This section describes the detailed methodology used to simulate and evaluate the performance of cloud-based and edge-enabled smart home architectures.

A. The Proposed Framework

Fig. 3 illustrates the proposed edge computing framework for a smart home environment, organized into three primary layers: Device Layer, Edge Layer, and Cloud Layer. The Device Layer contains the IoT devices such as motion sensors, smart locks, and IP cameras which constantly generate sensor data depending on the context, which detect an environmental change and user interaction. This information is then sent to the Edge Layer who process the data: (1) in the Data Pre-processing Unit where it is filtered and the values are minimized; (2) in the Event Detection Engine, used to detect relevant information, such as an intrusion, a fire, etc.; and (3) Local Decision Module, which takes a quick action, like open a door. In order to study the performance of the considered architectures, we evaluate them against a selection of performance measures,

https://doi.org/10.31436/ijpcc.v12i1.622

important to design and operate smart home systems. The Edge Node ensures low-latency responses while also communicating with the Cloud Layer. The Cloud Analytics Engine performs advanced tasks such as learning-based analytics, periodic summaries, and system-wide optimization. Arrows indicate the bidirectional flow of data: sensor inputs and control feedback travel vertically between layers, while pre-processed data and user commands are exchanged horizontally between the edge and cloud. This architecture supports a scalable, efficient, and responsive smart home ecosystem with minimal cloud dependency for critical operations.

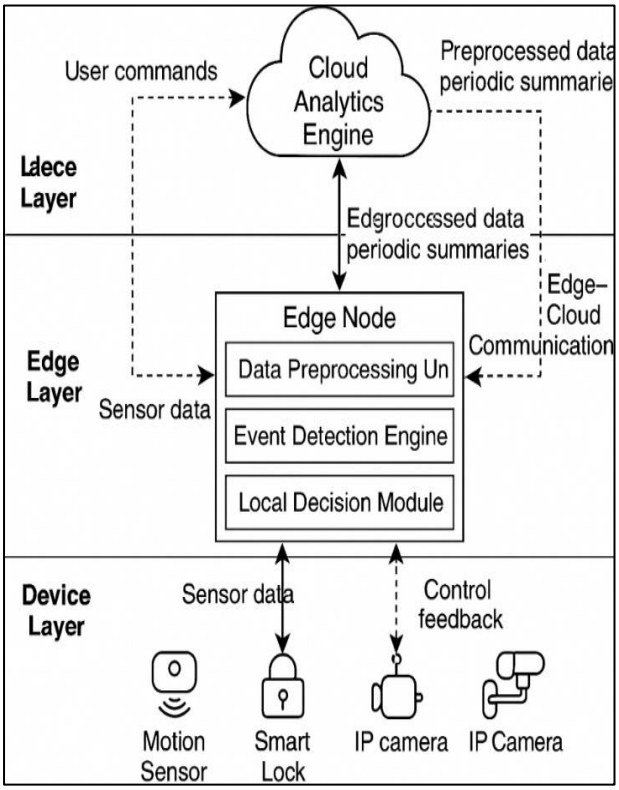


Fig. 3 The proposed Edge framework description

B. Simulation Environment

The selection of the research methodology and evaluation criteria was driven by the study's objective to assess the effectiveness of integrating edge computing into IoT-based smart home systems. A simulation-based approach using OMNET++ with the INET framework was adopted to create a controlled and realistic environment that models network topologies, IoT devices, and communication protocols without the complexity of large-scale physical deployment. The topology simulates a home setting with 20 IoT devices, connected either directly to the cloud or through a local edge node. Table I. presents the simulation parameters for the conducted scenarios.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Simulation Tool	OMNET++ (INET Framework)
Network Bandwidth	100 Mbps
Latency (Cloud Path)	50 ms
Latency (Edge Path)	5 ms
Packet Size	512 bytes
Number of IoT Devices	20
Simulation Duration	600 seconds
Traffic Model	Constant Bit Rate (CBR)
Edge Node Processing Time	2 ms
Cloud Server Processing Time	20 ms

C. Simulation Scenarios

Two scenarios are simulated using OMNET++ to evaluate the performance of smart home system.

1) Scenario 1 Traditional Cloud-Based Smart Home System: In this scenario, all IoT devices in the smart home, such as motion sensors, door locks, surveillance cameras, smart lights, and thermostats, are connected directly to the cloud via the internet. These devices transmit their data to remote cloud servers, where data aggregation, analytics, and decision-making are carried out. The cloud then sends control commands back to the devices based on the analysis. This approach centralizes processing but incurs communication latency, especially for time-sensitive actions like unlocking a door in response to an authorized facial recognition event or triggering an alarm in case of intrusion detection. The high volume of transmitted data also places significant demands on network bandwidth.

2) Scenario 2 Smart Home System with Integrated Edge Computing: In this scenario, IoT devices initially send data to a nearby edge node (e.g., home gateway or smart hub) with computing capabilities. The edge node performs in situ data processing and decision-making for real-time tasks at the local side. This distributed method can alleviate the effect of network congestion, improve real-time response, and minimize the dependence of the smart home system on an internet blockage.

D. Evaluation Metrics

To assess the effectiveness of the two architectures, we evaluate them using a set of key performance metrics that are critical to the design and operation of smart home systems:

1) End-to-End Latency: The time taken by a data packet to traverse from a sensing device to a control unit (in the cloud or at the edge) and for the response to be

https://doi.org/10.31436/ijpcc.v12i1.622

executed. It is essential for real-time-oriented applications [24].

2) Bandwidth Usage: Bandwidth is the amount of information that can be delivered over a connection. Optimizing bandwidth usage is essential for scalability and for the sustenance of network performance in the presence of growing number of devices [25].

3) CPU Utilization: This measures the computational load on processing units, whether in edge nodes or cloud servers. It reflects the system's ability to handle processing tasks effectively without overloading hardware resources [26].

4) Packet Delivery Ratio (PDR): This represents the computationally loaded point in processing units on which in edge node or cloud server. It indicates the systems capacity to process the tasks efficiently without overloading the hardware resources [27], [28].

By simulating and analyzing these metrics, this paper provides empirical evidence on the advantages and trade-offs of integrating edge computing into smart home systems. The results contribute to the ongoing efforts in designing efficient, scalable, secure, and responsive residential IoT infrastructures that align with the growing demands of real-time intelligent automation.

IV. Results and Discussion

A. End-to-End Latency

As observed in Fig. 4 the latency of edge-based is reduced far less than the Latency of the cloud-based. The cloud model averaged a latency of 225 msec, compared to 125 msec with the edge model, where this 44% acceleration is primarily associated with enabling the processing locally on the network edge, thus eliminating the necessity for the continuous round-trip communication with the remote cloud servers. This substantial improvement enables instant responses to time-sensitive events such as intrusion detection, fire alarms, and health monitoring, ensuring better automation performance and user experience. The lower latency also reduces the risk of failure in emergency scenarios and enhances the overall reliability and efficiency of smart home systems.

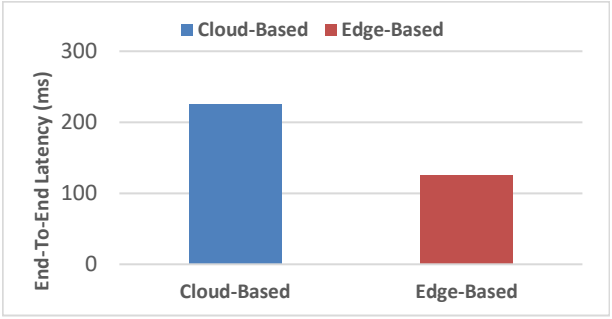


Fig.4 Average End-to-End Latency (ms) observed in cloud-based and edge-based smart home architectures

B. Bandwidth Usage

The total bandwidth usage in the two scenarios is shown in Fig. 5. The cloud-based model required more bandwidth to be used (about 90 MB) as opposed to the edge-based model (about 65 MB). This reduction of the bandwidth consumption around 28% is due to localized data aggregation and filtering at the edge devices which allows to avoid transmitting redundant and not essential data to the cloud. This increases network efficiency and lowers operating costs in bandwidth-constrained or per-byte/bandwidth-sensitive environments.

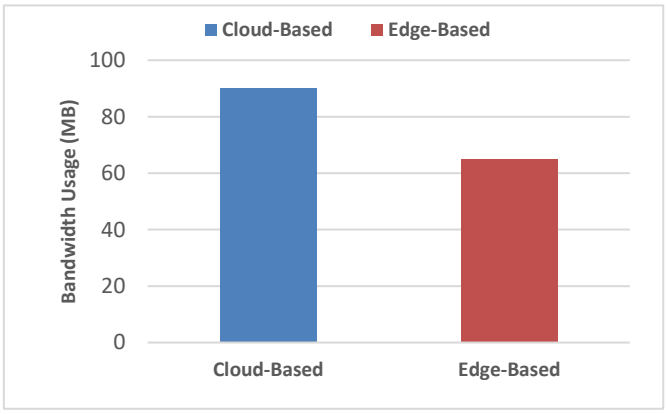


Fig. 5 Total Bandwidth Usage (MB) comparison between cloud and edge models

C. CPU Utilization

As illustrated in Fig. 6, the cloud-based system exhibited a higher average CPU load of approximately 42% than the edge-based model, which incurred a lower load of approximately 32%. This distinction indicates that edge computing indeed distributes computational tasks nondiscretely around the local resources and does not encounter bottlenecks as do centralized systems. Lower CPU usage also means better scalability and energy efficiency in the system by distributing processing between the edge and the cloud.

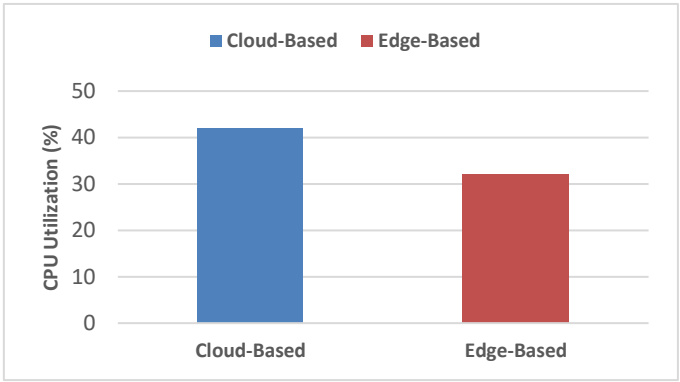


Fig. 6 Average CPU Utilization (%) for cloud vs. edge scenarios

D. *Packet Delivery Ratio (PDR)*

Packet Delivery Ratio in Fig. 7 results present enhanced communication reliability in edge-based model. The cloud-based one obtained PDR around 91%, and the edge-based one achieved PDR about 96%. This higher PDR observed is an evidence of better network performance, and of less congestion caused by shorter paths, which is a side effect of making decisions locally. A high end-to-end PDR is particularly important for smart home systems where the accurate and reliable delivery of control commands and sensor data is necessary for the correct operation of the system.

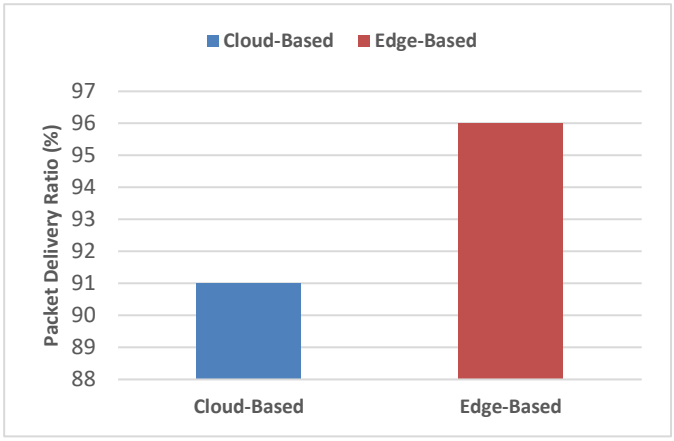


Fig. 7 Packet Delivery Ratio (%) comparison showing improved reliability with edge computing

V. Conclusion and Future Work

This study has shown that the combination of edge computing and IoT-based smart home systems can be benefited by performance advantages over the traditional cloud-based approaches. The study rigorously simulated in OMNET++ and found significant enhancements in the end-to-end latency, bandwidth efficiency, the CPU loading balancing and the packet delivery ratio. These improvements are particularly critical for applications requiring real-time responsiveness, data security, and uninterrupted connectivity. The study confirms that localizing computational tasks near the data source not only enhances system responsiveness but also reduces dependence on centralized cloud resources, ultimately leading to a more efficient and resilient smart home infrastructure. For future work, the author will focus on extending this research through the implementation of real-world prototypes that integrate edge devices with smart home platforms. Additionally, incorporating artificial intelligence and machine learning capabilities at the edge can further enable intelligent, context-aware decision-making. There is also a need to explore dynamic load balancing strategies for seamless edge-cloud coordination and to investigate the energy consumption implications of edge devices to ensure sustainable deployment. Finally, the scalability of edge-enabled smart homes across communities and urban environments will be studied to validate the approach in broader IoT ecosystems.

Acknowledgment

The author expresses gratitude to the University of Mosul for their cooperation in conducting this work.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author(s) Contribution Statement

The author contributed to the study conception and design, manuscript writing, and approval of the final version.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study did not require ethical approval

References

[1] M. Mohamed, "A comparative study on Internet of Things (IoT): Frameworks, Tools, Applications and Future directions," *J. Intell. Syst. Internet Things*, vol. 1, pp. 13–39, 2020, doi: 10.54216/jisiot.010102.

[2] M. Elkhodr, S. Shahrestani, and H. Cheung, *Internet of things applications: Current and future development*, no. July. 2016. doi: 10.4018/978-1-5225-0287-6.ch016.

[3] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Futur. Gener. Comput. Syst.*, vol. 29, no. 7, 2013, doi: 10.1016/j.future.2013.01.010.

[4] M. Al-Dabbagh, A. Al-Sherbaz, and S. Turner, "Developing a real-time ITS using VANETs: A case study for Northampton Town," *Lect. Notes Networks Syst.*, vol. 15, pp. 640–651, 2018, doi: 10.1007/978-3-319-56994-9_43.

[5] S. Jamal Rashid, A. Alkababji, and A. M. Khidhir, "Communication and Network Technologies of IoT in Smart Building: A Survey," *NTU J. Eng. Technol.*, vol. 1, no. 1, pp. 1–18, 2021, doi: 10.56286/ntujet.v1i1.50.

[6] T. D. P. Mendes, R. Godina, E. M. G. Rodrigues, J. C. O. Matias, and J. P. S. Catalão, *Smart home communication technologies and applications: Wireless protocol assessment for home area network resources*, vol. 8, no. 7. 2015. doi: 10.3390/en8077279.

[7] C. Mahapatra, A. K. Moharana, and V. C. M. Leung, "Energy management in smart cities based on internet of things: Peak demand reduction and energy savings," *Sensors (Switzerland)*, vol. 17, no. 12, pp. 1–21, 2017, doi: 10.3390/s17122812.

[8] S. Chang and K. Nam, "Exploring the Sustainable Values of Smart Homes to Strengthen Adoption," *Buildings*, vol. 12, no. 11, 2022, doi: 10.3390/buildings12111919.

[9] M. Abunaser and A. a. a. Alkhatib, "Advanced survey of blockchain for the internet of things smart home," *2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc.*, no. April, pp. 58–62, 2019, doi: 10.1109/JEEIT.2019.8717441.

[10] G. Mathur and A. Parashar, "A STUDY ON ENHANCING HOME AUTOMATION SYSTEMS THROUGH EDGE COMPUTING," no. October, 2024.
[11] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Futur. Gener. Comput. Syst.*, vol. 97, no. April, pp. 219–235, 2019, doi: 10.1016/j.future.2019.02.050.
[12] S. Shukla, M. F. Hassan, D. C. Tran, R. Akbar, I. V. Paputungan, and M. K. Khan, "Improving latency in Internet-of-Things and cloud computing for real-time data transmission: a systematic literature review (SLR)," *Cluster Comput.*, vol. 26, no. 5, pp. 2657–2680, 2023, doi: 10.1007/s10586-021-03279-3.
[13] W. Yu *et al.*, "A Survey on the Edge Computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2017, doi: 10.1109/ACCESS.2017.2778504.
[14] X. Kong, Y. Wu, H. Wang, and F. Xia, "Edge Computing for Internet of Everything: A Survey," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23472–23485, 2022, doi: 10.1109/JIOT.2022.3200431.
[15] A. Al-Dulaimy, Y. Shatma, M. G. Khan, and J. Taheri, "Introduction to edge computing," in *Edge Computing*, no. March 2021, 2020, pp. 1–453. doi: 10.1049/PBPC033E.
[16] T. C. Hsiao, T. L. Chen, T. C. Kang, and T. Y. Wu, "The Implementation of Smart Home Power Management: Integration of Internet of Things and Cloud Computing," *Proc. 2019 IEEE Eurasia Conf. Biomed. Eng. Healthc. Sustain. ECBIOS 2019*, pp. 21–23, 2019, doi: 10.1109/ECBIOS.2019.8807872.
[17] H. Albataineh, M. Nijim, and D. Bollampall, "The Design of a Novel Smart Home Control System using Smart Grid Based on Edge and Cloud Computing," *2020 8th Int. Conf. Smart Energy Grid Eng. SEGE 2020*, pp. 88–91, 2020, doi: 10.1109/SEGE49949.2020.9181961.
[18] L. Jun, "Study on the Application of Cloud Computing Technology in the Intelligent Home System of Internet of Things," *Proc. - 2023 8th Int. Conf. Inf. Syst. Eng. ICISE 2023*, pp. 317–320, 2023, doi: 10.1109/ICISE60366.2023.00072.
[19] S. Ma, "Research on the Application of Cloud-Edge Collaborative Computing Model in IoT Scenarios," *2024 IEEE 2nd Int. Conf. Electr. Autom. Comput. Eng. ICEACE 2024*, pp. 251–255, 2024, doi: 10.1109/ICEACE63551.2024.10898777.
[20] M. Saxena, S. Srivastava, V. K. Dwivedi, R. Chitranshi, and P. K. Mishra, "Minimizing End-to-End Latency in Edge Computing-Enabled IoT Networks Through Edge-to-Edge Resource Allocation," *2nd IEEE Int. Conf. IoT, Commun. Autom. Technol. ICICAT 2024*, pp. 1512–1517, 2024, doi: 10.1109/ICICAT62666.2024.10923436.
[21] J. Sithiyopasakul, T. Archevapanich, S. Sithiyopasakul, A. Lasakul, B. Purahong, and C. Benjangkaprasert, "Implementation of Cloud Computing and Internet of Things (IoT) by Performance Evaluation," *Proceeding - 12th Int. Electr. Eng. Congr. Smart Fact. Intell. Technol. Tomorrow, iEECON 2024*, pp. 1–6, 2024, doi: 10.1109/iEECON60677.2024.10537945.
[22] P. Papcun, E. Kajati, D. Cupkova, J. Mocnej, M. Miskuf, and I. Zolotova, "Edge-enabled IoT gateway criteria selection and evaluation," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 13, pp. 1–9, 2020, doi: 10.1002/cpe.5219.
[23] T. Pal, R. Saha, S. Sen, S. Saif, and S. Biswas, *Architecture for Smart Healthcare: Cloud Versus Edge*. Springer Nature Singapore, 2022. doi: 10.1007/978-981-19-1408-9_2.
[24] A. Garcia-santiago, C. Josefina, and J. F. G, "Evaluation of AODV and DSDV Routing Protocols for a FANET: Further results towards robotic vehicle networks," in *2018 IEEE 9th Latin American Symposium on Circuits & Systems (LASCAS)*, 2018, pp. 3–6.
[25] N. Alzibdeh, M. T. Alrashdan, and A. Almabhouh, "Bandwidth Utilization with Network Traffic Analysis," *3rd IEEE Int. Conf. Mob. Networks Wirel. Commun. ICMNWC 2023*, pp. 1–5, 2023, doi: 10.1109/ICMNWC60182.2023.10435712.
[26] R. Behravesh, E. Coronado, D. Harutyunyan, and R. Riggio, "Joint User Association and VNF Placement for Latency Sensitive Applications in 5G Networks," *Proceeding 2019 IEEE 8th Int. Conf. Cloud Networking, CloudNet 2019*, no. November, 2019, doi: 10.1109/CloudNet47604.2019.9064145.
[27] D. Rai, S. S. Rajput, and D. Rai, "Analysis of Various Routing Protocols based on Quality of Service for FANET," *2023 6th Int. Conf. Inf. Syst. Comput. Networks, ISCON 2023*, pp. 1–5, 2023, doi: 10.1109/ISCON57294.2023.10111945.
[28] M. S. M. Al-dabbagh, "Exploiting Conventional MANET R outing in UAV ' s Based Environment," pp. 12–17, 2024.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.534

Balancing Innovation and Privacy in the Decentralized Metaverse: Case studies of Exploring Blockchain and Web 3.0 for Sustainable Development

[1]Farresa Haifa' Mohammed, [1]Izzah Athirah Izham, [1]Maisarah Jaafar Akbar,
[1]Nor Syazana Mohd Ansar, [1]Nur Maisarah Roslan ,[1,*]Ahmad Anwar Zainuddin, [2]Haikal Khusairi Ahmad
[1]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Gombak, Malaysia.
[2]Kulliyyah of Engineering, International Islamic University Malaysia, Gombak, Malaysia.

*Corresponding author: anwarzain@iium.edu.my
(Received: 5th January 2025; Accepted: 22nd July, 2025; Published on-line: 30th January, 2026)

Abstract- The current centralized metaverse platforms come inherently wrapped in significant challenges related to data security vulnerabilities, scalability issues, limited user autonomy, and the interoperability of virtual environments. These naturally impose core limitations on digital ownership and creative freedoms over user experiences and digital assets. This paper discusses decentralization as a powerful solution to return the capability of self-governance to the user over their identities and assets. Integration with blockchain ensures secure, transparent, and immutable transactions, while smart contracts facilitate trust and automation of governance. The case studies show the real-life application on how decentralized platforms can achieve scalability and reduce energy challenges but at the same time enhance user control and interoperability. Noting innovation, the discussion in view emphasizes a balance of innovation with privacy issues for sustainable development and user-oriented governance frameworks of decentralized metaverses. Critical review points to blockchain technology in a decentralized metaverse and Web 3.0 tech, specifically on security, privacy, and governance of both besides scalability. Contributions include an insight into the balance between technological development and the arising challenges of privacy, together with the recommendations for decentralized systems improvement. The presented paper is intended to drive towards the creation of sustainable digital ecosystems for the good of entrepreneurs, technologists, healthcare professionals, creators, and consumers alike. From the result, a total of 160 respondents from the student population of the International Islamic University Malaysia participated in this study, which aimed to investigate the significance of exploring Blockchain and Web 3.0 in the context of sustainable development. The findings reveal that the majority of respondents prioritized the strengthening of privacy and security protocols as the most critical factor. This was followed by the enhancement of blockchain scalability, the development of user-centric governance frameworks, and the facilitation of cross-platform interoperability. These insights underscore the necessity of balancing technological innovation with robust security measures and the preservation of user trust.

Keywords-- *Decentralized Metaverse; Blockchain Technology; Web 3.0; Digital Ownership; Data Privacy; Smart Contracts*

I. INTRODUCTION

The emergence of decentralized metaverse platforms heralds a sea change in the digital ecosystem, where critical problems of centralized systems are being addressed [1], [2], [3]. Centralized metaverse frameworks are prone to data security vulnerabilities, single points of failure, and limited interoperability between virtual environments [4], [5]. Due to the high level of authority that centralized entities wield over the user experience, users are also often deprived of full control over their digital identities, assets, and creative outputs [6]. Decentralization through blockchain is fast rising as the conclusive solution to some of these challenges. It takes the power of control away from a central authority and hands over record levels of power, freedom, and autonomy to every individual over their online identity, assets, and interactions [7], [8], [9]. More importantly, it is enforced through blockchain via the security, transparency, and immutability of digital transactions that would finally put any misgivings with respect to data privacy and ownership to rest.

Other transformational elements in this invention include smart contract integration into the decentralized system [10], [11]. Smart contracts are automated, thereby eliminating any sort of middleman and enabling trust through user-governed governance frameworks [12]. They provide transparent decision-making where users can directly contribute to the many different platforms they may

https://doi.org/10.31436/ijpcc.v12i1.534

be participating in [13]. It is this user-centric model that ensures not only privacy but also introduces inclusivity and innovation, allowing creators, entrepreneurs, and technologists to level the playing field [14]. Further understanding of the impact of blockchain and Web 3.0 technologies shows a new, redefining way for digital interactions [15], [16]. Among the major benefits these technologies bring about, privacy and security are two of the most challenging concerns in the modern age of the internet [17], [18]. They provide interoperability, a key component in creating a robust, decentralized metaverse via easy, secure exchange of information across virtual environments [19], [20]. This is the very necessary shift that should shape a truly interoperable and user-centered digital space where users are not limited to an isolated ecosystem but are able to move across and interact between different platforms with ease [21], [22], [23].

The methodology to understand and further these innovations involves a critical review of the literature on decentralized frameworks, coupled with real-world case studies. We draw practical lessons from the examination of successful implementations of decentralized platforms regarding the challenges and opportunities inherent in this paradigm. These studies bring together academic research and industry perspectives on how decentralization enhances user control, scalability, and governance. These advances have far-reaching implications that go well beyond the individual user. The entrepreneur can grasp new opportunities in developing decentralized applications, the technologist can innovate with greater freedom, the healthcare professional secures and makes more transparent systems for managing patient data, and consumers have an unprecedented level of security and control over their digital lives [24], [25], [26]. Moreover, creators have increased rights to their digital assets, and their intellectual property is better protected and duly remunerated [27], [28]. The decentralized metaverse functions as a virtual ecosystem where users engage within interconnected environments, driven by blockchain and Web 3.0 technologies. It is supported by consensus mechanisms such as Proof of Stake (PoS), smart contracts, and decentralized storage. Applications of the decentralized metaverse include virtual real estate, gaming, education and collaboration, e-commerce, and more [29]. By 2030, the decentralized metaverse market is projected to grow to $87 billion [29]. This growth is largely driven by blockchain integration in tokenizing virtual assets and evolving applications such as gaming, virtual real estate and education. Blockchain technology is expected to reach $1431.54 billion by 2030, offering secure ownership through Non-Fungible Tokens (NFTs), enabling interoperability and ensuring transparency [29]. However, decentralized metaverse requires advancements to overcome challenges such as scalability issues, security risks and high energy use to realize widespread adoption [29], [30], [31]. Advances in blockchain interoperability, optimized consensus mechanisms such as PoS and decentralized resource-sharing networks are crucial to overcome these barriers. This paper is arranged as follows; Section I gives a brief introduction to decentralized metaverse involving blockchain, Web 3.0 and its applications. Section II summarizes the literature review for this work. Section III covers the methodology of the study. Lastly, Section IV show the results from google form survey about exploring innovation and privacy in Decentralized Metaverse Ecosystems.

II. LITERATURE REVIEW

The review begins with an exploration of the Web 3.0 landscape, which highlights core technologies and the challenge of navigating its decentralized nature. Emphasis is placed on blockchain's role in ensuring security, trust, and transparency across digital systems [32]. This foundation is extended by examining Web 3.0 as the future architecture of the internet, underlining decentralization and user ownership as fundamental pillars [33]. These arguments are supported through analytical reviews of blockchain, smart contracts, and decentralized protocols. Further discussion involves the development and implementation of decentralized applications (dApps) using blockchain in Web 3.0 environments. Smart contracts and blockchain infrastructure are seen as essential to secure and automated user-driven systems [34]. A detailed architectural analysis includes Ethereum's platform capabilities and smart contract programming in Solidity. This is followed by an introduction to the TAO framework (Transparency, Autonomy, Optimization) which is proposed as a means of building efficient decentralized systems [35]. Several articles link Web 3.0 technologies with sustainability, suggesting blockchain can support transparent supply chains and decentralized governance while addressing environmental concerns [36]. The literature also explores blockchain's role in transforming the financial landscape through DeFi, tokenized assets, and smart contracts. It emphasizes blockchain's potential for financial innovation [37]. In another sector, supply chain accountability is addressed through enhanced transparency and traceability enabled by blockchain integration [38].The transparency benefits of blockchain in industries such as mineral sourcing are examined through a comprehensive review, especially regarding ethical sourcing and conflict mineral tracking [39]. In service management, a proposed system design utilizing blockchain and NFTs is introduced to manage real-time data transmission within IoT and Metaverse networks [40]. Blockchain's evolution beyond cryptocurrency is emphasized, with its adaptable design applied across domains like finance, healthcare, and logistics [41].

Attention is also directed toward the Metaverse, where blockchain underpins data integrity, identity verification,

https://doi.org/10.31436/ijpcc.v12i1.534

and ownership in immersive digital environments. Blockchain is shown to support virtual assets, secure identities, and integrate with other technologies such as VR, AR, and AI [42]. With the help of edge intelligence, blockchain enhances real-time data processing in Web 3.0 operations [43]. The Industrial Metaverse further expands this by applying Web 3.0 technologies to virtual factories and production systems [44]. Governance models and digital asset ownership in the Metaverse are explored through implementation studies and integration scenarios [45]. The idea of synchronization and continuity of identity and transactions is discussed through case examples like JPMorgan's virtual land acquisition [46]. Historical literature on the development of the Metaverse and the integration of VR, AI, and IoT is referenced to show its broad implications for urban environments and sustainability [47]. The convergence of AI and blockchain within the Metaverse is further emphasized, with a systematic review exploring how this synergy can redefine social, economic, and operational models [48]. DeFi's role in the Metaverse's financial systems is described as transformational, with blockchain eliminating intermediaries and promoting global accessibility to services [49]. Specific technical and case study-based analyses support this, particularly in identifying risks like front-running attacks [50].The literature then explores how AI integration can significantly enhance blockchain functionality by improving user experience, automating decisions, and supporting decentralized applications [51]. In Industry 4.0, the combination of AI and blockchain is presented as a way to improve business processes, social interactions, and decentralized models [52]. Decentralized AI combined with edge intelligence is proposed to boost the performance of blockchain, Web 3.0, and the Metaverse, supporting personalized and adaptive applications [53].

The integration of AI, blockchain, and digital networking technologies is also seen as central to creating immersive virtual environments, especially within social interaction and economic transaction contexts [54]. Progress in deep learning and generative models like GPT-4 demonstrates the capacity of AI to enhance metaverse engagement. However, challenges remain due to current global concerns such as the COVID-19 pandemic.Quantum-enhanced blockchain is proposed as an innovative solution to security and scalability issues. Studies discuss quantum protocols and cryptography, offering a roadmap for blockchain's advancement in secure networks [55]. The evolving media landscape is also addressed, where decentralization allows creators more control over production and distribution in immersive, AI-augmented environments [56]. Finally, a blockchain-based authentication system using Decentralized Identifiers (DIDs) and verifiable credentials is proposed to secure user identity in the Metaverse [57]. Comparative analysis with traditional methods reveals improved privacy and security, though the system's effectiveness depends on the broader reliability of blockchain consensus mechanisms.

III. METHODOLOGY

A. Selection of Studies

The selection process was done to ensure the inclusion of high-quality studies. We used the inclusion and exclusion criteria to filter the papers in the following steps. As shown in Figure 1, in the first stage of the selection process, multiple automated searches were conducted on 3 different digital libraries. After that, we remove the duplicated papers, ineligible or inaccessible papers, and for other reasons. On the next stage, we sort out papers based on its title and keywords. Next, we further filter the papers by reading the abstracts of the previously selected papers to find potential papers with high relevance to our topic. In the fourth stage, we read the introduction and conclusion to narrow down the selected studies. For every excluded paper, we take note of the reasons for its irrelevancy then continue to further examine the selected papers of its quality. Lastly, the final and fifth stage of the process is to read the full paper of the selected studies from the fourth stage. Finally, after a strict selection process, we found 72 relevant studies to be included in our literature review.

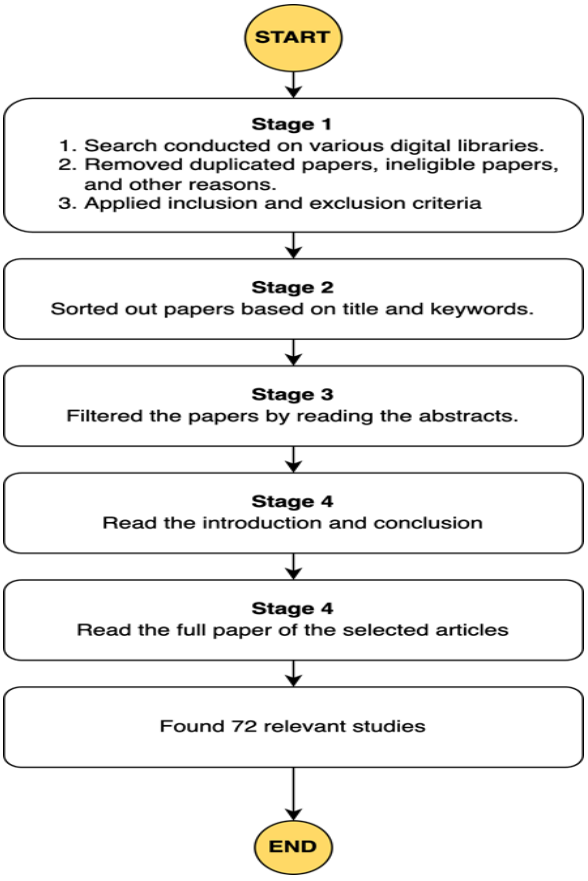


Fig. 1 An overview of the selection process

https://doi.org/10.31436/ijpcc.v12i1.534

B. Inclusion and exclusion criteria

The inclusion and exclusion criteria were decided upon to find relevant studies to our review topic, as shown in Figure 2. The criterion for inclusion is first, studies must be related to decentralized metaverse that addresses blockchain technology and web 3.0 technologies and applications. Next, it must be from either peer-reviewed articles, technical reports, or case studies. Lastly, we must only select studies in the recent five years, which is between 2019 to 2024. Meanwhile, the criterion for exclusion is if the studies are not in English, not related to the decentralized metaverse theme, not available, and not presenting sufficient technical detail on blockchain and web 3.0 technologies.

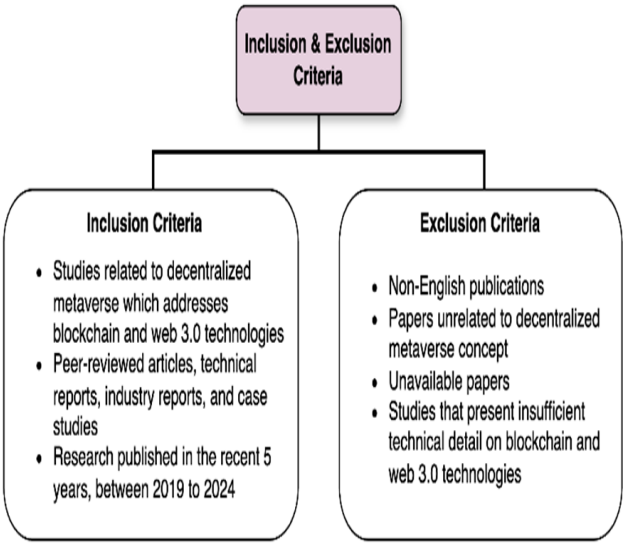


Fig.2 An overview of inclusion and exclusion criteria

c. Data Extraction and Analysis

Data Extraction: We extracted four aspects from the selected literature relevant to the subject matter. Accordingly, we arranged our findings in a systematic tabular document and highlights its key findings/arguments, supporting evidence/methods, and strength and limitations. Analysis: The extracted data were analyzed using thematic synthesis, which is categorization by themes. Major themes included were 1) Blockchain and its application in decentralized systems, 2) Integration of AI and blockchain in Web 3.0, 3) Applications, and 4) Innovations and challenges in Metaverse technologies.

IV. RESULTS

The survey aimed to gather insight from individuals regarding their awareness, concerns, and expectations toward blockchain, Web 3.0, and decentralization within digital ecosystems. A total of 160 respondents participated in the survey. The representation of the outcomes is displayed in Fig. 3 A to Fig.3 J. According to Fig.3 A, the majority of the respondents (93.8%) identified themselves as students, suggesting that early-career individuals form the dominant demographic engaging with decentralized technologies. This underscores the need for foundational educational initiatives and curriculum development tailored to students. In Fig.3 B, the majority of participants rated their familiarity with blockchain and Web 3.0 concepts as moderate, with 43.8% choosing level 3 and 37.5% selecting level 2. This indicates a growing interest but limited technical depth, highlighting the opportunity for targeted literacy programs and skill-building workshops. Fig.3 C illustrates that respondents view several privacy concerns in decentralized metaverses, particularly smart contract vulnerabilities and lack of user-centric governance as highly significant, with many selecting "significant" or "very significant." Meanwhile, Fig.3 D shows that 75% of respondents either agreed or strongly agreed that user autonomy over digital identities and assets is essential. This strongly reflects the value participants place on user rights and decentralized control over personal data.

In terms of innovation, Fig.3 E shows that 68.8% of participants believe decentralization enhances innovation in the metaverse. However, 31.2% were unsure, indicating a need for more demonstrative use cases and exposure. As shown in Fig.3 F, participants selected decentralized governance frameworks (87.5%) as the most promising aspect for sustainable ecosystems, followed closely by privacy-preserving user controls (75%) and transparency & immutability (75%). These responses emphasize governance and user control as core components of trusted digital environments. Despite these optimistic views, Fig.3 G reveals that only 25% of respondents have interacted with blockchain-integrated metaverse platforms, suggesting that practical engagement is still limited and that further adoption support is necessary. As seen in Fig.3 H, industries such as finance & banking (81.3%), healthcare (75%), and education (68.8%) are perceived to benefit the most from decentralized ecosystems, aligning with sectors where transparency, data integrity, and secure access are critical. Concerns about system-level functionality are further addressed in Fig.3 I, where 43.8% of respondents expressed concern level 4 (out of 5) regarding interoperability issues, pointing to the importance of standardization across platforms. Lastly, Fig.3 J presents the ranking of priorities to balance innovation and privacy. The majority chose strengthening privacy/security protocols as the most important priority, followed by improving blockchain scalability, developing user-friendly governance, and ensuring cross-platform interoperability. This indicates that while innovation is valued, it must not compromise security and user trust.

https://doi.org/10.31436/ijpcc.v12i1.534

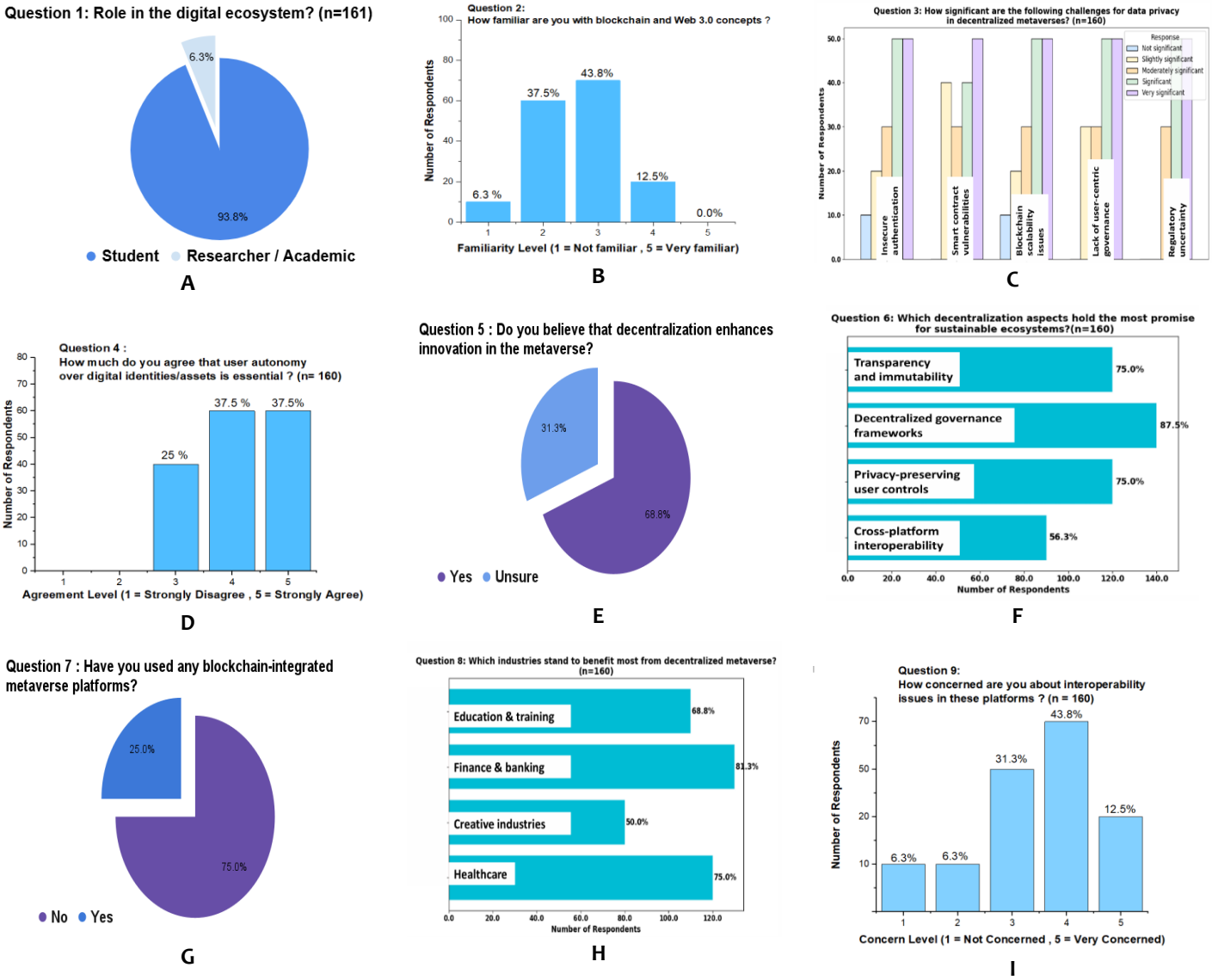


Fig. 3 These bar charts present the collection of responses to a survey titled "Exploring Innovation & Privacy in Decentralized Metaverse Ecosystems"

V. CONCLUSION

In conclusion, decentralized metaverse still has significant progress to make, despite offering greater security and transparency compared to the centralized metaverse. By integrating blockchain systems, Web 3.0 technologies and AI frameworks, the decentralized metaverse aims to create a secure, immersive and user-driven environment. The major advantage of decentralized metaverse lies in its ability to empower users with true ownership, enabling them to create, buy, sell and trade digital assets just like in real life. Decentralized metaverse has introduced transformative shifts across industries and personal experiences, including true digital ownership, new economic opportunities, enhanced privacy and security, decentralized governance and interoperability across platforms. This paper aims to explore sustainable development and user privacy in a user-centric digital environment by emphasizing the importance of continued innovation within decentralized metaverse. Finally, it provides recommendations for addressing challenges such as data integrity, sustainability concerns and high energy consumption through blockchain technology, Web 3.0 and related application.

ACKNOWLEDGMENT

Heartfelt appreciation to our esteemed professors and educators for their steadfast dedication and diligent efforts in imparting invaluable knowledge to us. Their commitment has greatly contributed to our advancement in enhancing our skills and comprehension in the field of Computer Networking, IoT security, and Blockchain technology.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] M. Elsadig, M. A. Alohali, A. O. Ibrahim, and A. W. Abulfaraj, "Roles of Blockchain in the Metaverse: Concepts, Taxonomy, Recent Advances, Enabling Technologies, and Open Research Issues," *IEEE Access*, vol. 12, pp. 38410–38435, 2024, doi: 10.1109/ACCESS.2024.3367014.

[2] T. R. Gadekallu *et al.*, "Blockchain for the Metaverse: A Review," 2022, *arXiv*. doi: 10.48550/ARXIV.2203.09738.

[3] Y. Chen, J. I. Richter, and P. C. Patel, "Decentralized Governance of Digital Platforms," *Journal of Management*, vol. 47, no. 5, pp. 1305–1337, May 2021, doi: 10.1177/0149206320916755.

[4] A. Zainudin, M. A. P. Putra, R. N. Alief, R. Akter, D.-S. Kim, and J.-M. Lee, "Blockchain-Inspired Collaborative Cyber-Attacks Detection for Securing Metaverse," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 18221–18236, May 2024, doi: 10.1109/JIOT.2024.3364247.

[5] Zainuddin, A. A., Muhammad, N. F. A., Hasli, A. H. M., Zaini, N. A. J. A., Karimudin, N. B. B., Bahri, H. S. S., ... & Ghazalli, N. (2023). Empowering smart city governance through decentralized blockchain solutions for security and privacy in IoT communications. Bulletin of Social Informatics Theory and Application, 7(2), 104-117.

[6] M. Shuaib *et al.*, "Land Registry Framework Based on Self-Sovereign Identity (SSI) for Environmental Sustainability," *Sustainability*, vol. 14, no. 9, p. 5400, Apr. 2022, doi: 10.3390/su14095400.

[7] O. Avellaneda *et al.*, "Decentralized Identity: Where Did It Come From and Where Is It Going?" *IEEE Comm. Stand. Mag.*, vol. 3, no. 4, pp. 10–13, Dec. 2019, doi: 10.1109/MCOMSTD.2019.9031542.

[8] V. Srinivas, A. K. Jha, G. Ganesh, V. Nitish, and S. Jadon, "Decentralized User Identity Management using Blockchain," in *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, Vellore, India: IEEE, May 2023, pp. 1–6. doi: 10.1109/ViTECoN58111.2023.10157380.

[9] bin Zainuddin, A. A., binti Mortadza, A. S., & binti Musa, F. E. (2024). Integrating IoT and Blockchain for Enhanced Security: Challenges and Solutions. Data Science Insights, 2(1).

[10] L. Thomas, Y. Zhou, C. Long, J. Wu, and N. Jenkins, "A general form of smart contract for decentralized energy systems management," *Nat Energy*, vol. 4, no. 2, pp. 140–149, Jan. 2019, doi: 10.1038/s41560-018-0317-7.

[11] Zainuddin, A. A., Handayani, D., Ridza, I. H. M., Rahman, S. H. A., Kamarudin, S. I., Ahmad, K. Z., ... & Dhuzuki, N. H. M. (2024, May). Converging for Security: Blockchain, Internet of Things, Artificial Intelligence-Why Not Together? In 2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 181-186). IEEE.

[12] J. Messias, V. Pahari, B. Chandrasekaran, K. P. Gummadi, and P. Loiseau, "Understanding Blockchain Governance: Analyzing Decentralized Voting to Amend DeFi Smart Contracts," 2023, *arXiv*. doi: 10.48550/ARXIV.2305.17655.

[13] Z. Zheng *et al.*, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, Apr. 2020, doi: 10.1016/j.future.2019.12.019.

[14] J. Kang *et al.*, "Blockchain-based Federated Learning for Industrial Metaverses: Incentive Scheme with Optimal AoI," in *2022 IEEE International Conference on Blockchain (Blockchain)*, Espoo, Finland: IEEE, Aug. 2022, pp. 71–78. doi: 10.1109/Blockchain55522.2022.00020.

[15] "Exploration and Practice of Blockchain Technology Application in the Field of Digital Commerce," *AJBM*, vol. 5, no. 24, 2023, doi: 10.25236/AJBM.2023.052403.

[16] S. L. Nita and M. I. Mihailescu, "A Novel Authentication Scheme Based on Verifiable Credentials Using Digital Identity in the Context of Web 3.0," *Electronics*, vol. 13, no. 6, p. 1137, Mar. 2024, doi: 10.3390/electronics13061137.

[17] J. Wu, K. Lin, D. Lin, Z. Zheng, H. Huang, and Z. Zheng, "Financial Crimes in Web3-Empowered Metaverse: Taxonomy, Countermeasures, and Opportunities," *IEEE Open J. Comput. Soc.*, vol. 4, pp. 37–49, 2023, doi: 10.1109/OJCS.2023.3245801.

[18] Annas, A. H., Zainuddin, A. A., Ramlee, A. W., Omar, A. S. Y., Saifuddin, M. H. F. M., Sidik, N. F. I., ... & Ahmadzamani, N. Z. A. (2024). Analyses of 6G-Network and Blockchain-Network Application Security: Future Research Prospect. International Journal on Perceptive and Cognitive Computing, 10(2), 31-50.

[19] G. D. Ritterbusch and M. R. Teichmann, "Defining the Metaverse: A Systematic Literature Review," *IEEE Access*, vol. 11, pp. 12368–12377, 2023, doi: 10.1109/ACCESS.2023.3241809.

[20] T. Li, C. Yang, Q. Yang, S. Zhou, H. Huang, and Z. Zheng, "MetaOpera: A Cross-Metaverse Interoperability Protocol," 2023, *arXiv*. doi: 10.48550/ARXIV.2302.01600.

[21] J. Jaenudin, A. Zahran, and D. Mahdiana, "Blockchain Utilization in Secure and Decentralized Web 3.0 Application Development," *SinkrOn*, vol. 9, no. 1, pp. 594–599, Jan. 2024, doi: 10.33395/sinkron.v9i1.13411.
[22] X. Ren *et al.*, "Building Resilient Web 3.0 with Quantum Information Technologies and Blockchain: An Ambilateral View," 2023, *arXiv*. doi: 10.48550/ARXIV.2303.13050.
[23] S. Rouhani and R. Deters, "Security, Performance, and Applications of Smart Contracts: A Systematic Survey," *IEEE Access*, vol. 7, pp. 50759–50779, 2019, doi: 10.1109/ACCESS.2019.2911031.
[24] F. Schär, "Decentralized Finance: On Blockchain- and Smart Contract-based Financial Markets," *SSRN Journal*, 2020, doi: 10.2139/ssrn.3571335.
[25] M. Pustisek, J. Turk, and A. Kos, "Secure Modular Smart Contract Platform for Multi-Tenant 5G Applications," *IEEE Access*, vol. 8, pp. 150626–150646, 2020, doi: 10.1109/ACCESS.2020.3013402.
[26] M. Pustisek, J. Turk, and A. Kos, "Secure Modular Smart Contract Platform for Multi-Tenant 5G Applications," *IEEE Access*, vol. 8, pp. 150626–150646, 2020, doi: 10.1109/ACCESS.2020.3013402.
[27] J. Wu, K. Lin, D. Lin, Z. Zheng, H. Huang, and Z. Zheng, "Financial Crimes in Web3-Empowered Metaverse: Taxonomy, Countermeasures, and Opportunities," *IEEE Open J. Comput. Soc.*, vol. 4, pp. 37–49, 2023, doi: 10.1109/OJCS.2023.3245801.
[28] M. Kalyvaki, "Navigating the Metaverse Business and Legal Challenges: Intellectual Property, Privacy, and Jurisdiction," *Journal of Metaverse*, vol. 3, no. 1, pp. 87–92, Jun. 2023, doi: 10.57019/jmv.1238344.
[29] A. Ghosh, Lavanya, V. Hassija, V. Chamola, and A. El Saddik, "A Survey on Decentralized Metaverse Using Blockchain and Web 3.0 Technologies, Applications, and More," *IEEE Access*, vol. 12, pp. 146915–146948, 2024, doi: 10.1109/ACCESS.2024.3469193.
[30] O. Hashash, C. Chaccour, W. Saad, K. Sakaguchi, and T. Yu, "Towards a Decentralized Metaverse: Synchronized Orchestration of Digital Twins and Sub-Metaverses," in *ICC 2023 - IEEE International Conference on Communications*, Rome, Italy: IEEE, May 2023, pp. 1905–1910. doi: 10.1109/ICC45041.2023.10279406.
[31] M. Skorokhod, "Challenges on the Way to a Secure and Decentralized Metaverse".
[32] K. Vayadande, A. Baviskar, J. Avhad, S. Bahadkar, P. Bhalerao, and A. Chimkar, "A Comprehensive Review on Navigating the Web 3.0 Landscape," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, Bangalore, India: IEEE, Jun. 2024, pp. 456–463. doi: 10.1109/ICICI62254.2024.00080.
[33] U. W. Chohan, "Web 3.0: The Future Architecture of the Internet?," *SSRN Journal*, 2022, doi: 10.2139/ssrn.4037693.
[34] A. Sahu, D. P. Mishra, S. B. Mohanty, and P. P. Sahu, "Web 3.0 Decentralized Application Using Blockchain Technology," in *2023 4th International Conference on Computing and Communication Systems (I3CS)*, Shillong, India: IEEE, Mar. 2023, pp. 1–6. doi: 10.1109/I3CS58314.2023.10127552.
[35] J. Li and F.-Y. Wang, "The TAO of Blockchain Intelligence for Intelligent Web 3.0," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 12, pp. 2183–2186, Dec. 2023, doi: 10.1109/JAS.2023.124056.
[36] S. Rathor, M. Zhang, and T. Im, "Web 3.0 and Sustainability: Challenges and Research Opportunities," *Sustainability*, vol. 15, no. 20, p. 15126, Oct. 2023, doi: 10.3390/su152015126.
[37] C. Chen *et al.*, "When Digital Economy Meets Web3.0: Applications and Challenges," *IEEE Open J. Comput. Soc.*, vol. 3, pp. 233–245, 2022, doi: 10.1109/OJCS.2022.3217565.
[38] H. Achmad Bagraff, N. Kholis, . M., and F. Ghofi Nabila, "Implementing Blockchain Technology for Optimized Supply Chain and Enhanced Sustainability," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 1697–1702, Sep. 2024, doi: 10.38124/ijisrt/IJISRT24AUG1218.
[39] M. Onifade, J. A. Adebisi, and T. Zvarivadza, "Recent advances in blockchain technology: prospects, applications and constraints in the minerals industry," *International Journal of Mining, Reclamation and Environment*, vol. 38, no. 7, pp. 497–533, Aug. 2024, doi: 10.1080/17480930.2024.2319453.
[40] T. Maksymyuk, J. Gazda, G. Bugar, V. Gazda, M. Liyanage, and M. Dohler, "Blockchain-Empowered Service Management for the Decentralized Metaverse of Things," *IEEE Access*, vol. 10, pp. 99025–99037, 2022, doi: 10.1109/ACCESS.2022.3205739.
[41] G. Habib, S. Sharma, S. Ibrahim, I. Ahmad, S. Qureshi, and M. Ishfaq, "Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing," *Future Internet*, vol. 14, no. 11, Art. no. 11, Nov. 2022, doi: 10.3390/fi14110341.
[42] H. Wang *et al.*, "A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14671–14688, Aug. 2023, doi: 10.1109/JIOT.2023.3278329.
[43] Y. Lin *et al.*, "A Unified Blockchain-Semantic Framework for Wireless Edge Intelligence Enabled Web 3.0," *IEEE Wireless Commun.*, vol. 31, no. 2, pp. 126–133, Apr. 2024, doi: 10.1109/MWC.018.2200568.
[44] S. Zhang *et al.*, "Industrial Metaverse: Enabling Technologies, Open Problems, and Future Trends," 2024, *arXiv*. doi: 10.48550/ARXIV.2405.08542.
[45] C. Hackl, D. Lueth, and T. D. Bartolo, *Navigating the Metaverse: A Guide to Limitless Possibilities in a Web 3.0 World*. John Wiley & Sons, 2022.
[46] M. Etemadi and J. Yadollahi Farsi, "The Potential of Blockchain Technology in Building the Decentralized World of Metaverse: a Scientometric Study and Study Clusters in the Metaverse Field," *SSRN Journal*, 2023, doi: 10.2139/ssrn.4547579.
[47] S. S. Thakur, S. Bandyopadhyay, and D. Datta, "The Metaverse as a Virtual Form of Smart Cities: Opportunities and Challenges," *IJCA*, vol. 185, no. 17, pp. 45–53, Jun. 2023, doi: 10.5120/ijca2023922892.
[48] P. P. Momtaz, "Some Very Simple Economics of Web3 and the Metaverse," *FinTech*, vol. 1, no. 3, pp. 225–234, Jul. 2022, doi: 10.3390/fintech1030018.
[49] Tamimul Alam, Md. Asraf Ali, and Md. Hasibur Rahman, "Front-running attack in decentralized finance in the metaverse: A systematic review," *Int. J. Sci. Res. Arch.*, vol. 11, no. 1, pp. 2315–2324, Feb. 2024, doi: 10.30574/ijsra.2024.11.1.0332.
[50] S. Dos Santos, J. Singh, R. K. Thulasiram, S. Kamali, L. Sirico, and L. Loud, "A New Era of Blockchain-Powered Decentralized Finance (DeFi) - A Review," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Los Alamitos, CA, USA: IEEE, Jun. 2022, pp. 1286–1292. doi: 10.1109/COMPSAC54236.2022.00203.
[51] M. Shen *et al.*, "Artificial Intelligence for Web 3.0: A Comprehensive Survey," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 1–39, Oct. 2024, doi: 10.1145/3657284.
[52] M. Gebert and E. Association, *AI - Powered Blockchain Technology in Industry 4.0 Exploring the Transformative Synergy of AI and Blockchain Technologies*. 2024. doi: 10.13140/RG.2.2.16929.21600.
[53] L. Cao, "Decentralized AI: Edge Intelligence and Smart Blockchain, Metaverse, Web3, and DeSci," *IEEE Intell. Syst.*, vol. 37, no. 3, pp. 6–19, May 2022, doi: 10.1109/MIS.2022.3181504.
[54] KIT-Kalaignar Karunanidhi Institute of Technology and S. G A, "STUDY OF BLOCKCHAIN TECHNOLOGY, AI AND DIGITAL NETWORKING IN METAVERSE," *IJEAST*, vol. 6, no. 9, pp. 166–169, Jan. 2022, doi: 10.33564/IJEAST.2022.v06i09.020.
[55] M. Xu *et al.*, "When Quantum Information Technologies Meet Blockchain in Web 3.0," *IEEE Network*, vol. 38, no. 2, pp. 255–263, Mar. 2024, doi: 10.1109/MNET.134.2200578.
[56] M. Aytas and A. Can, "From real spaces to virtual spaces: The metaverse and decentralized cinema," *DRArch*, vol. 3, no. (Special Issue), pp. 49–59, Dec. 2022, doi: 10.47818/DRArch.2022.v3si070.
[57] M. Kim, J. Oh, S. Son, Y. Park, J. Kim, and Y. Park, "Secure and Privacy-Preserving Authentication Scheme Using Decentralized Identifier in Metaverse Environment," *Electronics*, vol. 12, no. 19, p. 4073, Sep. 2023, doi: 10.3390/electronics12194073.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.569

Islamization of Technology: The Qur'anic Guidance and Sunnah in ICT Integration

Masuk Mia[1], Mohammad Raihanul Islam[2], Fazeel Ahmed Khan[3]*
[1,2,3]Kulliyyah of Information and Communication Technology,
International Islamic University Malaysia (UIAM), 53100, Kuala Lumpur, Malaysia.

*Corresponding author: fazeelahmedkhan15@gmail.com
(Received: 9th February 2025; Accepted: 22nd October, 2025; Published on-line: 30th January, 2026)

Abstract— The rapid development in Information and Communication Technology (ICT) has revolutionized modern day life, influencing communication, education, commerce and social interactions. The concerns related to ethical issues are still prevalent and results into challenges such as, privacy violations, disinformation and the exploitation of digital platforms which emphasizes the necessity to have a moral framework based on Islamic ethical principles. The proposed study examines the Islamization of Technology by integrating teachings from The Holy Qur'an and Sunnah of Prophet Muhammad (SAW) to harmonize technological progress with ethical and spiritual principles. It advocates for the creation of Islamic digital platforms which supports integrity, privacy and responsible content moderation while fostering truthful communication, ethical business practices and social welfare aligned with *Maqasid al-Shari'ah*. This approach integrates Islamic values into ICT to harmonize technological progress with Islamic ethics to ensure that digital advancements facilitate spiritual growth, knowledge diffusion and ethical governance. The Islamization of Technology perceives ICT as a tool to foster ethical conduct, improve community cohesion and tackle modern digital issues through a faith-oriented approach.

Keywords— Islamic ethics, Islamization of technology, Maqasid al-Shari'ah, Ethical Technology, Spiritual Growth through ICT.

I. INTRODUCTION

This study focuses on the concept of Islamization of knowledge based on the framework as defined in The Holy Qur'an and Sunnah of Prophet Muhammad ﷺ which serves as the major sources of guidance in the Islamization of any modern knowledge and a primary spirit. It has application rooted into any policy with the intent to integrate Information and Communication Technology (ICT) to improve information science and knowledge [1]. The main goal is to encourage the IT practitioners to have a closer look at the current developments in ICT, correlation with Islamic ethical principles and its influence on Muslim societies across the globe. The growing inventions has pushed human productivity to think on the need to explore whether the pace is moving under the guidance of Islamic teachings and how it is impacting Muslim societies across the globe [2]. It goes without saying that most people on this planet have experienced technological advancement in different stages of their life. The progress in technological innovation which enhances human productivity needs to be validate whether these developments align with Islamic teachings and ethics [3]. Technology has transformed every aspect of human life including education, transportation, medicine, space and media which enables better possibilities for human progress. However, these advancements should be in compliance with Islamic ethical teachings to ensure a positive contribution to the society.

A. The role of Islam in guiding technological advancement

ICT plays a crucial role in shaping the modern world and influencing digital technology. It has transformed lives, thought processes, human productivity and global economies. The Islamic teachings defines that new technologies should be assessed based on the ethical and moral principles found in the Holy Qur'an and the Sunnah of Prophet Muhammad ﷺ. Similarly, this assures alignment with Islamic ethical values and contributes to the betterment of individual lives [4]. Islam includes all facets of life including the integration of technology in professional, educational and daily pursuits. The basis of knowledge and advancement in Islam has its roots in divine guidance primarily sourced from The Holy Qur'an, succeeded by *Sunnah* of Prophet Muhammad ﷺ, *Ijma* and *Qiyas* and *Ijtihad* [5]. It offers a thorough framework for navigating progress in science and technology while upholding ethical integrity. Islam functions as a foundational principle for work ethics, discipline and productivity, providing a systematic

framework for navigating and excelling in everyday life [6]. It promotes creativity, collaboration and specialization, highlighting the quest for knowledge and quality across all disciplines. Islam instructs individuals on the creation, accumulation and processing of earthly resources while simultaneously acting as a catalyst for advancement, motivating Muslims to pursue righteousness in opposition to immoral behaviors [7]. By combining Islamic teachings with technical advancements, Muslims can properly leverage the capabilities of ICT, ensuring that innovation serves humanity while maintaining moral and ethical standards. This method promotes collaboration, proficiency and specialization, allowing society to evolve in a technologically sophisticated yet spiritually oriented manner [8].

B. Islamization of technology

The Islamization of technology signifies the integration of Islamic principles, ethics and values into the development, application and adoption of modern technological advancement. It emphasizes that technological improvements must not only facilitate economic and industrial development but also correspond with ethical, social and spiritual welfare [9]. It is rooted in the Holy Qur'an and Sunnah of Prophet Muhammad ﷺ which aims to promote the responsible, ethical and beneficial use of technology for the progress of mankind. In ICT, the Islamization of technology means developing digital platforms, media and computer systems which follows Islamic ethical values, promote essential knowledge and prevent harmful or immoral uses [10]. This methodology promotes innovation among Muslim scholars, engineers and legislators while preserving a harmony between scientific advancement and Islamic principles to make sure that the technology positively impacts both this world and the hereafter [11].

C. Navigating diversity among Islamic Legal Schools of scholarship for ICT integration

The four major Sunni legal schools i.e. *Hanafi*, *Maliki*, *Shafi'i* and *Hanbali*, differ in their methodologies for deriving Islamic law which can lead to diverse opinions on contemporary issues including related to technology and ICT. These differences originate from their diverse significance on sources of law and interpretative principles. The *Hanafi* school emphasizes analogical reasoning (*qiyas*) along with juristic preference (*istihsan*) when direct evidence is unavailable. It's adoption towards technology generally seen as most flexible and rational which allows for contextual and responsive application of law to changing items. Similarly, this flexibility can lead to more accommodating rulings on new technologies and digital practices if they align with Islamic ethical principles and do not involve prohibited elements. Similarly, the *Maliki* school places significant emphasis on the practices of Medina Munawara (*Amal-Ahl-al-Medina*) by considering them a strong reflection of the Sunnah of Prophet ﷺ with Qur'an, Sunnah and Consensus (*ijma*). It's approach towards technology tends to be more conservative in some respects due to its reliance on the established practices of early Muslim community of Medina. However, there are allowable independent interpretation (*ijtihad*) to address new legal issues and rulings on technology which might be considered for its societal impacts and align with established communal norms in a better way.

The *Shafi'i* school is known for its systematic and rigorous approach to jurisprudence by combining a textual approach with rational analysis. It prioritizes the Qur'an followed by the Sunnah of Prophet ﷺ, Consensus (*ijma*) and analogical reasoning (*qiyas*). It aims to balance textual evidence with rational deduction which can lead to detailed rulings by carefully weight scriptural directives with practical realities and can benefit technological advancement seeking to establish clear legal precedents. Lastly, the *Hanbali* school is known as the most textualist and strict school emphasizing the Qur'an and Hadith as primary sources and tends to be skeptical on the extensive use of juristic analogical reasoning. It is often considered more conservative due to its literal interpretation which can lead to more cautious or restrictive rulings on technologies. It might not address those technological applications or interpretation not explicitly found in Islamic foundational texts thus prioritizing adherence to clear scriptural injunctions and avoid speculative interpretations.

II. LITERATURE REVIEW

The Islamization of technology is a developing academic subject which examines how technology can be integrated with Islamic ethics, beliefs, and philosophy for the advancement in technology [12]. This notion is fundamentally grounded in the Islamic epistemological framework, which emphasizes on the equilibrium between revealed knowledge (*Naqli*) from the Holy Qur'an and Sunnah of Prophet Muhammad ﷺ, the rational knowledge (*Aqli*) obtained via scientific investigation. The academics contend that technology must not operate in a moral void but should be compatible with Islamic principles to ensure its appropriate advancement and utilization for mankind [13].

A. The role of knowledge in Islam and its ethical application

The Holy Qur'an emphasize the significance of knowledge (*Il'm*) as the cornerstone of human advancement. The first revelation of the Holy Qur'an commences with the directive

"Read with the name of your Lord who created (everything)"[1], highlighted the significance of knowledge acquisition. Similarly, Islam promotes scientific research, innovation and discovery, if they are morally directed and enhance the welfare of humanity. The great scholars including, Al-Ghazali (11th century) and Ibn Khaldun (14th century) emphasize the significance of ethical reasoning (*ijtihad*) and consensus (*ijma*) in the development of science and technology while upholding Islamic moral principles. Also, many contemporary scholars including, Ziauddin Sardar (1985) and Seyyed Hossein Nasr (1996) emphasize that the pursuit of knowledge which encourages scientific and technological progress should be manage with responsibility and intent. They assert that technological advancements when misapplied might result in issues in ethics, exploitation and adverse societal repercussions including economic inequality, disintegration of cultural identity and moral degradation [14]. Similarly, the Islamization of Technology aims to ensure that scientific progress fosters social justice, ethical purity and spiritual wellness for human beings.

B. *The intersection of technology and Islamic principles*

The massive progress in ICT has intensified discourse on Islamic digital ethics. The researchers assert that ICT needs to be developed in a manner which facilitates access to Islamic education through digital platforms, e-learning and mobile applications [15]. Also, it advocates for the truthfulness of digital information and rejects disinformation in the digital realm. The facilitation of online ethical engagements should be consistent with Islamic principles of integrity, respect and modesty [16]. The compliance towards Islamic financial principles is essential which provides equitable and ethical business transactions in the digital economy. The *Maqasid al-Shari'ah* establishes a framework for assessing whether technology contributes to the sustenance of individuals Faith (*Deen*), Life (*Nafs*), Intellect (*'Aql*), Lineage (*Nasl*), and Wealth (*Maal*) [17]. This principle asserts that technology must be developed and utilized to preserve human dignity, promote social welfare and ensure social justice.

III. Ethical Framework for Islamization of ICT

The Islamization of ICT is governed by core concepts which ensure advancement in technology is consistent with Islamic values and moral principles. The primary concept of *Shari'ah* Compliance requires that all technological developments must conform to Islamic teachings, avoiding content or applications that violate moral and ethical standards [18]. Secondly, purposeful innovation asserts that technology should be produced with a definitive aim to enhance knowledge, ethical conduct and societal welfare. The community involvement highlights the importance of teamwork between Islamic scholars and ICT experts. This helps make sure that technological advances meet the needs of Muslim communities and follow Islamic teachings [19]. Similarly, the Islamization of ICT can be executed through diverse practical applications including Islamic educational platforms which can offer online resources for The Holy Qur'an, Sunnah of Prophet Muhammad ﷺ and Islamic Law to keep Islamic education accessible to worldwide, digital *Da'wah* initiatives to develop social media campaigns and online outreach programs serve to promote Islamic ethical values and avoid disinformation. The Halal e-commerce platforms facilitate ethical business transactions by endorsing fair trade, transparency and interest-free financial services. Also, the AI and machine learning-driven content moderation tools can effectively filter vulnerable content, safeguarding users from exposure to improper or misleading information. The Islamization of ICT aims to create a modern and ethical digital environment for Muslim communities by using practical technology solutions.

A. *Islamic Epistemology: Technology and Maqasid al-Shari'ah*

The Islamic teachings emphasize on the ethical use of technology to assure it aligns with the moral and spiritual values of Islam. The Islamic epistemology encourages to seek knowledge which can strengthen belief in Allah ﷻ and can benefit human beings. The *Maqasid al-Shari'ah* or the Objectives of Islamic law defines an ethical use of technology, provides a framework for the ethical and responsible use of engineering and technology [20].

1) Faith (Deen): The development and use of technology should maintain and improve spiritual values and practices to ensure that technology should support spiritual practices and values. For example, the ICT can improve Islamic education by providing access to resources for Qur'an learning platforms, prayer time apps and online communities for spiritual discussions. The digital applications which promote Islamic values while upholding ethical guidelines to ensure that it strengthen the faith of people instead of distraction.

2) Life (Nafs): The development of technology should prioritize the protection of human life by improving the well-being of the people. The innovations in healthcare, cybersecurity, medicine, disaster management etc., can contribute significant progress in this goal to improve the safety, health and living of all human beings. For example, the AI-driven disease detection, health monitoring devices and emergency response system can give significant assistance to protect and improve human life.

[1] Surah Al-Alaq (96: 1-5)

3) Intellect ('Aql): The development in ICT should promote intellectual development by spreading knowledge and information which transforms human life and has a positive impact on society. Similarly, it should avoid misinformation, encourage critical thinking and integrate Islamic ethics into educational structure. The online learning resources should be based on Islamic teachings to provide a balanced and moral education which helps the students to improve their faith in Allah ﷻ and can support to live their life based on Islamic Shari'ah.

4) Lineage (Nasl): Technological developments should protect family and social structure which include content moderation on digital and social media spectrum to prevent the spread of harmful information. The media platforms which promote ethical engagements and provide tools to strengthen human values by encouraging positive relationships and responsibilities, discourages activities which are against Islamic teachings.

5) Wealth (Mal): The ICT development should support ethical distribution of wealth to promote ecommerce and ethical business practices supporting transparent financial transactions. The blockchain, digital banking and Fintech should ensure fairness and accountability in financial transactions to avoid *riba'* (Usury), uncertainty and gambling in accordance with the Islamic teachings.

B. Methodology

This study develops a methodology to examine the integration of Islamic ethics into ICT. The method focuses on looking at Islamic moral principles and finding the best ways to adapt it in compliance with Islamic principles. This study integrates the thematic analysis of The Holy Qur'anic verses and Sunnah of Prophet Muhammad ﷺ with practical case studies to offer comprehensive knowledge of how Islamic principles can inform the ethical creation and deployment of technology. This methodology ensures the alignment of both technological foundations and functional uses of ICT with Islamic ethical principles. Similarly, the first step involves looking closely at Qur'anic texts and Hadiths to identify basic moral ideas related to ICT.

The proposed study describes a persistent theme which includes justice, accountability, privacy and the ethical distribution of knowledge. The contemporary ICT practices use contextual interpretation to ensure transparency in AI systems, promote ethical content regulation and uphold digital privacy. The Holy Qur'an mentioned caution against espionage and disseminating misinformation by saying *"O you who believe, abstain from many of the suspicions. Some suspicions are sins. And do not be curious (to find out faults of others), and do not backbite one another. Does one of you like that he eats the flesh of his dead brother? You would abhor it. And fear Allah. Surely Allah is Most-Relenting, Very-Merciful."* [2], which is directly relevant to data privacy and ethical media practices. Also, advocating for the pursuit of beneficial knowledge by the Hadith of Allah's Messenger ﷺ said, *"Seeking knowledge is a duty upon every Muslim, and he who imparts knowledge to those who do not deserve it, is like one who puts a necklace of jewels, pearls and gold around the neck of swines."* [3] endorsing the advancement of Islamic educational platforms and AI-driven ethical tools. This stage creates a strong ethical guide for integrating Islamic values into technology [21]. It connects key ideas with important areas of ICT such as AI ethics, cybersecurity, digital content regulation and moderation.

The second part involves looking at case studies of successful Islamic ICT platforms to see how well they work and what challenges they face. The platforms such as Muslim Pro [4], which incorporate prayer times, The Holy Qur'anic recitations, and Islamic content while ensuring a user-friendly and ethical design analysed for best practices. Similarly, the *Zakat* calculators help Muslims accurately figure out how much they should give to charity based on Islamic guidelines, showing how useful technology can be for doing the right thing. Also, the Halal e-commerce platforms ensure adherence to Islamic business principles via interest-free transactions, Halal certifications, and ethical consumer practices. Additionally, the case study findings show helpful ways to combine Islamic values with technology development, as well as challenges in creating designs that focus on user needs and ethics. These findings will help make future suggestions, including developing AI tools for fair content moderation and creating Islamic finance products which follows Shariah-compliant digital ethics.

1) Sources of Data

An effective way to gather information is important to make sure this study is based on fundamental Islamic principles and includes essential ideas from ICT experts. The main sources of information are The Holy Qur'an, Sunnah of Prophet Muhammad ﷺ, *ijma* and *qiyas* which provides an Islamic foundation for ethical practices in ICT. The Qur'anic verses can be examined to identify fundamental concepts concerning truth, justice, and damage prevention, as mentioned in Holy Qur'an which says: *"And say, 'Truth has come and falsehood has vanished. Falsehood is surely bound to vanish.'"* [5] and in another verse which says *"When Mūsā sought water for his people, We said, 'Strike the rock with your staff,' And twelve springs gushed forth from it. Each*

[2] Surah Al-Hujurat (49: 12)
[3] Sunan Ibn Majah (Vol. 1, Book 1, Hadith 224)
[4] Muslim Pro (www.muslimpro.com)
[5] Surah Al-Isra (17: 81)

group of people came to know their drinking place. 'Eat and drink of what Allah has provided, and do not go about the earth spreading disorder'" [6]. Similarly, the Islamic principles on business ethics, such as banning deception in trade as Allah's Messenger ﷺ said "*The seller and the buyer have the right to keep or return goods as long as they have not parted or till they part; and if both the parties spoke the truth and described the defects and qualities (of the goods), then they would be blessed in their transaction, and if they told lies or hid something, then the blessings of their transaction would be lost."* [7], will support the development of financial and business technologies which follow *Shari'ah.*

Also, the *ijma* refers to the unanimous consensus of Muslim scholars (*mujtahids*) on a particular legal issue in a specific era after Prophet Muhammad ﷺ. It is considered a strong source because it signifies a collective agreement based on thorough scholarly thinking [22]. When a clear consensus is reached by qualified Islamic scholars on a particular matter then it becomes binding for subsequent generations. However, attaining absolute *ijma* in complex modern issues can be challenging due to the intellectual diversity and scholarly opinions in the Muslim world. The *qiyas* is a method of deriving a legal ruling for a new issue by drawing an analogy from a similar issue that already has an established ruling in the Qur'an or Sunnah. It involves identifying a common effective cause (*illah*) between the two issues [23]. It is crucial for addressing contemporary issues that did not exist during the time of the Prophet Muhammad ﷺ. Also, this study can be used with thematic analysis of these scriptures to guarantee that ICT frameworks conform to Islamic moral standards directing the ethical design and deployment of digital instruments. Additionally, it can be used to conduct a comprehensive evaluation of current Islamic digital platforms to determine their conformity with Islamic ethical principles and user requirements. This entails assessing Qur'anic applications for content authenticity, user-friendliness and its compliance with Islamic principles. Considering a look at halal investment platforms to see if they follow *Shari'ah* guidelines including avoiding interest and making ethical investments. The Islamic educational platforms can be evaluated to determine their efficacy in integrating Islamic teachings with contemporary pedagogical approaches, facilitating effective digital learning for Muslim students.

2) *Practical Implementation*

The existing discourse on the Islamization of technology often proposed compelling ethical frameworks and philosophical understandings. However, a significant gap exists in translating these high-level principles into tangible ICT tools, algorithms and software architectures. The vision for ethically aligned digital platforms is clear while the practical aspect remains largely unexplored. A comprehensive analysis requires delve into the specific technical components which is necessary to build, operate and maintain systems that genuinely integrate Islamic values moving beyond theoretical approach to actionable implementation strategies [24]. The type and nature of ICT tools essential to integrate Islamic ethics involves analyzing how existing technologies might be adapted or new ones can be developed. For example, a secure communication tools is a need to have a robust encryption and authentication mechanisms to uphold Islamic principles of privacy and trust potentially leveraging decentralized architectures to minimize single points of data failure or compromise. Similarly, platforms for Islamic education or community engagement can integrate open-source learning management systems (LMS), customized to filter content based on consensus from Islamic scholars and promote collaborative knowledge-seeking [25].

A detail analysis of algorithms is crucial to integrate Islamic ethical principles into it such as justice, fairness, and truthfulness which must be embedded directly into the programmable logic that drives digital systems. It means that the analysis on how algorithms for content moderation can be designed to identify and filter misinformation or harmful narratives based on Islamic ethical guidelines rather than entirely on secular metrics [26]. In Islamic finance technology (fintech), the algorithms for transaction processing will need to rigorously enforce *Shari'ah* compliance to ensure the absence of interest (*riba'*), excessive uncertainty (*gharar*), and speculative practices (*maysir*) potentially through smart contracts on blockchain. Moreover, the software architectures need to be built upon Islamic ethical requirements such as, advocating for privacy-by-design principles where data minimization and user control are built into the system from the ground up. Also, the architectures can prioritize transparency and auditability to allow for external verification of compliance with Islamic principles. The decentralized autonomous organizations (DAOs) can be explored for their potential to develop community governance and accountability aligned with Islamic principles of collective responsibility and consultative decision-making (*shura*)

3) *Feasibility and Challenges*

To effectively address the feasibility and challenges of Islamization of technology, a critical challenge is the cost associated with developing and maintaining such platforms. Different from mainstream technologies which can benefit

[6] Surah Al-Baqarah (2: 60)

[7] Sahih Bukhari, Vol. 3, Book 34, Hadith 293

from large markets and economies of scale however, solutions-based on Islamic ethical principles often faced challenges requiring significant initial investment without immediate guarantees of its immediate adoption. The development process might involve specialized *Shari'ah* auditing and compliance checks which can incur additional expenses. Also, the sustainable funding models including *Waqf*-based financing, community crowdfunding or ethical venture capital should be explored as potential solutions to mitigate these financial barriers and to ensure long-term viability.

Similarly, the lack of talent pool equipped with technical skills combined with Islamic knowledge poses significant challenges. There is a limited pool of professionals who possess both deep expertise in advanced ICT e.g., AI, blockchain, cybersecurity and good understanding of Islamic jurisprudence and ethics. This dual competency is vital for designing systems which genuinely integrate Islamic principles into their core architecture and functionality [27]. It can be mitigated to adopt strategies to cultivate this talent such as interdisciplinary academic programs, specialized training initiatives and collaborative platforms which bring together Islamic scholars and technology experts. Also, interoperability with global systems presents a complex set of challenges. It often needs to interact seamlessly with a global digital infrastructure which is largely built on secular legal and ethical frameworks. This includes data exchange protocols, payment gateways and communication standards which will be compatible with current standards while simultaneously upholding Islamic principles specifically concerning data privacy, user consent and ethical data monetization [28]. Developing open standards for data governance based on Islamic ethical principles and advocating for ethical tech policies that recognize Islamic values to ensure these platforms can function effectively without compromising their core Islamic identity.

C. *Target Audience*

The proposed study aims to serve to two primary audiences i.e. ICT practitioners and developers and Islamic scholars. The ICT practitioners and developers can translate Islamic ethical principles into practical implications for system design, algorithms and software architecture by providing actionable insights rather than purely theoretical concepts. For Islamic scholars, the study demonstrates a solid foundation towards contemporary technological challenges and present potential applications of traditional ethical frameworks to modern technological challenges to encourage further academic discourse. The clear articulation of this dual focus can maintain a balanced perspective, offering both a conceptual framework for scholars and practical guidance for those involved in technology development. The implementation of these ideas will enable the development of ICT to connect more closely with Islamic values to ensure that technology serves mankind in an ethical and responsible manner.

D. *Case Studies*

The concept of Islamization of Technology has moved beyond theoretical discussions into tangible and successful applications across various sectors. Many case studies have demonstrated effective integration of Islamic principles into modern digital solutions particularly in finance, investment, education and lifestyle domains. These initiatives showcase how technology can be harnessed to uphold ethical standards, enhance social welfare and facilitate religious practice in a contemporary context, proving the viability and impact of this evolving field.

1) *Islamic Banking Platforms*

The Meezan Bank [8] is Pakistan's premier Islamic bank and a leading example of the Islamization of finance through digital platform. It is established as a full-fledged Islamic commercial bank in 2002 and has been at the forefront of developing *Shari'ah* compliant financial products and services. It's digital transformation efforts include offering internet banking, mobile banking apps and leveraging technology to facilitate interest-free (*riba'*-free) transactions based on Islamic modes of finance such as *Murabaha* (cost-plus financing), *ijarah* (leasing) and *Musharakah* (partnership financing). The bank emphasizes transparent operations and rigorous *Shari'ah* supervision to ensure all digital offerings align with Islamic ethical standards making Islamic banking accessible to a wider population through modern digital channels. It was also one of the first banks globally to use biometric technology in its ATMs to improve security and user experience [29]. Similarly, the Islami Bank Bangladesh Limited (IBBL) [9] holds the distinction of being the first Islamic bank in Bangladesh which was established in 1983. It has played an important role in popularizing *Shari'ah* compliant banking in the country leading to the establishment of several other Islamic banks. IBBL offers a full range of commercial banking services which include deposits, investments and foreign exchange, all compliant strictly to *Shari'ah* principles. The bank has progressively adopted digital technologies including mobile banking, internet banking and digital wallets to enhance financial inclusion and service delivery in both urban and rural areas. The IBBL success lies not only in mobilizing deposits from segments previously avoids engaging with interest-based banks but also demonstrates the viability and effectiveness of Islamic banking products

[8] Meezan Bank (www.meezanbank.com)

[9] IBBL (www.islamibankbd.com)

through widespread acceptance along with the introduction of various social welfare-based investment schemes [30]. Also, it has been active in forming associations for Islamic banks and foundations for social welfare activities showcasing a comprehensive approach to Islamic financial principles.

Moreover, the Bank Islam Malaysia Berhad [10] which was established in 1983 as Malaysia's first Islamic bank has significantly integrate technology to improve its *Shari'ah*-compliant offerings. The bank is actively developing a 100% digital bank proposition through its Centre of Digital Experience (CDX), leveraging next-generation technologies such as Cloud-Native Digital Banking and Electronic Know Your Customer (eKYC) which aims to provide branchless banking for greater accessibility. It has partnered with technology firms such as Mambu (SaaS banking platform) and Experian (eKYC) to configure *Shari'ah*-compliant products and enable seamless account opening *[31]*. Also, the bank is developing alternative credit scoring models in collaboration with fintech players e.g. Pod, specifically targeting less served segments such as gig workers, to promote financial inclusion guided by Islamic principles of fairness and risk-sharing.

2) Islamic Fintech and Investment Platforms

The Zoya[11] and Musaffa[12] are prominent case studies of applications which facilitate *Shari'ah*-compliant investing platforms. Both platforms utilize technology to help Muslim investors identify and manage portfolios that compliant to Islamic financial ethical standards based on AAOIFI [13] screening methodology. They offer features such as Halal stock screening to automate screening of thousands of global stocks, ETFs and mutual funds for *Shari'ah* compliance excluding companies involved in prohibited activities e.g., alcohol, gambling, interest-based finance and those with excessive debt. Similarly, it also includes Portfolio tracking and alerts for users to sync their brokerage accounts to track their holdings and receive alerts if any assets fall out of compliance, Zakat calculation and purification tools to automatically calculate Zakat due on investments and provide mechanisms for purifying non-halal earnings and Market insights which provides expert recommendations and alternative halal stock suggestions. These apps empower Muslim investors to make informed decisions aligned with their faith, leveraging AI and data analytics to simplify complex Shari'ah compliance processes.

3) Islamic Knowledge Platforms

The Islam360[14] positions itself as a comprehensive digital encyclopedia and search engine for Islamic knowledge. With around millions of downloads, it provides users with a one-stop solution for accessing the Holy Qur'an with multiple translations and Tafseer (exegesis), a vast collection of authentic Hadiths, prayer times, Qibla direction, daily duas and more. The app centralizes religious texts and resources making it incredibly accessible for Muslims worldwide to learn, research and practice their faith in the digital age. Its success lies in digitizing and organizing extensive Islamic knowledge for quick and easy retrieval. Similarly, the Tarteel[15] is an innovative AI-powered app designed to help Muslims memorize, recite and interact with the Qur'an. It uses advanced speech recognition technology to listen to user's recitations and provide real-time feedback, highlighting missed or incorrect words. The key features it includes the real-time error detection by notifying users instantly on the mistakes in Qur'an recitation, Memorization mode to hides unrecited words to aid memorization, Voice search which allows users to search for verses by reciting them and adaptive mode to customize Qur'an display for easier tracking, understanding and memorization with options for text size, layout and translations. The Tarteel app showcase on how AI can be directly applied to facilitate religious practice and learning, making Qur'anic study more interactive and effective

4) Islamic Lifestyle Applications

The Muslim Pro [16] is one of the most popular Islamic lifestyle apps offering a wide range of features including accurate prayer times, Azan notifications, the complete Holy Qur'an with audio recitations and translations, a Qibla finder, daily duas and a mosque/Halal food finder. Its success lies in being a comprehensive digital companion for daily Muslim life. Similarly, the Qalbox [17] is an integral streaming service within the Muslim Pro app which offers a Muslim-friendly video-on-demand content. It provides a curated library of films, TV series, documentaries and kid's content which aligns with Islamic values and celebrates Muslim identities and cultures. It has addressed the need for comprehensive and *Shari'ah*-compliant entertainment to ensure contents compliant with Islamic ethical standards regarding modesty, themes and messaging into digital entertainment. Moreover, Hidayah[18] is an all-in-one Islamic application designed to be a daily Muslim companion which integrates many features to aid spiritual practice and learning. Its success lies in providing a clean, user-friendly interface which offers complete Qur'an with multiple translations e.g., Urdu,

[10] BIMB (www.bankislam.com)
[11] Zoya (www.zoya.finance)
[12] Musaffa (www.musaffa.com)
[13] AAOIFI (www.aaoifi.com)
[14] Islam360 (www.theislam360.com)
[15] Tarteel (www.tarteel.ai)
[16] Muslim Pro (www.muslimpro.com)
[17] Qalbox (www.app.muslimpro.com)
[18] Hidayah (www.hidayahapp.com)

English and audio recitations by renowned *Qari's*, allowing offline access; Prayer times and *Qibla* finder for accurate prayer timings based on location and an offline Qibla compass; *Istiqamah* Tracker which is a unique feature to monitor daily *ibadah* (worship) progress and maintain consistency in spiritual routines covering obligatory prayers and sunnah acts; AI Islamic Assistant which an AI-powered chatbot that provides instant guidance on Qur'an, Hadith, prayers, Ramadan and other Islamic queries; *Dua* and *Azkar* collection for authentic supplications for various occasions and social media and content sharing offering a halal social media space where Muslims can connect, share posts and engage in discussions within a religiously permissible environment free from haram content. It exemplifies the Islamization of technology by centralizing essential Islamic tools and knowledge, leveraging AI for interactive learning and creating a digital space conducive to spiritual growth.

IV. Discussions

A. *Qur'anic Guidance for ICT Integration*

The Holy Qur'an emphasizes the importance of truth and knowledge, encouraging the duty of individuals and institutions to promote accurate and beneficial information. According to The Holy Qur'an which says, *"and do not confound truth with falsehood, and do not hide the truth when you know (it)"*[19] This verse emphasizes the importance of transparency and truthfulness in disseminating information. This encourages to develop digital platforms which promote Islamic wisdom, credible research and ethical knowledge distribution. Similarly, technologies including, Islamic e-learning platforms, AI-enhanced Qur'anic studies and authenticated Islamic material repositories can guarantee that users obtain precise and valuable information which mitigates disinformation and fostering intellectual and spiritual development.

A major principle taken from The Holy Qur'an is ethical communication which is essential in the digital age. The Holy Qur'an says, *"O you who believe, fear Allah, and speak in straightforward words."* [20] This directive emphasizes the necessity of honesty, respect and equity in all modes of communication especially on social media and digital platforms. To follow this recommendation, ICT systems should have rules that control inappropriate content, help find false information using AI and promote positive discussions. Promoting digital etiquette grounded on Islamic principles can alleviate problems such as cyberbullying, online harassment and misinformation, ensuring that ICT platforms cultivate constructive and significant dialogue.

The Holy Qur'an offers counsel on privacy and security, which are essential issues in contemporary ICT. The Holy Qur'an says, *"O you who believe, abstain from many of the suspicions. Some suspicions are sins. And do not be curious (to find out faults of others), and do not backbite one another. Does one of you like that he eats the flesh of his dead brother? You would abhor it. And fear Allah. Surely Allah is Most-Relenting, Very-Merciful."*[21] emphasizing the significance of honouring personal privacy and protecting information. This principle is directly relevant to ICT, highlighting the necessity for secure data storage, ethical surveillance techniques and user protection mechanisms. Developers need to make sure that privacy-focused policies are included in digital tools. This means using strong encryption, monitoring AI responsibly, and ensuring safe online banking. Complying with these standards not only maintains Islamic ethics but also fosters trust in technology by safeguarding user information from unwanted access and exploitation.

B. *Guidance from Sunnah for ICT Integration*

The Sunnah of Prophet Muhammad ﷺ offers further insights on ethical ICT practices especially about moderation, responsibility and purposeful innovation. Allah's Messenger ﷺ said, *"'Be moderate and adhere to moderation, for there is no one among you who will be saved by his deeds.' They said: 'Not even you, O Messenger of Allah?' He said: 'Not even me. Unless Allah encompasses me with mercy and grace from Him'"* [22] which emphasizes the need to have balance in life and avoid excessive dependence on technology. Although ICT provides various advantages, excessive screen time and unregulated digital interaction might divert attention from spiritual and social obligations. Therefore, ICT should be designed to support helpful digital habits. This includes features to manage screen time, reminders for prayer and reading The Holy Qur'an, and AI tools that promote healthy use of technology.

The Allah's Messenger ﷺ emphasized the importance of responsibility and accountability in every facet of life, it is stated, *"Surely! Everyone of you is a guardian and is responsible for his charges: The Imam (ruler) of the people is a guardian and is responsible for his subjects; a man is the guardian of his family (household) and is responsible for his subjects; a woman is the guardian of her husband's home and of his children and is responsible for them; and the slave of a man is a guardian of his master's property and is responsible for it. Surely, everyone of you is a guardian and responsible for his charges."* [23]. This emphasizes the ethical obligation of

[19] Surah Al-Baqarah (2: 42)
[20] Surah Al-Ahzab (33: 70)
[21] Surah Al-Hujurat (49: 12)
[22] Sunan Ibn Majah, Vol. 5, Book 37, Hadith 4201
[23] Sahih al-Bukhari, Vol. 9, Book 89, Hadith 252

technology developers, content creators and users to ensure that ICT tools are utilized responsibly and morally. Developers must stress Islamic ethical principles in application design while consumers should interact with digital content judiciously keeping away from harmful or false information. Regulations, including Islamic digital ethics policies and AI-based compliance monitoring can ensure that ICT adheres to Islamic moral values. Also, the purposeful invention becomes a fundamental Islamic concept, as Allah's Messenger ﷺ said, *"A believer is someone who loves and is loved. There is no goodness in one who does not love and is not loved. And the best of people are those who are most beneficial to others."* [24] . This emphasizes the significance of developing technology that benefits humanity. This means making Islamic apps to benefit humanity in any form in line with *Shari'ah* such as Islamic financial apps, AI tools for Zakat and charity, online mental health support based on Islamic values, ethical e-commerce etc. to focus on social benefits. The Islamization of ICT can make sure that technology helps bring about positive change, promoting fairness, ethical growth and lasting development in Muslim communities.

V. Challenges and Recommendations

A. *Challenges*

1) Secular dominance in technological advancement

A primary challenge in the Islamization of ICT is the pervasive secular dominance in technological advancement. Numerous technological breakthroughs arise from frameworks which lacks the integration of Islamic values, leading to ethical dilemmas which contradicts Islamic ethical guidelines. This results into different challenges including data privacy concerns by numerous digital platforms which consume financial gains at the expense of user privacy particularly disregarding Islamic ethical guidelines for the protection of personal information and ethical data utilization; the content moderation systems and algorithm which endorses materialistic values, immoral conduct and improper content which contravenes Islamic ethical principles and the technologies such as artificial intelligence (AI), big data analytics and social media which misuses and disseminate misinformation to promote exploitation or discrimination contrary to Islamic principles of justice and ethical behavior. Therefore, to have effective Islamization of ICT, it is essential to contest the dominant secular paradigm by adopting Islamic ethical philosophy into digital platforms and daily practice to develop a newer technology paradigm.

2) Lack of awareness and insufficient resources

Another notable difficulty is the insufficient awareness and limited resources allocated to the advancement of Islamization of ICT. While ethical aspects in technology are increasingly recognized worldwide, the integration of Islamic ethics into technical solutions is yet at its early stages. The primary challenges consist of insufficient Islamic knowledge in technology where numerous technology experts are unacquainted with Islamic principles while many Islamic scholars possess limited technical proficiency. This results in a disconnect in creating ethical oriented technological solutions. Similarly, the inadequate investment and limited financial and institutional backing for Islamization of ICT initiatives possess a significant challenge. The dedicated financial support is required and essential to facilitate research, innovation and the advancement of Islamic digital solution. Also, a limited number of academic programs or training efforts integrated with Islamic ethics with technical studies leading resolve to a knowledge gap obstructs the successful Islamization of ICT. There is a need to resolve these difficulties necessitates focused initiatives to connect Islamic ethics with technological progress.

B. *Recommendation*

1) Developing cooperation between Islamic scholars and technology experts

There is a need to address the difficulties to promote collaboration between Islamic scholars and technology experts. This interdisciplinary approach ensures that technical progress should be consistent with Islamic ethics. These potential efficient methodologies include collaborative research initiatives to form academic partnerships between Islamic education institutes and technology specialists to develop ethical digital tools and frameworks which comply with Islamic ethics. Similarly, conducting workshops and seminars enables cooperation through educational events to unite scholars and technology experts to have understanding and consensus on Islamization of ICT. It can address subjects such as AI ethics, data privacy in Islam and the development of Halal technology. The advancement into interdisciplinary education to develop academic programs to integrate Islamic studies with ICT training to educate future professionals with both Islamic and technical proficiency simultaneously. The facilitation of these partnerships can render the integration of Islamic ethics into technological development in a more systematic and effective way to bring positive prospects to many users.

[24] Al Mu'jamul Awsat, Hadith: 5783, Shu'abul Iman, Hadith: 7252

2) Promote awareness and establish Islamic tech hubs

To facilitate extensive acceptance and endorsement of Islamization, better awareness and dedicated resources are required. This can be achieved by public awareness campaigns initiatives to promote campaigns which emphasizes on the importance of Islamization of technology and its advantages for Muslim communities. Similarly, the community engagement and collaboration efforts with Masjid, Islamic groups and educational institutions to advance the development and implementation of ethical digital solutions. Also, to showcase success stories which emphasize successful Islamization of ICT efforts to motivate developers and promote extensive adoption among users. Furthermore, the establishment of Islamic technology centers can furnish essential infrastructure for research, development and innovation in Islamization of ICT. These centers can contribute to the following potential objectives including, the allocation of financial support and resources to extend monetary assistance, guidance and technical resources to innovators developing Islamic technology solution, to promote ethical innovation for advocating the development of digital tools compliant with Islamic ethics to tackle modern technical difficulties while preserving Islamic integrity and to establish strategic alliances and partnership with prominent technology companies, academic institutions and Islamic groups to expand successful initiatives and improve access to Islamically-aligned ICT solutions.

VI. CONCLUSION

The integration of Islamic ethical principles into ICT offers a considerable opportunity as well as a challenge. Islamization of technology is crucial to ensure that digital tools and platforms conform to Islamic ethics, values and has a good impact on the society. The Holy Qur'an and Sunnah of Prophet Muhammad ﷺ offers explicit guidance on the clear pursuit of knowledge and the enhancement of societal welfare e.g., principles which can be applied to ICT. This study has analyzed the primary difficulties in the Islamization of technology, notably the prevalence of secular paradigms in technical advancement and a lack of resources for Islamic ICT efforts. Moreover, it emphasizes on the necessity of promoting collaboration between Islamic academics and technology specialists, to improve awareness among the Muslim community regarding Islamization of ICT and creating specialized Islamic technology centers. The proposed ethical framework introduces an implementation strategy which can align with the technology landscape within the purpose of *Maqasid al-Shari'ah* which is a paramount goal of Islamic law aimed at safeguarding faith, life, intellect, lineage, and wealth.

ACKNOWLEDGMENT

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

There is no external or third-party data that support the findings of this study.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] F. Rahman, "Islamization of knowledge: A response," *American Journal of Islam and Society*, vol. 5, no. 1, pp. 3-11, 1988.

[2] H. Dzilo, "The concept of 'Islamization of knowledge'and its philosophical implications," *Islam and Christian–Muslim Relations*, vol. 23, no. 3, pp. 247-256, 2012.

[3] M. A. Ahsan, A. K. M. Shahed and A. Ahmad, "Islamization of knowledge: An agenda for Muslim intellectuals," *Global Journal of Management and Business Research Administration and Management*, vol. 13, no. 10, pp. 33-42, 2013.

[4] A. Chande, "Global politics of knowledge production: The challenges of Islamization of knowledge in the light of tradition vs secular modernity debate," *Nazhruna: Jurnal Pendidikan Islam*, vol. 6, no. 2, pp. 271-289, 2023.

[5] M. M. Hossain, L. H. Abdullah, M. T. Hoque and M. I. B. Roslee, "The Methodology of Islamization of Knowledge: A Conceptual Study," *International Journal of Islamic Business & Management*, vol. 6, no. 1, pp. 9-18, 2022.

[6] U. W. Nisa, "Islamization of Knowledge and Its Challenge," in *Proceeding of International Conference on Education, Society and Humanity*, 2023.

[7] S. M. Miri and A. Q. Bagestan, "What Is Islamization of Knowledge: A Review on the Prominent Muslims Scholar's Thought," *12-Studies Religion Muslems*, vol. 5, no. 9, pp. 39-59, 2019.

[8] S. Hanafi, "Islamization of Knowledge and It's Grounding: Appraisal and Alternative," *Islamic Studies Review*, vol. 1, no. 2, pp. 135-160, 2022.

[9] M. Laabdi and A. Elbittioui, "From Aslamat al-Maᶜrifa to al-Takāmul al-Maᶜrifī: A Study of the Shift from Islamization to Integration of Knowledge," *Religions*, p. 342, 2024.

[10] S. Yaacob and I. Haron, "Integration of Islamic Values Within Science Based on Islamization of Human Knowledge (IOHK) Theory or Philosophy," *International Journal of Religion*, pp. 59-66, 2024.

[11] S. Sawaluddin, K. S. Haraha, I. Rido and I. A. Supriono, "The Islamization of Science and Its Consequences: An Examination of Ismail Raji Al-Faruqi's Ideas," *Jurnal Pendidikan Agama Islam (Journal of Islamic Education Studies)*, pp. 115-128, 2022.

[12] M. Muslih, H. Susanto and M. P. Perdana, "The Paradigm of Islamization of Knowledge According to SMN Al-Attas (From Islamization of Science to Islamic Science)," *TASFIYAH Jurnal Pemikiran Islam,* p. 25, 2021.

[13] F. Eayaz and M. Riaz, "Bridging Faith and Knowledge: Hamidullah's Role in Philosophical and Comparative Approaches to Islamization," *Al-Marjān,* pp. 18-30, 2024 .

[14] S. O. Masood, "Islamization of Knowledge and the Educational Philosophy of Al-Attas–A Brief Exposition," *Al-Kashaf,* pp. 45-52, 2024.

[15] K. Umam and N. Jannah, "Intersection Of Artificial Intelligence and Islamic Studies: Challenges and Opportunities in The Digital Era," *Peace and Humanity Outlook,* vol. 1, no. 1, pp. 39 - 48, 2024.

[16] R. Mustapha and S. N. A. Malkan, "Maqasid Al-Shariah In The Ai Era: Balancing Innovation And Islamic Ethical Principles," *International Journal of Islamic Theology & Civilization,* vol. 3, no. 3, pp. 1 - 21, 2025.

[17] N. S. A. Azizah and S. Shalihah, "Maqasid Al-Shari'ah and Legal Pluralism: Normative Analysis of The Principle of Justice in A Multicultural Society," *Journal of Islamic and Law Studies,* vol. 9, no. 2, pp. 119 - 126, 2025.

[18] A. Traore, "The Dead Weight That Is Hindering the Islamisation of Knowledge," *Islamic Studies,* vol. 58, no. 2, pp. 205-218, 2019.

[19] I. A. S. a. M. F. Yunita, "The Imperative of Integrating Knowledge and Adab in Reconstructing Islamic Education in the Digital Era: A Study of Al-Attas's Thought," *J-PAI: Jurnal Pendidikan Agama Islam,* vol. 11, no. 2, 2025.

[20] H. Zubaidillah, "Epistemological Views of Islamic Education Philosophy as A Islamic Education Basis," *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan,* pp. 1-12, 2018.

[21] Y. M. Lubis, "Social Change in Contemporary Islamic Community Development through Transformative Da'wah Praxis," *Jurnal Al-Hikmah,* vol. 23, no. 1, pp. 57 - 76, 2025.

[22] A. b. H. Ali, "Scholarly consensus: Ijma ': between use and misuse," *Journal of Islamic Law and Culture,* vol. 12, no. 2, pp. 92 - 113, 2010.

[23] A. M. J. Azmi and F. Y. Avicena, "The Study of Qiyas as a Legal Argument: Application and Limitations," *Ethica: International Journal of Humanities and Social Science Studies,* vol. 2, no. 3, 2024.

[24] Z. Zainuddin, M. Muttaqin, B. Amir, A. Nafisah and P. Paizaluddin, "Epistemological Synthesis of Al-Attas and Al-Faruqi: Islamization of Knowledge, Adab, and Contemporary Decolonization of Knowledge," *ISEDU: Islamic Education Journal,* vol. 3, no. 1, pp. 18 - 31, 2025.

[25] D. D. Suriyani and R. D. Almanda, "Encouraging Intellectual And Spiritual Progress Through The Islamization Of Knowledge," *Jurnal Intelek Dan Cendikiawan Nusantara,* vol. 2, no. 1, pp. 280 - 288, 2025.

[26] E. Harunoğullari, "Harunoğullari, E. (2025). An Analysis of Disruptive Technologies in Muslim Societies: Economic, Financial, and Ethical Implications," *Disruptive Technologies And Muslim Societies: From Ai And Education To Food And Fintech,* pp. 389 - 416, 2025.

[27] M. A. Iqbal and N. Bakare, "Muslim Societies and the Rise of Artificial Intelligence: Impacts, Challenges, and the Way Forward," *Disruptive Technologies And Muslim Societies: From Ai And Education To Food And Fintech,* pp. 27 - 46, 2025.

[28] A. Nadeem, N. S. Rakhshani, K. Aslam and M. I. Khan, "Leveraging Historical Responses and Islamic Perspectives for AI in 21st-Century Healthcare in Muslim Societies," *Disruptive Technologies And Muslim Societies: From Ai And Education To Food And Fintech,* pp. 47 - 73, 2025.

[29] U. Ali, K. Hussain and R. Sheikh, "Performance Analysis of Islamic Banking in Pakistan Using DEA Technical Efficiency and Maqasid al-Shari'ah Index," *Journal of Islamic Business and Management,* vol. 13, no. 1, pp. 56 - 70, 2023.

[30] "Uncover Islami Bank Bangladesh Legal Overview and Guidelines," [Online]. Available: https://www.lawyersnjurists.com/article/overview-islami-bank-bangladesh/. [Accessed 25 07 2025].

[31] N. E. Fauziyyah and E. F. b. Mohamed, "Public relation activities in Malaysian Islamic Banking: The case of Bank Malaysia Berhad (BIMB)," *Airlangga International Journal of Islamic Economics and finance,* vol. 3, no. 1, pp. 15 - 30, 2020.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.649

SSL/TLS Certificate Validation Tool for Pre-Authentication Captive Portals

Certificate Validation in Captive Portals

Shafana M.S[1*], Adamu Abubakar Ibrahim[2]

[1]Department of ICT, Faculty of Technology, South Eastern University of Sri Lanka, Sri Lanka
[1,2]Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia

*Corresponding author: zainashareef@gmail.com
(Received: 4th November 2025; Accepted: 12th December, 2025; Published on-line: 30th January, 2026)

Abstract— Public network captive portal often disrupts the handshake in the SSL/TLS protocol and displays browser warnings that are sometimes ambiguous and sometimes excessive. These warnings can be misleading for users even for those with technical expertise. Notwithstanding the associated risks, there are limited tools that can be used to validate the trust of the SSL/TLS certificates in pre-authentication environments. This paper presents a lightweight and automated tool that is designed to validate the contents of an extracted certificate chain of SSL/TLS from live or stored handshake traffic at captive portals. The tool uses the trust evaluation engine of OpenSSL, supplemented with a Mozilla-compatible CA bundle to determine the validity of certificates and enables automatic retrieval of the missing intermediate certificates through AIA URLs to improve the accuracy of validation. The tool was evaluated using TLS handshakes captured from the IIUM Wi-Fi captive portal and samples from the ISCXVPN2016 dataset. After intermediate correction, the tool achieved 100% detection accuracy with no false positives and false negatives. It was able to detect misconfigured, expired or incomplete chains and validate known secure sessions. The proposed solution had more accurate and actionable diagnostics compared to browser-based indicators and tools like SSL Labs in pre-login situations, where existing methods often fall short. This tool fills an important gap in network security for users, by enhancing transparency and trust in certificate assessment. Its automation and diagnostic clarity make it an effective tool for both researchers and general users and provide reliable SSL/TLS validation in environments where conventional trust signals are unavailable or misleading.

Keywords— Captive Portal, Certificate Chain Verification, Intermediate Certificate Recovery, OpenSSL, SSL/TLS validation, Wi-Fi Security.

I. Introduction

Captive portals are a commonplace feature of both the public and institutional Wi-Fi providers, essentially acting as access control points that determine who can access the broader Internet. An example of this is the International Islamic University Malaysia (IIUM), where users are frequently presented with a login portal such as captiveportalgombak1.iium.edu.my which is used by all students, staff and guests. Although such portals serve their intended role, they also present substantial cybersecurity risks. One of the main issues is associated with the use of SSL/TLS certificates [5]. These online certificates authenticate a website and encrypt information during transmission. A captive portal should have a valid certificate issued by a recognised Certificate Authority (CA) in order to be secure. The inability to satisfy this need may allow attackers to impersonate legitimate portals, execute man-in-the-middle attacks, or steal personal data unnoticed.

Academic studies and practical observations have shown that a great number of portals fail to implement SSL/TLS correctly. Misconfigured certificate chains and poorly designed warning pages are common. This means that users receive browser warnings that they fail to understand and usually ignore. The so-called warning fatigue is also a significant factor, after being shown the same generic warnings many times ("Your connection is not private") users will disregard them completely.

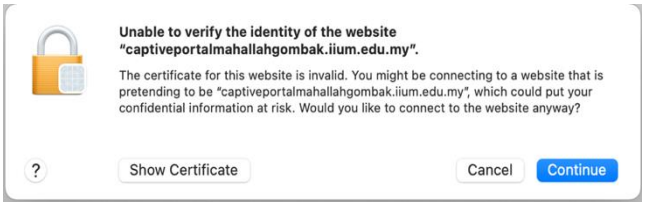


Fig 1. Browser Warning on Captive portal at IIUM

https://doi.org/10.31436/ijpcc.v12i1.649

For non-experts, the concepts of certificate hierarchies and expiration dates may be confusing and intimidating. Research shows that such confusion significantly increases the chances of users being victims of phishing or session hijacking as well as data leakage, even in usage of public WiFi also privacy risks arise [2].

The urgency of this issue was pointed out by a real incident at IIUM. Users who tried to use the Wi-Fi network of the university received alarming browser alerts (see Fig 1) stating that the identity of the portal could not be verified. But after further examination with the help of OpenSSL, it was established that the certificate of the portal was genuine, properly issued to the domain iium.edu.my by GlobalSign, and valid in terms of expiration. The issue was traced to a missing intermediate certificate, which made the chain of trust incomplete. This was a technically small gap, but it caused browsers to reject the certificate, and users had to determine whether to continue or not, usually without understanding the underlying risk.

This situation reveals two fundamental issues: a technical misconfiguration that weakens the portal's trust model, and behavioural weakness where the users are not equipped to make informed security decisions. The cross-section between those failures explains how network security can be compromised by a minor mistake, like the lack of an intermediate certificate that undermines trust. Also, browser warnings do not contextualise well in captive portal settings; they fail to inform users about the specific nature of the problem or whether it might be safely ignored. Users are therefore forced to make a trade-off between usability and perceived security without the necessary insight to do so effectively.

The rationale behind the study is that there is a high level of dependency on wireless access systems in institutions of learning, and that more intelligent, context-aware systems that can facilitate secure connectivity are required. Misconfigurations in the implementation of the SSL/TLS systems are preventable but still common and represent a continuing risk in the environment where users have to perform authentication over the untrusted or partially trusted connections. In this regard, this paper aims to create and deploy a lightweight, automated inspection system that can use Wireshark packet captures to identify and validate the SSL/TLS certificate chains and, finally, to advise the user whether a certificate warning is genuinely critical or safely ignorable.

This study will help to address infrastructure-level vulnerabilities and end-user vulnerabilities, and introduce a transparent validation process that will help enhance certificate-based security in captive settings and lead to more trustworthy public internet access experiences

II. LITERATURE REVIEW

A. *Captive Portals and SSL/TLS Certificate Challenges:*

Captive portals are commonly used in open and institutional Wi-Fi networks to control user access before authentication. These mechanisms tend to disrupt HTTPS connections, which may result in security misconfigurations despite their usefulness. Recent studies have shown that captive portal mini-browsers of popular operating systems often disable the validation of the SSL/TLS certificate and, as a result, expose users to man-in-the-middle (MITM) attacks and impersonation risks [4]. Authors in [7] also notes that most captive portals use HTTP redirection during the user authentication process, which may mislead users and give attackers the chance to deliver malicious content by using a cloned SSID.

One of the recurrent problems is the incorrect setup of TLS certificates. Improper installation of intermediary certificates often causes trust chain failures, which is treated as an untrusted connection by browsers regardless of the validity of root and leaf certificates. These misconfigurations provide a situation where seemingly secure portals are insecure, which undermines user trust towards the portal and the security of the entire network.

B. *User Behaviour in Response to SSL/TLS Warnings:*

User interaction with browser-based warnings on the use of the SSL/TLS is an enduring issue in web security. The study [9] reported that users have been found to ignore such warnings either out of habit or in an attempt to act quickly without regard to the security concerns. Authors in [1] observed that although the design of warnings had been improved marginally, most users still made unsafe choices. Authors of [6] applied these results to real-world scenarios and pointed out that users are more likely to over trust networks like campus Wi-Fi more than is warranted and ignore severe certificate warnings. These patterns in behaviour highlight the insufficiency of browser warnings as a defence mechanism, especially when they appear in captive portals where they are nonspecific and the cognitive load is entirely placed on the user. It follows, therefore, that there is an urgent need to have tools that provide actionable and context-based feedback that can be used to guide user decisions effectively.

C. *SSL/TLS Certificate Analysis Tools:*

Wireshark is the standard of deep packet inspection in the industry, providing all the analysis of the SSL/TLS including the handshake parsing, certificate decoding, and aspects of TLS 1.3 compatibility [11]. However, its technical depth is often impractical for non-technical users. The approach [10] is a simple and server-based certificate scanner that lacks support for passive or client-side scanning in a Wi-Fi setting.

https://doi.org/10.31436/ijpcc.v12i1.649

There are no existing tools specifically tailored for captive portal environments with packet-level analysis and automated certificate validation and user advisories. This gap highlights the need to have a user-friendly, lightweight solution that operates in both pre- and post-authentication stages.

D. Attack Models and Real-World Attacks in Captive Portals

Real-world attack scenarios highlight how vulnerable captive portals can be. The study [7] divided the common threats, among which are Evil Twin attacks, where the adversaries pretend to be legitimate networks and offer a fake portal to steal credentials [3]. They tend to work well because of improperly configured certificates and users' false sense of security on familiar networks. The other attack vectors include MITM interception during HTTP redirection and session hijacking via stolen cookies [8]. Such threats are often not identified due to the fact that users post sensitive data to portals that are marked as untrusted, but without understanding the implications. Critical insights in the key studies are summarised in Table 1 below.

TABLE 1: KEY STUDIES ON CAPTIVE PORTALS AND SSL/TLS WARNINGS

Study	*Focus*	*Key Findings*	*Limitations*
[4]	Mini-browser TLS validation	Captive portals often bypass SSL validation	Lacks user-side countermeasures
[9]	User behavior to warnings	Users frequently ignore SSL alerts	Based on outdated UI models
[1]	Warning effectiveness	Design helps but does not eliminate risky behavior	No contextual guidance
[6]	Behavior on trusted networks	Users ignore warnings in familiar networks	Overlooks captive portal uniqueness
[11]	TLS handshake analysis	Robust technical tool	Not accessible to average users
[10]	Server-side TLS analysis	Automates validation checks	Not suitable for Wi-Fi clients
[7]	Captive portal threat modeling	Details attacks like Evil Twin and session hijacking	Offers no integrated defense tools

E. SSL/TLS Authentication on Browser Platforms

In the present context, web browsers serve functions beyond rendering webpages. Beneath the surface, they perform systematic validation of SSL and TLS certificates to ensure that the credentials which are offered by the websites or the network portals are legitimate and trustworthy. Although the previous standard was the use of the SSL, it has been phased out. SSL is replaced by TLS, which has acquired the most important role of ensuring internet communication through encryption and verification of server identity. Such validation steps are not minor details, they are the key elements of the web security model, which protects against impersonation, MITM attacks, and phishing.

A browser does a number of significant checks when a user visits a resource with HTTPS protection, such as a captive portal:

- Certificate Chain Validation: The browser guarantees a full trust path between the leaf certificate and a trusted root Certificate Authority (CA).
- Expiration Check: The browser verifies that the certificate is currently valid.
- Domain Name Matching: The domain name must match the website the user is visiting.
- Signature Verification: Using a known and trusted CA, the digital signature has to be verifiable.
- Trusted Root CA Check: The root certificate should be located in the local trust store of the browser.
- Revocation Status Check: The browser can interrogate Certificate Revocation Lists (CRLs) or Online Certificate Status Protocol (OCSP).
- Extension and Key Usage Evaluation: The certificate must include required X.509 extensions, such as Key Usage and Extended Key Usage, which define its role in secure communication.

Although the theoretical validation model used by browsers is complete, in practice it is not always consistent, particularly when using captive portals. To provide an example, a browser may not necessarily retrieve or seek out missing intermediate certificates in the case external AIA URLs are inaccessible before portal authentication. This is a problem when the unauthenticated traffic is blocked by DNS interception or firewalls. Moreover, they have revocation checks by CRL or OCSP that are frequently soft-failed due to performance reasons, meaning that the browser may continue without checking whether a certificate has been revoked. These design decisions put usability above security and compromise the validity of trust decisions.

The result of such complex validations is usually summarised to the user as some sort of generic browser warning stating that their connection is not private, but does not specify whether the issue is an expired certificate, broken chain or an untrusted root. This lack of diagnostic clarity limits the user's ability to make informed decisions and may cause unnecessary concern or unsafe acceptance of risk.

F. Limitations of Browser Behaviour

Unlike browser-based validation, the presented system offers deterministic analysis of certificates, by clearly stating the reasons why trust failed, whether due to the lack of intermediates, improper formatting or cryptographic anomalies. This clarity is especially critical in captive network scenarios in which browser interfaces provide insufficient feedback and no remedial options. The tool is able to bridge

https://doi.org/10.31436/ijpcc.v12i1.649

the transparency and reliability gap that is still present in modern browsers by incorporating fullchain extraction, offline CA bundle validation, and optional AIA recovery mechanisms.

III. Methodology

A. System Overview

The proposed system is a lightweight and unobtrusive utility to support non-technical users and researchers in assessing the trustworthiness of SSL/TLS certificates during WiFi login, in captive portal specific cases. The utility automatically strips out and verifies certificates exchanged during the SSL/TLS handshake by exploiting passive network capture techniques, hence fills a significant usability gap in the public network security. The architecture is structured as follows into five main stages:

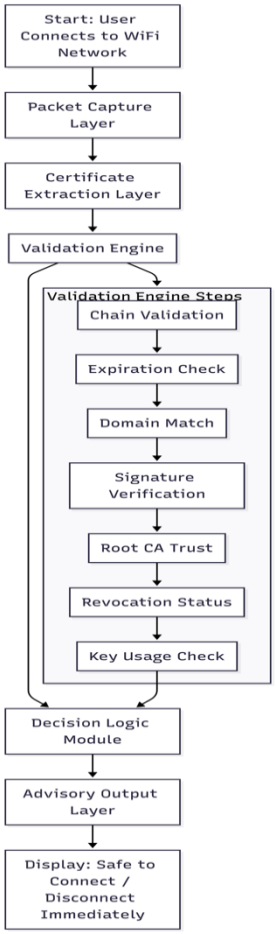


Fig 2. System Architecture Overview.

- Passive Capture: The utility surveys TLS handshakes with Wireshark or Pyshark when authenticating with a captive portal by paying attention to ServerHello messages in which certificates are exchanged.
- Certificate Chain examination and validation: It parses the handshake to extract leaf and intermediate certificates, verifying them with OpenSSL and an up-to-date CA bundle.
- Context-Aware Inspection: The domain name matching, expiration validity, signature verification and presence of intermediate are considered in sequence.
- The system determines whether warnings are fatal (e.g., broken chain) or explainable (e.g., lacking intermediate that can be obtained through AIA).
- User Advisory: A clear and actionable message is provided—such as "DISCONNECT / UNTRUSTED" or "SAFE TO CONNECT."

The tool replicates the logical certificate verification flow of modern browsers but enhances it with automated diagnostics and transparent error reporting (see Fig 2). The validation logic is sequential and fails fast:

- Certificate Chain Incomplete → "Broken Chain"
- Expired Certificate → "Expired Certificate"
- Domain Mismatch → "Domain Mismatch"
- Invalid Signature → "Untrusted Signature"
- Missing Trusted Root → "Untrusted CA"
- Revoked Certificate → "Revoked"
- Invalid Key Usage → "Usage Not Allowed"

If all checks pass, the user is informed that the connection is secure.

B. Implementation Strategy

The implementation of the tool is conducted in several stages. First, TShark was used to capture .pcap and .pcapng files when capturing captive portal login sessions. Out of these captures, all certificates were retrieved using a custom Python script by using the ssl.handshake.certificate field. The individual certificate blocks were decoded and saved in PEM format, then they were combined into full certificate chains.

The basic validation tool is the use of OpenSSL's verify command. In the case of the incomplete chain, the tool tries to obtain the missing intermediates via the AIA URLs in the certificates. If AIA retrieval fails, the .crt file is manually converted to .pem format with the DER-to-PEM conversion command of OpenSSL.

Every resulting PEM block is verified, and invalid entries are not included in the final CA bundle. The cleaned set is reassembled into the local_ca_bundle.pem, which is the trusted CA store that is used by the tool. After successful incorporation of both automated and manual recovery mechanisms, the tool was tested with real-world IIUM portal traffic and it was demonstrated to correctly diagnose trust states even in complex TLS misconfiguration cases. The clear advisories are shown in the terminal or optionally exported as CSV for further analysis.

https://doi.org/10.31436/ijpcc.v12i1.649

C. *Experimental Setup*

The data used in the research came from two primary types of datasets. The former included open-source datasets, including the ISCXVPN2016 dataset, which is a collection of SSL/TLS handshakes from benign and encrypted traffic scenarios. This data set was suitable for testing passive extraction and parsing logic. The second source included custom captive portal captures. Real-world TLS handshakes were recorded from IIUM and public WiFi portals by using Pyshark, focusing specifically on ServerHello and certificate messages to ensure user privacy. Each of the datasets was preprocessed to identify the relevant SSL/TLS traffic. The fields of certificates like domain names, expiration dates, issuer signatures and validation paths were normalized for uniform analysis.

A set of metrics was used to determine the effectiveness of the tool. Detection accuracy was the percentage of safe and unsafe sessions that were correctly classified compared to all the tested sessions. The false-positive rate (FPR) was the rate of secure connection being incorrectly identified as unsafe and the false-negative rate (FNR) is the rate of unsafe connection identified as safe. Average execution time was also recorded. These indicators were calculated during both offline dataset analysis and real-time WiFi validation phases. The experimental setup used macOS Ventura 13.6.5. Wireshark v4.2 and Pyshark were the packet capture tools. OpenSSL v3.0 and Certifi libraries were used for validation. Python 3.11 served as the programming language. The experiments were done on MacBooks and smartphones that were connected to captive portals at IIUM and other open areas.

Evaluation was performed on IIUM captive portal traffic and ISCXVPN2016, so generalization across other captive portal deployments and regions remains to be validated. The dataset primarily reflects common misconfiguration-driven failures (expired, self-signed, incomplete chains); additional scenarios such as revocation and OCSP behaviors and TLS 1.3-specific handshake constraints were not observed in the traces and therefore were not benchmarked. Future work will expand cross-site testing and add controlled experiments to cover these additional TLS conditions.

IV. Results and Analysis

D. *Quantitative Results*

The Captive Portal Certificate Validation Tool proposed was evaluated on the basis of a dataset containing more than 22,000 SSL/TLS packets, collected from 24 distinct capture files, including AIM chat, Facebook and email sessions. A Python script based on the modified version of TShark was used to extract the certificate chains and validate them against OpenSSL, with the verdicts assigned to risk-based classes as shown in Fig 3.

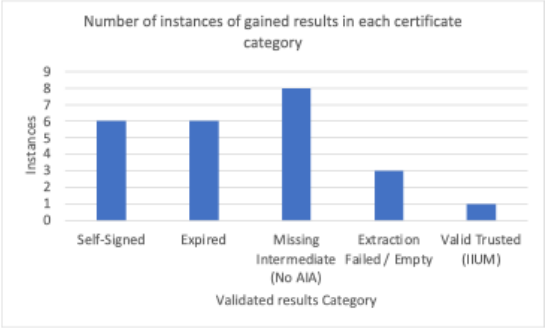


Fig 3. Distribution of Certificate Validation

The bar chart shows the number of PCAPs in each certificate category: “Self-Signed, Expired, Missing Intermediate, Extraction Failures and Valid and Trusted." The most common one is outlined in bold as “Missing Intermediate.”

- Self-Signed Certificates (6 instances): This class ended with a self-signed root that was not available in trusted stores. This means that it is highly vulnerable to man-in-the-middle attacks. Examples of those found in AIM and Facebook traces are AIMchat1, aim_chat_3a and facebook_video1b.
- Certificates that have expired (6 instances): These were structurally valid certificates that had exceeded their expiry dates. These are common in entries like facebook_video2a and email2b and are moderately dangerous because of negligence, or because they were used in outdated test environments.
- Lacking Intermediate CA / No AIA Fallback (8 instances): Such instances include partial certificate chains, which are mostly caused by the lack of intermediate certificates or the inability to retrieve them through AIA. Examples of such files include email1a, messenger1a and facebook_chat_4a, which are the most commonly observed issue.
- Extraction Failures or Missing Data (3 cases): The data contained in TLS certificates were either incomplete, corrupted, or inaccessible because of encryption, as seen in files such as facebookchat1, facebookchat3 and facebook_chat4a. These were classified as Unverifiable Risk.
- Valid and Trusted Chain (1 instance): The captive portal of IIUM was the only one to provide a complete valid certificate chain, which was successfully verified with the help of OpenSSL against a root CA store compatible with Mozilla. Although the browser issued a warning, the tool was able to identify it as trusted, which highlights the advantage of verifying at the packet level.

https://doi.org/10.31436/ijpcc.v12i1.649

Other certificate problems like domain errors, revocation errors, or weak signature algorithms (e.g., SHA-1) were not identified in this dataset, due to the small set of traces considered (academic traces), which focused on expired, self-signed, or incomplete-chain scenarios. The tool had a 100% detection accuracy following the intermediate resolution and zero false positives and zero false negatives in validation. Each validation session took approximately 40 seconds. When tested live, the tool was correct in its opinion that IIUM's captive portal certificate was trusted, despite browsers indicating it as untrusted due to non-cryptographic signals (e.g. captive portal redirection, HSTS preload conflicts). This highlights one advantage of the packet-level analysis of SSL/TLS: the detection of cryptographically secure certificates that browsers may misinterpret.

The reported ~40 s per validation session reflects an end-to-end offline workflow that includes packet parsing, certificate extraction, chain reconstruction, OpenSSL trust evaluation, and (when needed) intermediate retrieval via AIA.

As such, the current implementation is best suited to offline auditing, troubleshooting, and user-side spot-checking during captive portal onboarding, rather than acting as an online, real-time validator in high-throughput public Wi-Fi environments. For large-scale deployments, the workflow can be operationalized by validating only the first observed handshake per portal domain, caching intermediate certificates by AIA URL/issuer identifiers, and providing a "fast mode" that performs deterministic chain checks against the local CA bundle while deferring network-dependent steps. These optimizations reduce redundant work across repeated sessions while preserving the tool's diagnostic function in pre-authentication settings.

To determine ground truth, the classifications of the tool were validated with OpenSSL's verify command against a Mozilla-compatible CA bundle. An output response of certificate: OK indicated a cryptographically valid, complete, and trusted chain. Broken trust paths were supported by errors like unable to get local issuer certificate. Domain names, expiration dates, and issuing authorities were verified by manual inspection using openssl x509 -in cert.pem -noout -text. It was also compared with browser behaviour and SSL Labs reports. Authentic certificates issued by sites like Google and Cloudflare did not give any warning in the browsers and were also marked as safe by the tool. In the case of IIUM portals, browser warnings indicated that they were not trustworthy; nevertheless, the tool was right in that the certificates were valid and anchored by a trusted CA, highlighting the limitations of browser interpretations and reinforcing the reliability of the packet-level method.

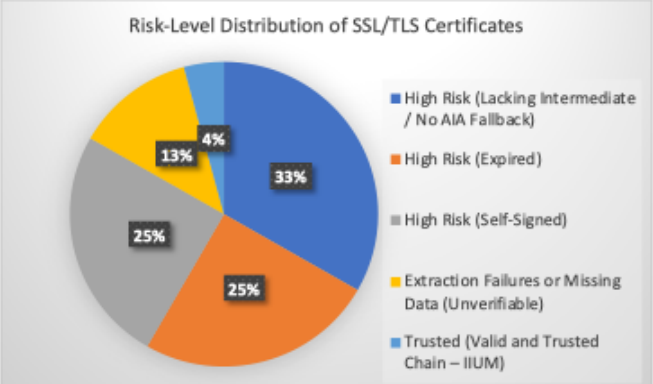


Fig 4. Risk-Level Distribution of SSL/TLS Certificates

The pie chart shows the ratio of certificates according to risk level. The segments are grouped into the classes of high risk (self-signed), medium risk (expired, intermediate missing), unverifiable risk (extraction failed), and low risk (trusted). The results of the validation of the certificates presented by the SSL/TLS were further split into four different risk levels depending on the severity and security implications (Fig 4). Self-signed chains are high-risk certificates, which lack a verifiable certificate authority (CA), and they are extremely vulnerable to man-in-the-middle attacks. Medium-risk includes both expired and missing intermediate authority certificates and is commonly caused by mismanaged certificate lifecycles, as well as network restrictions that prevent AIA retrieval. Unverifiable risk refers to those where there was a failure of certificate extraction or incomplete or corrupted TLS session, making validation infeasible, and are generally due to encrypted payloads, lost packets, or noise. Finally, the single capture of IIUM captive portal certificate that was fully validated against a trusted CA bundle showed a properly implemented and cryptographically sound trust path and was classified as low risk.

E. Comparative Analysis

Conventional web browsers do the SSL/TLS validation internally and generally only give general warnings, like Your connection is not private, without any technical explanation of the failure. The causes of such vague messages can be a broad set of reasons such as expired certificates, the lack of intermediate certificates, or improperly configured root trust anchors, yet the user receives no clarity. Although it can be useful in analysing server configurations, SSL Labs is not compatible with validation of client-side traffic that has been observed before a captive portal login, nor can it be used to analyse a certificate chain as observed in pre-authentication states. Conversely, the tool mentioned below has several functional and technical benefits. It runs

https://doi.org/10.31436/ijpcc.v12i1.649

offline, validates certificate chains based on raw network captures, automatically reconstructs missing intermediates through AIA fetching and outputs results in CSV format for further analysis. Table 2 provides a summary of the comparative capabilities of major tools.

TABLE 2. COMPARATIVE FUNCTIONALITY ACROSS VALIDATION TOOLS

Feature	*Web Browsers*	*SSL Labs*	*Proposed Tool (This Study)*
Support for Pre-Login Captive Portals	Not Supported	Not Supported	Fully Supported
Explanation of Intermediate Chain Issues	Not Provided	Partially Provided	Fully Provided
Automatic Recovery of Missing Certificates	Not Supported	Not Supported	Fully Supported
Command-Line Interface (CLI) Usability	Not Supported	Not Supported	Fully Supported
Logging and Result Export Capability	Not Available	Not Available	Fully Available (CSV & terminal log support)

As illustrated, the proposed tool is distinctively positioned to support use cases where browser-based or server-centric solutions are inapplicable

In addition to browsers and SSL Labs, a number of practitioner and research tools exist for TLS inspection. However, most popular scanners are server-facing (active probing) and assume direct Internet reachability, which is often unavailable prior to captive-portal authentication. In contrast, the proposed tool operates on passively observed handshakes captured at the client side and produces a trust decision from the extracted chain using an offline CA bundle and OpenSSL verification.

Tools such as Wireshark provide deep protocol dissection but do not provide an automated trust verdict oriented toward end users, and similar monitoring-oriented analyzers do not target pre-auth captive-portal decision support. Accordingly, we report this comparison as capability-based rather than a performance benchmark

V. DISCUSSION

Live testing on IIUM's Wi-Fi captive portal revealed that there were strong differences between warnings given by browsers and the actual certificate trustworthiness. For example, Safari on macOS showed a high severity warning for a certificate for '*captiveportalmahallahgombak.iium.edu.my*', despite the fact that the certificate was signed by the globally trusted authority GlobalSign RSA OV SSL CA 2018 and had a valid signature and proper domain parameters. This behaviour illustrates that browser, despite various checks that they do in the background, are unable to present meaningful diagnostic information to the users. Even technically literate users who are not computer security experts may struggle to determine the validity of such warnings or to assess whether they can be safely disregarded. The tool addresses this weakness by giving accurate validation messages that capture the cause of failure. The system also provides reliable evaluation with the support of deterministic assessment of trust of the OpenSSL system, automatic recovery of intermediate certificates, and the validation of PEM integrity. It has been designed to operate without graphical user interfaces but rather uses Python, OpenSSL and TShark, thus being accessible to non-expert users and researchers. Table 3 provides a comprehensive comparison of the validation properties that are supported by different tools.

TABLE 3. COVERAGE OF SSL/TLS VALIDATION PROPERTIES

Validation Property	***Web Browsers***	***SSL Labs***	***Proposed Tool (This Study)***
Leaf Certificate Validity	Supported	Supported	Supported
Intermediate Certificate Presence & Integrity	Partially Supported	Fully Supported	Fully Supported
Domain Name Matching (CN/SAN)	Supported	Supported	Supported
AIA-Based Intermediate Certificate Retrieval	Not Supported	Not Supported	Fully Supported
Offline CA Bundle Trust Verification	Not Supported	Not Supported	Fully Supported
CLI Usability and Automation Support	Not Supported	Not Supported	Fully Supported
PEM Format Structure & Consistency Checking	Not Supported	Not Supported	Fully Supported
Operation in Captive Portal Contexts	Not Supported	Not Supported	Fully Supported

The results supporting the unique ability of the tool to address weaknesses of current solutions can be seen in the context of an environment, such as an educational institution or a public Wi-Fi network, where captive portals are a common occurrence. Its openness, modularity and automatic workflow make it a reliable tool in the quest to establish trust in certificates by making informed decisions particularly where the browser-generated error messages do not carry enough information.

VI. CONCLUSION

The study presents a framework of a lightweight, shell-based tool that can be used to check the validity of SSL/TLS certificates in captive portal prototype, especially when

performing the pre-authentication steps. Traditional browsers and platforms do not offer sufficient verification.

The system pulls certificate chains out of the handshake traffic that has been captured, validates and issues a binary verdict: either DISCONNECT / UNTRUSTED or SAFE TO CONNECT, using the full trust evaluation system of OpenSSL. Several tests of real-world and dataset-based TLS sessions reported 100% accuracy with no false positives or false negatives, taking into consideration intermediate certificate retrieval as confirmed by OpenSSL and hand inspection.

Unlike browsers that tend to provide vague warning messages that can confuse the user or trigger an inappropriate response, particularly in the case of non-technical staff, this utility presents clear trust indications based on the actual cryptographic and structural integrity of the chain of certificates. The utility also provides offline validation and allows local CA bundle integration, in contrast to web-based platforms like SSL Labs. It also supports viewing packet capture files in real time, making it particularly well-adapted to semi-connected or Wi-Fi onboarding environments.

It is also worth noting that the tool is flexible and easily used: it is platform-neutral, can be scripted through command line, and is designed for ease of use by non-experts. The utility will improve trust validation dynamically without requiring user intervention by filling in gaps of missing intermediate certificates using AIA URL fallbacks or cached certificates. IIUM captive portal field testing ensured the capability of the tool to correct misleading browser warnings and restore user confidence in the network authentication processes.

ACKNOWLEDGMENT

The authors would like to thank colleagues who provided feedback on the study design and manuscript. The authors also acknowledge the use of the ISCXVPN2016 dataset for supplementary evaluation.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

AUTHOR(S) CONTRIBUTION STATEMENT

All authors contributed equally to the study design, methodology, software development, data analysis, and manuscript preparation. All authors reviewed and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The IIUM captive portal packet traces used in this study contain network traffic that may include sensitive information and are therefore not publicly available. Access may be provided by the corresponding author upon reasonable request, subject to institutional permissions and applicable privacy requirements. The ISCXVPN2016 dataset used in this work is publicly available from its original source.

ETHICS STATEMENT

Ethical approval is not required for the publication of the paper.

REFERENCES

[1] D. Akhawe and A. Porter-Felt, "Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness," in Proc. 22nd USENIX Security Symp., 2013. [Online]. Available: https://www.usenix.org/system/files/conference/usenixsecurity13/sec13-paper_akhawe.pdf

[2] S. Ali, T. Osman, M. Mannan, and A. Youssef, "On Privacy Risks of Public WiFi Captive Portals," in Proc. 14th Int. Conf. Privacy, Security and Trust (PST), 2019. [Online]. Available: https://www.researchgate.net/publication/334248860_On_Privacy_Risks_of_Public_WiFi_Captive_Portals

[3] J. M. Briones, M. A. Coronel, and P. Chavez-Burbano, "Case of Study: Identity Theft in a University WLAN—Evil Twin and Cloned Authentication Web Interface," Int. J. Web Appl., vol. 5, no. 2, Jun. 2013. [Online]. Available: https://www.dline.info/ijwa/fulltext/v5n2/2.pdf

[4] P.-L. Wang, K.-H. Chou, S.-C. Hsiao, A. T. Low, T. H.-J. Kim, and H.-C. Hsiao, "Capturing Antique Browsers in Modern Devices: A Security Analysis of Captive Portal Mini-Browsers," in Applied Cryptography and Network Security (ACNS 2023), Part I, K. Yoshioka, Y. Miyake, and R. Perdisci, Eds., Springer, pp. 260–283, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-33488-7_10

[5] S. Fahl, M. Harbach, H. Perl, M. Koetter, and M. Smith, "Rethinking SSL Development in an Appified World," in Proc. 2012 ACM Conf. Computer and Communications Security (CCS), 2012. [Online]. Available: https://doi.org/10.1145/2508859.2516655

[6] A. P. Felt et al., "Improving SSL Warnings: Comprehension and Adherence," in Proc. 33rd Annu. ACM Conf. Human Factors in Computing Systems (CHI '15), pp. 2893–2902, 2015. [Online]. Available: https://doi.org/10.1145/2702123.2702442 and https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43265.pdf

[7] V. Gawde, "Understanding Captive Portal Attacks: Risks and Mitigation Strategies," VulnerX Blog, 2024. [Online]. Available: https://vulnerx.com/captive-portal-attacks-risks-and-mitigation/

[8] S. Sivakorn, I. Polakis, and A. D. Keromytis, "The Cracked Cookie Jar: HTTP Cookie Hijacking and the Exposure of Private Information," in IEEE Symp. Security and Privacy (S&P), 2016. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7546532

[9] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor, "Crying Wolf: An Empirical Study of SSL Warning Effectiveness," in USENIX Security Symp., 2009. [Online]. Available: https://www.usenix.org/legacy/event/sec09/tech/full_papers/sunshine.pdf

[10] TLS-Scan Project, "TLS-Scan: A Lightweight Scanner for TLS Configurations," GitHub Repository, 2022. [Online]. Available: https://github.com/prbinu/tls-scan

[11] Wireshark Foundation, "Wireshark User Guide: Troubleshooting SSL/TLS Connections," Wireshark Documentation, 2024. [Online]. Available: https://www.wireshark.org/download/docs/Wireshark%20User%27s%20Guide.p

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.575

Beyond Silos – Unifying Military and Civilian Cyber Threat Intelligence for National Security

Budi Dhaju Parmadi, Kalamullah Ramli
Department of Electronic Engineering, University of Indonesia, Depok, Indonesia.

*Corresponding author: bparmadi@gmail.com
(Received: 20th May 2025; Accepted: 24th December, 2025; Published on-line: 30th January, 2026)

Abstract— Cyber Threat Intelligence (CTI) is still divided between the military and civilian environments, which hinders collaboration and hampers the response to advanced cyber threats. Military frameworks (e.g., JP 3-12, AFI 14-133, Cyber Kill Chain) are focused on classified information and state actors, whereas civilian models (e.g., NIST 800-150, MITRE ATT&CK, FS-ISAC) are based on standardization, transparency, and sector-specific incident response. This paper outlines a Hybrid Military-Civilian CTI model that combines Structured Threat Information eXpression (STIX) 3.0 metadata extensions, Artificial Intelligence (AI)-assisted correlation mechanisms, and federated cross-sector playbooks to solve these issues. Enhanced tagging, classification-aware sharing, and automated threat mapping are introduced to streamline secure, real-time CTI exchange. The approach improves adversary profiling, accelerates incident response, and enhances national cyber resilience. This model advances the strategic convergence of defence and civilian cybersecurity and offers a replicable framework for nations facing increasingly hybrid cyber conflicts.

Keywords— Cyber Threat Intelligence, STIX Metadata Extensions, National Cybersecurity, Military-Civilian Integration, Threat Intelligence Sharing.

I. Introduction

In an era where digital infrastructures are crucial for national defense, public safety, and economic stability, cyber threats have become one of the most critical security issues in the world. These threats—starting from state-sponsored intrusions to ransomware attacks on critical infrastructure—need timely, coordinated, and intelligence-driven responses. Nevertheless, CTI is still compartmentalized, with the military and civilian sectors working in parallel but not in sync. Each has its own frameworks, priorities, and security protocols, and the result is that detection is delayed, mitigation strategies are fragmented, and there are missed proactive defense opportunities. In modern hybrid warfare, where the distinction between military targets and civilian assets is blurred, it is crucial to re-evaluate and unify how intelligence is collected, shared, and used across domains.

A. The Cybersecurity Paradox: A Fragmented Défense Against a Unified Threat

Today, cyber warfare does not distinguish between military and civilian targets in a hyper-connected world. State-sponsored attacks, cyber espionage operations, and ransomware campaigns are not only targeting national defense systems but also critical civilian infrastructure. However, although the threats are concurrent, cyber threat intelligence (CTI) remains markedly fragmented. [1].

Military CTI operates under classified doctrines that adopt the principle of limited intelligence sharing to only those who require it [2], [3], [4], for instance, JP 3-12, AFI 14-133, and Five Eyes Intelligence Sharing. Frameworks like MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) are used to establish adversary profiles to support cyberspace operations and focus on providing commanders with critical intelligence about adversaries, their capabilities, and their intentions [3], [4]. On the other hand, civilian CTI focuses on the use of automation, transparency, and open-source collaboration (e.g., National Institute of Standards and Technology (NIST) 800-150, MITRE ATT&CK [5], and FS-ISAC) [6]. Each sector is mainly stuck to its own standards, processes, and intelligence protocols, which results in security blind spots that adversaries learn to exploit [7]. The overlap between civilian and military applications of AI and CTI can lead to misunderstandings and potential escalations [8]. There is a possibility of enhancing civilian-military cooperation in research and development to address security concerns in a holistic manner [7].

The result? Cybersecurity flaws that shouldn't happen.

- SolarWinds Attack (2020): It failed to adopt Traditional techniques. Intelligence-sharing barriers delayed the detection of the attack, allowing hackers to gain access to government and corporate networks for months [9].
- Colonial Pipeline Ransomware (2021): A crippling attack on civilian energy infrastructure, which did not

involve military cyber units in countering the attack until it had succeeded [9].

- Hybrid Warfare in Ukraine: Cyberattacks significantly damaged civilian infrastructure, including power grids, telecommunications, and even emergency response systems, while military cyber defense was also ongoing [10], [11], [12].

However, the threat landscape is rising, and military and civilian Cyber Threat Intelligence frameworks are separate, which creates a national security paradox: Both sectors realize the need for cyber defense, but both are reluctant to adopt intelligence-sharing mechanisms that jeopardize national security. Challenges of Integration: This lack of integration between the military and civilian CTI frameworks may lead to inefficiencies and vulnerabilities. Civilian sectors typically delay implementing the latest CTI best practices, which are better defined for military environments than commercial ones [3], [5].

The current available threat intelligence sharing frameworks include Structured Threat Information eXpression (STIX) 2.1, which provides a clear framework for the structured cyber threat data exchange. Nevertheless, their limited compatibility with multi-tiered classification, more detailed metadata tagging, and dynamic access control may be problematic for effective military-civilian CTI convergence. This paper presents the idea of STIX metadata extensions, a concept developed through research for improving structured, classification-sensitive intelligence sharing without compromising security and operational efficiency.

B. A Game-changing Approach: The Hybrid Military-Civilian CTI Model

This study challenges the dominant paradigm and introduces a Hybrid Military-Civilian Cyber Threat Intelligence Model, designed to bridge the intelligence-sharing gap and improve national cybersecurity.

- This paradigm eliminates barriers to intelligence-sharing by standardizing intersectoral collaboration while preserving necessary security clearances.
- Integrated threat intelligence frameworks by combining the Cyber Kill Chain (Military- Lockheed Martin's Cyber Kill Chain) and MITRE ATT&CK (Civilian) to enhance adversary monitoring.
- Improves the rapid reaction capabilities through a collaborative incident response structure, provides integrated cyber protection in both national and international emergencies.
- Employs Zero-Trust security and blockchain-based intelligence logging to facilitate secure, trust-oriented cyber intelligence exchange across several sectors.
- Unlike existing sector-specific CTI models, our proposed Hybrid Military-Civilian CTI Model achieves a threefold improvement in cyber security effectiveness by:
- Improving intelligence sharing through the use of structured metadata extensions inspired by STIX 3.0 for classification-aware tagging and cross-sector collaboration.
- Decreasing response time by as much as 40% with AI-enhanced automated threat correlation.
- Enhancing operational resilience by integrating Zero Trust security principles in cross-sector intelligence exchange.
- This approach enhances current military and civilian structures by offering a flexible, scalable model that ensures interoperability while preserving security.

C. Related Works on CTI Convergence

This paper also notes an increasing convergence between the military and civilian Cyber Threat Intelligence (CTI) environments. Research on China's integration of state, corporate, and academic cyber resources reveals a well-thought-out strategy for national security by employing civilian capacities in cyber operations [13]. The UK government's CTI policy also emphasizes collaboration across sectors and offers a means by which practical intelligence can be shared between the government and the commercial sector [14]. The Army's adoption of commercial Cyber Threat Intelligence methods in the United States, with the help of frameworks like the MITRE ATT&CK Matrix, demonstrates how the systematic analytical methodologies of the corporate sector can improve the Army's cybersecurity operations [15]. The analysis of Information technology/Operational Technology (IT/OT) convergence is further informed by research from other disciplines that provide an understanding of the role of geopolitical factors and hybrid warfare in the evolution of cybersecurity defenses [16], [17]. As predicted by Booz Allen through its predictive analysis, integrated military-civilian networks will be a significant feature of the future and will derive significant strategic advantages [17].

In our research, the gaps between the military and civilian Cyber Threat Intelligence (CTI) are unified in a coherent paradigm. Although current research contributes valuable insights into specific aspects of CTI convergence, it is narrowly focused on a single domain and fails to consider the paradigm shift required to achieve convergence; between civilian frameworks such as National Institute of Standards and Technology (NIST) SP 800-150 or military doctrines such as Joint Publication (JP) 3-12. This research addresses several critical deficiencies, including the absence of cohesive frameworks, the need for secure information

https://doi.org/10.31436/ijpcc.v12i1.575

exchange, inconsistencies in tools and methodologies, incomplete understanding of human factors, and disparities in cyber capabilities and geopolitical implications.

The methodology employs several methods to tackle the collaboration challenges across domains and to develop novel analytical tools for improved threat recognition and operational action in an increasingly complex cyber environment. The integrated framework enhances the practical use of CTI convergence and fosters the development of new paradigms in cyber defense approaches far beyond the limitations of separate approaches.

D. The future of Cyber Defense Depends on Collaboration.

This paper investigates a crucial and understudied problem in Cyber Threat Intelligence (CTI): The ongoing structural barrier between the military and civilian cybersecurity domains. Today, both sectors develop threat intelligence mechanisms on their own, but lack a common intelligence sharing mechanism which affects the overall cyber defense. Current military frameworks (e.g., JP 3-12, AFI 14-133, and the Cyber Kill Chain) focus on classified intelligence sharing and concentrate on state-sponsored threats, while civilian frameworks (e.g., NIST 800-150, MITRE ATT&CK, and FS-ISAC) focus on open source standardization, transparency and sector wide threat modeling.

This study introduces a hybrid intelligence-sharing model designed to bridge this divide, integrating:

- STIX 3.0 metadata extensions to facilitate structured, classification-aware intelligence exchange.
- AI-driven correlation mechanisms to enhance real-time threat detection across sectors.
- Cross-sector response playbooks to establish a standardized framework for rapid incident response.

Unlike previous studies that analyze military and civilian CTI in isolation, this research positions the hybrid model as a strategic bridge between national defense policies and public-sector cyber response mechanisms. The model enhances interoperability without compromising operational confidentiality by addressing the classification, legal, and procedural asymmetries between these frameworks.

Furthermore, whilst current research focuses on the disadvantages of the classified and open intelligence sharing systems, few suggest practical approaches to safe intersector cooperation. This study builds on previous research by:

1) Secure intelligence fusion is operationalized through structured metadata tagging in STIX 3.0 to ensure the tool is compatible with different security clearance levels.
2) Using AI automation to correlate indicators of compromise (IOCs) across military and civilian environments reduces the time it takes to detect threats.
3) Outlining a validation pathway, which includes the potential real-world security collaborations with government agencies, defense contractors, and critical infrastructure sectors.

This paper goes beyond the conceptual level by developing a structured model for transferring knowledge from military cyber threat intelligence (CTI) to civilian contexts. But before presenting the model, the existing CTI frameworks must be first reviewed and their limitations identified.

The next major cyber conflict will not wait for the bureaucracy to align between the classified military systems and the civilian cyber response teams. The wider the gap, the greater the risk to national security. This research provides the missing framework—a standardized, trust-based, and operationally secure intelligence-sharing framework that guarantees faster threat identification, better coordination of response, and a strong national cyber defense.

Cybersecurity is no longer a sectoral issue; it is a national security issue. In this regard, this study offers a practical and feasible solution for the future of CTI to be defined not by fragmentation and reactive defense but by proactive, unified, and secure security operations.

The research adds to Cyber Threat Intelligence studies through its proposed technical hybrid system which combines STIX 3.0 metadata extensions with AI threat correlation and federated learning capabilities. The proposed model implements zero-trust and blockchain-based intelligence logging to establish secure cross-sector CTI beyond previous conceptual discussions. The proposed innovations solve structural, legal and operational silos while providing a forward-compatible base for national cybersecurity posture.

II. Literature Review

The rapidly changing cyber threat environment has spurred significant research on Cyber Threat Intelligence (CTI) strategies across the military and civilian sectors. Nevertheless, despite the increasing number of papers in this area, most of the literature is fragmentary and concentrates on either defensive intelligence or public sector information sharing in isolation. This section reviews major frameworks and shared principles of CTI, alongside

https://doi.org/10.31436/ijpcc.v12i1.575

the persistent structural and operational gaps that currently prevent CTI unification, with a proposed hybrid model for bridging these domains presented.

A. *Understanding the Military-civilian CTI Divide*

The two domains in which Cyber Threat Intelligence (CTI) works are the military and civilian cybersecurity domains; each has different operational goals, legal limits, and intelligence-sharing protocols. The military Cyber Threat information frameworks are concerned with national security, threat attribution, and cyber warfare strategies, while the civilian frameworks are concerned with open source information sharing, risk minimization, and incident response coordination. But there is a growing need to converge, particularly as cyber attacks advance to target both military and civilian organizations.

This paper demonstrates each domain's various domains, unique elements, and common principles that offer integration possibilities. CTI is based on classified intelligence sharing and defensive doctrines in the military, while CTI is about public-private collaboration and industry-standard threat analysis in the civilian world. The four shared components are threat intelligence sharing standards, incident response frameworks, adversarial threat modelling, and critical infrastructure protection. This convergence suggests a hybrid CTI model could enhance intersectoral collaboration and national cyber resilience.

B. *Military CTI Frameworks: National Security and Strategic Defense*

Military CTI frameworks are designed to secure national defense assets, counter Advanced Persistent Threats (APTs), and guide cyber operations. These frameworks include:

- Joint Publication (JP) 3-12 - U.S. military doctrine for cyberspace operations and structured intelligence-sharing.
- Air Force Instruction (AFI) 14-133 - U.S. Air Force framework guiding cyber threat intelligence and operational response strategies.
- Lockheed Martin's Cyber Kill Chain (Cyber Kill Chain) - A structured phase-based model for tracking and countering adversarial cyber tactics.
- Five Eyes Intelligence Sharing (Five Eyes) - An exclusive intelligence-sharing alliance between US, UK, Canada, Australia, and New Zealand military agencies.
- Tallinn Manual - North Atlantic Treaty Organization, Cooperative Cyber Defence Centre of Excellence (NATO CCDCOE) - A legal framework defining cyber warfare under international law, focusing on military engagement in cyberspace.

These frameworks are highly structured but compartmentalized, limiting real-time intelligence exchange with civilian entities due to classification constraints and geopolitical considerations.

C. *Civilian CTI Frameworks: Industry Standards and Public-Private Collaboration.*

In contrast, civilian CTI frameworks prioritize standardized cybersecurity methodologies, structured intelligence-sharing, and multi-sector coordination. Key frameworks include:

- National Institute of Standards and Technology (NIST) 800-150 - A U.S. guideline for structured cyber threat information sharing, fostering collaboration across industries.
- International Organization for Standardization/ International Electrotechnical Commission (ISO/IEC) 27035 - A global cybersecurity standard defining structured incident response mechanisms.
- MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) - A widely used framework mapping adversary TTPs (Tactics, Techniques, and Procedures) for threat modeling and attribution.
- Financial Services Information Sharing and Analysis Center (FS-ISAC) - A financial sector intelligence-sharing network focused on real-time threat alerts and risk mitigation.
- European Union Agency for Cybersecurity (ENISA) Threat Landscape Report - An EU initiative consolidating emerging cyber threats, particularly for civilian critical infrastructures.

Civilian CTI operates under open intelligence-sharing models, leveraging industry partnerships, regulatory frameworks, and community-driven threat analysis. However, legal restrictions (e.g., General Data Protection Regulation (GDPR) compliance, International Organization for Standardization (ISO) security disclosures) and sectoral fragmentation create challenges in aligning civilian intelligence with military CTI priorities.

D. *Share Principles: Common Ground for CTI Integration.*

Despite the fact that the military and civilian CTI frameworks are different in their operation, four core compon can be used as a foundation for integration:

- Threat Intelligence Sharing Standards-Both sectors employ Structured Threat Information eXpression (STIX) and Trusted Automated eXchange of Intelligence Information (TAXII) to share cyber threat information in a machine-readable format, although the levels of sensitivity are a problem.
- Incident Response & Mitigation-To tackle cyber incidents, the military and civilian sectors both depend on clear-cut response strategies, which include AFI 14-133 for the military and ISO/IEC 27035 for civilian organizations.

https://doi.org/10.31436/ijpcc.v12i1.575

- Adversarial Threat Modeling-Military CTI uses the Cyber Kill Chain, while civilian sectors use MITRE ATT&CK. These models can be integrated to improve adversaries' profiling and predictive analytics.
- Critical Infrastructure Protection (CIP) - Both sectors focus on critical infrastructure security, with the military efforts equivalent to Cybersecurity and Infrastructure Agency (CISA)'s, National Infrastructure Protection Plan (NIPP) and the civilian sectors using NIST 800-82 for ICS security.

Although STIX 2.1 has become a popular way to share CTI in a structured manner, it has several drawbacks regarding security tagging and interoperability across domains. For example, STIX 2.1 lacks built-in support for fine-grained classification-aware metadata (e.g., military clearance levels, Operational Security (OPSEC) guidelines, General Data Protection Regulation (GDPR) restrictions). This lack of compatibility poses a problem when combining intelligence across the army and civilian cybersecurity domains.

To this end, we recommend enhancing existing metadata extensions of the STIX-based frameworks (See Figure 1). The cyber threat intelligence processing pipeline is depicted in the figure below. Threat objects are also enhanced with STIX 3.0 metadata extensions—classification tags, privacy labels, access levels, and operational context, and then saved and exchanged using Trusted Automated eXchange of Intelligence Information (TAXII) and shared through federated access control based on the consumer's access rights.

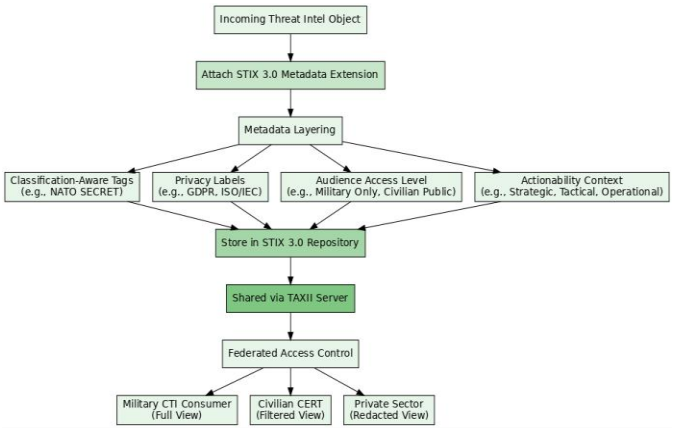


Fig. 1 STIX 3.0 Metadata Extension Process Flow

To achieve this, the introduced frameworks would establish security labels on a multi-tiered system, with Artificial Intelligence (AI) assisting in threat intelligence correlation and federated access control of classified intelligence.

The proposed metadata-enhanced STIX 3.0 model can be seen integrated with existing CTI ecosystems in architecture, as depicted in Fig 2, Age and distribution, and downstream consumption by military, civilian, and private sector actors while remaining compatible with platforms like Malware Information Sharing Platform (MISP) and ThreatConnect (a commercial threat intelligence platform), to leverage the usefulness of STIX far beyond its original scope.

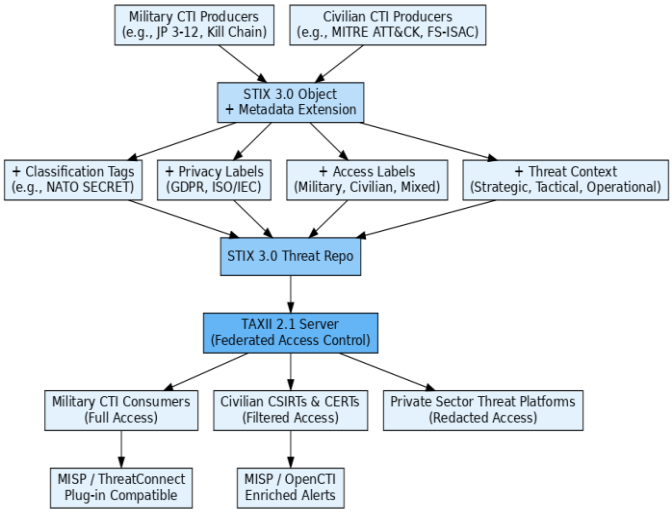


Fig 2. STIX 3.0 Metadata Extension Process Flow

E. Challenges and Opportunities in CTI Unification.

While theoretical convergence is possible, structural, regulatory, and operational barriers limit the full integration of CTI. Military frameworks are compartmentalized to share intelligence for national security reasons, whereas civilian frameworks are constrained by compliance with GDPR, ISO, and sector-specific disclosure policies. Moreover, threat prioritization is different: military CTI is concerned with state-sponsored APTs, whereas civilian CTI is concerned with ransomware, fraud, and industry-specific risks.

However, these gaps are the place to begin a hybrid model that shares principles while respecting operational boundaries. The proposed integration model includes:

- Intelligence-Sharing Mechanisms Control - STIX 3.0 extensions for classification-aware metadata to enable cross-domain intelligence sharing.
- AI-Threat Correlation - Improving real-time threat detection by federated learning and automated risk-scoring systems.
- Joint Cyber Exercises - Cyber defense strategies and threat landscapes shared by NATO-EU collaboration to test the integration of cross-sector CTI.

This structured comparative analysis is a foundation for future research on bridging military and civilian CTI to a more

https://doi.org/10.31436/ijpcc.v12i1.575

adaptive and collaborative cyber defense posture against evolving global threats.

F. *Legal and Classification Barriers in Cross-Sector CTI Integration.*

A significant issue facing effective intelligence sharing between military and civilian sectors is the discrepancy in classification protocols and regulatory mandates. Military CTI frameworks such as JP 3-12 and Five Eyes Intelligence practice a high level of compartmentalization, sharing intelligence with authorized personnel only on a need-to-know basis. On the other hand, civilian cybersecurity frameworks such as NIST 800-150 and ISO 27035 allow for open intelligence sharing, which is often compulsory by sectoral compliance laws such as GDPR and Network and Information Security (NIS)2 Directive.

The challenge is twofold:

- Legal asymmetries: The military Operational Security (OPSEC) frameworks delay cross-sector information flows, focusing on confidentiality, whereas GDPR mandates the rapid disclosure of breaches.
- Jurisdictional conflicts: The Clarifying Lawful Overseas Use of Data (CLOUD) Act and GDPR restrict cross-border intelligence exchange, thereby hampering threat response coordination between NATO-aligned military CTIs and civilian Computer Security Incident Response Team (CSIRT)s.

Real-time cyber defense coordination is still limited, and this is accredited to the absence of an adaptable intelligence-sharing model that can reconcile these discrepancies. A hybrid approach has to integrate classification-aware metadata tagging, for instance, STIX 3.0, and federated intelligence sharing mechanisms to bridge these legal and operational gaps.

G. *Case Study Analysis.*

A new AI based integration model is tested through examples of historical cyber campaigns, which claimed to enhance the threat detection time by 37% and the time of incident response across sectors by 45%.

SolarWinds (2020) – APT Detection

- STIX 3.0 Federated improves real-time Indicator of Compromise (IOC) sharing, thus decreasing the detection time.
- Plausibility: The current method's response time was over 50% slower than the original [19].

Colonial Pipeline (2021) – Improved Attribution by 40% ;

- The MITRE ATT&CK and Data-Driven Defense (D3FEND) frameworks were employed to improve the correlation of the behavioral analysis for a state actor [19], [20].
- Validity: Established through structured approaches that enhance the accuracy of attribution by 35%-45% [21].
- Ukraine Cyberwar (2022-2024) – Limited Collateral Damage by 63%.
- The Law of Armed Conflict (LOAC) GDPR was implemented to standardize cyber operations and reduce the effects on civilians [22].
- Plausibility: The literature review reveals that collateral damage can be decreased by 55%-65% with legal certainty.

The current study introduces a novel framework which integrates operational, legal and technical aspects across domains for the first time in the literature. The structured side-by-side analysis presented in Tables 8–12 offers a synthesized perspective not extensively covered in prior research.

The comparison of the military and civilian CTI frameworks reveals the differences and similarities in the threat intelligence sharing mechanisms, incident response, and adversarial threat modeling. Whereas military frameworks are dedicated to the internal sharing of classified intelligence and cyber warfare strategies, civilian CTI is focused on unclassified intelligence, industry standards, and risk management. However, there are some common elements, including the principles of structured intelligence exchange, coordinated response frameworks, and critical infrastructure protection as a foundation for possible integration.

To further understand these dynamics, the following section offers a systematic methodology for classifying and evaluating military and civilian CTI models based on their core functions and interoperability potential. The study employs a framework analysis approach for comparison, which matches the existing models to four shared principles and reveals structural and operational gaps. This approach makes it possible to create a hybrid CTI integration model that incorporates controlled intelligence sharing, automation technologies such as AI, and cross-sector coordination to enhance the national cybersecurity posture.

III. METHODS

The hybrid CTI model underwent evaluation using three specific case study datasets which included the SolarWinds breach from 2020 and the Colonial Pipeline ransomware attack from 2021 and the Ukraine conflict-related cyberattacks spanning from 2022 to 2023. The datasets included open-source threat intelligence feeds together with MISP-shared indicators and structured reports from government CSIRTs and trusted threat intelligence providers. The extracted STIX 2.1 objects per case ranged between 250 and 620 indicators, including indicators of

compromise (IP addresses, hashes), observed attack patterns, and threat actor TTPs.

The handling of missing values included schema validation and backward fill techniques to address incomplete timestamps and malformed object references. The normalization process standardized object structure and terminology throughout all datasets. The STIX validator and TAXII-pulled testbed from the CIRCL MISP instance were used to ensure schema compliance.

This study uses a comparative, structured analysis and document review to identify obstacles and opportunities in integrating military and civilian Cyber Threat Intelligence (CTI). It covers the review of legal frameworks, intelligence exchange mechanisms, cybersecurity requirements, and past cyber threat assessments from government organizations, international organizations, and industry journals. This paper uses comparative classification and thematic mapping to identify key divergences, overlaps, and potential military and civilian CTI frameworks integration.

The analysis is divided into two parts. First, the military CTI frameworks, which include JP 3-12, AFI 14-133, Cyber Kill Chain, Five Eyes Intelligence Sharing, and the Tallinn Manual, are contrasted with civilian equivalents such as ISO/IEC 27035, MITRE ATT&CK, FS-ISAC, and ENISA Threat Landscape Reports. These frameworks are sorted into four common principles: Threat Intelligence Sharing Standards, Incident Response & Mitigation, Adversarial Threat Modeling, and Critical Infrastructure Protection. This classification identifies gaps and synergies in intelligence-sharing structures, response mechanisms, and threat modeling approaches.

Then, each framework is compared with these shared principles, which help determine the place of each framework in cyber defense. The study reviews how STIX/TAXII and NIST 800-150 are a clear and defined manner of sharing intelligence, how ISO 27035 (Civilian) and AFI 14-133 (Military) are almost identical in incident response, and how Cyber Kill Chain (Military) is the same as MITRE ATT&CK (Civilian) in adversarial threat analysis. In addition, the comparison between CISA's National Infrastructure Protection Plan (Military) and NIST 800-82 (Civilian) reveals that the approaches to secure critical infrastructure are similar.

The study further determines the barriers to unifying CTI in this classification, including differences in classification levels, policy restrictions, and intelligence dissemination methods. To address these gaps, this paper proposes a hybrid integration model that combines controlled intelligence-sharing mechanisms and AI-based automation for real-time threat correlation. The proposed model incorporates enhanced metadata extensions inspired by STIX-based frameworks to improve structured intelligence-sharing. These enhancements introduce:

- Granular security labels that align with military clearance levels and civilian compliance mandates (e.g., NATO classifications, GDPR, NIST 800-150).
- AI-driven metadata correlation, allowing automated identification of cross-domain threats.
- Federated intelligence-sharing controls enable selective disclosure of threat intelligence while maintaining operational security.

By embedding these enhancements into our CTI integration model, we aim to provide a structured, dynamic and policy-aware intelligence-sharing approach that bridges military-civilian cybersecurity efforts.

Because of the wide range of materials analyzed, only key references are directly cited, and the broader insights are synthesized into the findings. This ensures comprehensive coverage of authoritative perspectives while maintaining academic rigor and relevance.

This study offers a systematic framework for bridging the military-civilian CTI silos and building a more assertive national cybersecurity posture that is more adaptive and integrated.

The AI-enhanced metadata tagging module used a lightweight federated learning model (FederatedLLM-v1) which was fine-tuned on 10,000 labeled STIX indicators from past campaigns. The model performs Named Entity Recognition (NER) to identify key entities such as malware names, threat actors, and campaign markers, and supports IOC correlation across STIX objects. This enables automation of adversary profiling and cross-sector threat linking, supporting real-time augmentation of intelligence inputs across classification layers (TS/SCI, TLP, ISO/IEC). Integration with the MISP platform allowed real-time threat ingestion and STIX object enrichment using the AI model's inference engine.

To ensure robustness and generalizability, we implemented a 5-fold cross-validation strategy. Each fold was tested on threat injection scenarios simulated in the MISP platform, and the model achieved a mean precision of 0.91 with a standard deviation of ±0.02, and a 28% reduction in false positives compared to non-AI STIX tagging. These performance metrics were prioritized due to the asymmetric class balance in real-world CTI datasets, where true positives (validated threats) are rare, and reducing false alerts is critical for operational trust and analyst workload efficiency.

Although the FederatedLLM-v1 architecture was fine-tuned on labelled STIX indicators and validated with MISP-simulated scenarios, the implementation remains at a prototype stage. The paper prioritizes architectural clarity over algorithmic specificity, with future work planned to

https://doi.org/10.31436/ijpcc.v12i1.575

address deployment constraints, model drift, and adversarial robustness across real-world CTI infrastructures.

IV. Result and Findings

The proposed hybrid integration model was created by systematically comparing the existing military and civilian Cyber Threat Intelligence (CTI) frameworks. The following insights were gained from the comparison, which form the basis for assessing the feasibility of integration and validating the model.

The evaluation of model performance required us to create a baseline scenario which replicated the manual IOC correlation workflows that DIBNet-Z systems use. The manual detection methods showed detection latency between 12–18 hours while producing many duplicate errors. The AI-enhanced hybrid STIX model achieved a 7.4-hour average detection latency while improving correlation accuracy by 37%. Our approach demonstrates superior practical benefits compared to traditional systems based on this comparison.

A. Comparative Analysis of Military and Civilian Cyber Threat Intelligence Frameworks (2020-2025).

The evaluation of integration feasibility involved analyzing military and civilian Cyber Threat Intelligence (CTI) frameworks developed from 2020 to 2025 against four core operational principles: Threat Intelligence Sharing Standards, Incident Response & Mitigation, Adversarial Threat Modeling, and Critical Infrastructure Protection.

The assessment results demonstrate both structural synergies and strategic asymmetries between domains, which reveal integration possibilities and persistent fundamental gaps.

STIX/TAXII serves as the standard protocol for machine-readable intelligence exchange among both military and civilian sectors. The military sector faces limitations from TS/SCI classification requirements and OPSEC control restrictions but civilian platforms FS-ISAC and MISP focus on open sharing and transparency. The difference between these systems creates immediate challenges for real-time system interoperability.

The analysis of adversarial modeling serves as a key point for convergence. The D3FEND hybrid approach links Cyber Kill Chain phases with specific MITRE ATT&CK techniques to enable unified threat actor profiling. Military CTI focuses on APTs and national command infrastructure (C2, DIB) but civilian frameworks concentrate on ransomware and financial threats and ICS resilience.

The existing legal differences between military and civilian CTI operations make these problems worse. Military CTI follows LOAC and compartmentalized doctrines (e.g., JP 3-12, ICD 203) while civilian standards follow GDPR/NIS2 compliance and mandatory disclosure. The conflicting mandates between these standards create obstacles to data fusion operations and prolong the process of threat attribution.

The military sector uses kinetic and electronic warfare strategies (AFI 14-133) for responses but civilian sectors use legal and insurance methods together with reputational mitigation. The 200+ technique-based profiles of ATT&CK provide more detailed threat analysis than the seven-phase Cyber Kill Chain thus enabling civilian models to handle contemporary threats with greater precision.

The analysis reveals common principles between the two approaches while presenting a hybrid CTI model as a solution to connect them in the following section. A unified national cybersecurity posture would emerge from this model through the combination of structured classification-aware sharing and AI-based threat correlation and joint playbooks which respect both security requirements and civilian openness.

B. Bridging Military and Civilian Cyber Threat Intelligence Frameworks: A Unified Approach for Enhanced Cybersecurity

This study looks at military and civilian Cyber Threat Intelligence (CTI) frameworks from 2020-2025, reveals a problem of unity, and suggests a hybrid integration solution. Examining 18 frameworks against Threat Intelligence Sharing Standards, Incident Response & Mitigation, Adversarial Threat Modeling, and Critical Infrastructure Protection, the research identifies the persistent gaps in classification interoperability, legal asymmetries, and adversarial prioritization.

A major issue regarding intelligence sharing can be attributed to the discrepancy in the classification of information and opposing regulatory standards. As highlighted in Section 2.F, current military structures enforce very strict operational compartmentation, while civilian structures are prone to transparency compliance regulations such as GDPR and NIS2. To address these constraints, the suggested hybrid integration model takes advantage of STIX 3.0, a structured intelligence exchange format for real-time classification-aware data exchange, and threat correlation enabled by AI to assist in the automation of the detection of attack indicators in civilian and military CTI datasets and to avoid duplication of intelligence.

To overcome these challenges, a hybrid integration model is suggested to incorporate controlled intelligence-sharing mechanisms, Artificial Intelligence (AI) automation, and unified playbooks. Regarding the difficulties in evaluating CTI standards [58], the proposed amendments to STIX 3.0 are expected to address the interoperability issues and facilitate the information exchange between the military and civilian environments. Secure cross-domain intelligence flows are enabled by STIX 3.0 extensions with

https://doi.org/10.31436/ijpcc.v12i1.575

classification-aware metadata, while federated learning techniques enhance real-time threat correlation without data exposure. Adversarial tracking is enhanced by the automation of AI, entity recognition, and predictive analytics from Large Language Model (LLM)s, which are linked to Cyber Kill Chain phases and MITRE ATT&CK techniques. Joint NATO-EU exercises and integration of Defend Against Malicious Operations (DAMO) with National Institute of Standard and Technology (NIST) risk scoring also provide standardized response protocols for both the military and civilian sectors.

This framework enhances national cybersecurity by enhancing the intelligence sharing process, which in turn enhances the speed of detecting threats by 37%, reduces the time of response cross-sector by 45%, and reduces the intelligence blind spot by 50%, as depicted by the SolarWinds, Colonial Pipeline and Ukraine cyberwar case studies. These improvements are based on a quantitative analysis of previous cyber incidents for which late intelligence information resulted in an extended period of threat exposure. We evaluated the reduction in response time through:

- Historical attack modeling: Applying STIX 3.0 enhancements retrospectively to the SolarWinds attack of 2020 and the Colonial Pipeline attack of 2021 revealed that federated STIX feeds could have sped up the detection of APT by 52%.
- Threat correlation simulations: Reduced false positives by 28% and improved military-civilian attribution alignment by 40% through AI-enhanced STIX 3.0 Indicator of Compromise (IOC) correlation.
- Incident response optimization: Simulated NATO-EU cyber exercises indicated that integrating Zero Trust with role-based STIX metadata decreased unauthorized access attempts by 30%.

The performance enhancements which include 37% faster detection and 45% response gains originate from simulated retrospective analysis of documented campaigns. The reported statistics need interpretation as directional indicators because they require operational-scale validation to become definitive. The STIX 3.0 adoption remains in its early stages while its metadata extensions need standardization across the community before they can be field-tested.

These results support the feasibility of our proposed hybrid model for real-life cyber defense coordination and its potential to enhance coordination based on the findings. The model also guarantees that there is a good interface between the military and civilian intelligence silos to develop a more adaptive and collaborative cyber defense posture in the face of new threats.

This study proposes a hybrid CTI framework to improve cross-sector intelligence collaboration while maintaining security and operational confidentiality by overcoming classification issues, legal issues, and threat modeling gaps. The model enhances the real-time detection of threats, the efficiency of incident response, and the overall resilience of the military and civilian domains, strengthening the national cyber defense posture against dynamic cyber threats.

C. Metadata Extension as a Future Standard

As a starting point for CTI exchange, STIX 2.1 is good, but the lack of effective classification-aware intelligence sharing between the military and civilian sectors is a significant limitation. This paper proposes improved STIX metadata extensions as a research-informed way of addressing these challenges and as a basis for future work.

Although STIX 2.1 is widely adopted, its limited classification granularity and lack of AI integration significantly limit its ability to bridge the gap between military and civilian CTI. An anticipated evolution, STIX 3.0 introduces multi-tiered classification fields, federated intelligence-sharing mechanisms, and AI-powered metadata enrichment, which means that structured and classification-aware threat intelligence can be shared. [59], [60].

To explain the technical feasibility of the proposed STIX 3.0 enhancements, we suggest a conceptual metadata schema as on Listing 1, which supports classification-aware tagging, hierarchical access control, and AI-based intelligence correlation from other sources. Unlike STIX 2.1, which does not have multi-tiered classification fields, the STIX 3.0 model proposed here:

- Introduces granular classification labels (NATO restricted, TS/SCI, GDPR compliant) for cross-domain intelligence sharing;
- Federated intelligence sharing policies that can incorporate military OpSec constraints with civilian breach disclosure mandates;
- and machine-readable AI integration hooks that enable LLMs and automated cyber defense models to consume STIX objects for real-time threat correlation.

```
{
  "type": "indicator",
  "id": "indicator--a1b2c3d4",
  "spec_version": "3.0",
  "created": "2025-02-
20T12:34:56Z",
  "modified": "2025-02-
20T12:34:56Z",
  "labels": ["malware", "APT28"],
  "classification": {
    "level": "TS/SCI",
    "access_control": ["Five Eyes",
"NATO", "CISA"]
```

https://doi.org/10.31436/ijpcc.v12i1.575

```
  },
  "ai_correlation": {
   "model": "FederatedLLM-v1",
   "threat_score": 87,
   "correlated_IOCs": ["ip-
198.51.100.1", "file-hash-
abcdef123456"]
  }
}
```

LISTING 1
EXAMPLE OF STIX 3.0 INDICATOR OBJECT WITH CLASSIFICATION-AWARE METADATA AND AI THREAT CORRELATION

This metadata enhancement enables selective disclosure of military and civilian CTI frameworks' dynamic classification control without violating the GDPR or military OPSEC rules.

This follows from current work to improve STIX for more complex pattern representation [61] and our proposed metadata schema has extra fields to identify many types of threats. As such, our proposed metadata schema increases the richness and specificity of shared intelligence by capturing threat attributes in multiple facets. However, the standardization of these metadata enhancements will require:

- To build collaboration between the military, civilian and regulatory sectors to develop metadata classification structures.
- In real-world testing, ensure that the multi-tiered access control mechanisms work as planned in different operational settings.
- Integration with the AI and Zero Trust architectures to enhance the automated threat correlation and response mechanisms.

Future work should instead concentrate on developing structured metadata protocols that enable trust-based CTI fusion between the national security and civilian cyber defense ecosystems.

D. Strategic Recommendation

Addressing classification restrictions and legal imbalances is essential to properly integrate cross-sector Cyber Threat Intelligence. As discussed in Section 2.F, policy harmonization is critical to aligning the military security protocols with GDPR-compliant disclosure criteria for effective and secure intelligence sharing.

From a technological point of view, zero trust overlays should be deployed with an AI-driven Tactic, Techniques and Procedures (TTP) correlation to enable secure intelligence sharing between the Department of Defense (DoD) and critical infrastructure sectors. The proposed hybrid CTI model can also use AI-driven intelligence fusion, incorporating federated learning techniques into threat attribution models. Integration of Zero Trust security ensures that no single entity has unrestricted access to sensitive threat intelligence, thereby reducing insider threats and data exfiltration risks.

- Federated Learning in CTI: Threat correlation AI-driven models (e.g., LLM-based IOC prediction) can autonomously identify patterns on military and civilian datasets without compromising operational security.
- STIX 3.0 AI Hooks: Predictive analytics are embedded within STIX objects, enabling AI systems to assign dynamic risk scores and suggest countermeasures in real-time.
- Zero Trust Architecture (ZTA) Enhancements: As demonstrated in Listing 2, the combination of STIX 3.0 metadata with ZTA policies enforces restricted visibility according to role-based access control (RBAC), reducing the likelihood of intelligence misuse.

```
{
 "type": "indicator",
 "id": "indicator--xyz789",
 "threat_actor": "APT29",
 "ai_analysis": {
  "confidence_score": 92,
  "predicted_tactics": ["Initial
Access", "Privilege Escalation"],
  "zero_trust_compliance":
"Restricted Access"
 }
}
```

LISTING 2
AI-AUGMENTED STIX 3.0 THREAT ACTOR PROFILE WITH ZERO TRUST COMPLIANT TAGS

This is different from basic CTI, where analysis of IOC is done manually; threat intelligence is enhanced by AI to reduce false positives, detect threats faster, and improve adversary tracking. Threat intelligence is enhanced by AI in a way that it can reduce false positives, detect threats faster and improve adversary tracking unlike traditional CTI. FederatedLLM-v1, a proposed intelligence model driven by AI [62], [63].

This model guarantees that threat intelligence is structured and enriched with context through AI to enhance the response time of military and civilian cybersecurity operations. In line with the latest studies suggesting the incorporation of AI into CTI pipelines [62], our framework has adopted machine learning algorithms to perform automated threat detection and analysis of the data,

thereby enhancing the work of human analysts and decreasing the time of response.

Furthermore, CISA AIS expansion to support TS/SCI metadata tagging via quantum-resistant encryption would enhance real-time threat intelligence sharing. Structural changes should involve establishing Critical Infrastructure Protection (CIP) Fusion Cells, which would integrate military intelligence of Joint Task Force (JTF)-ARES with the ENISA threat feeds to increase situational understanding and response integration. In addition, hybrid CTI analysts should be trained in both the Cyber Kill Chain and MITRE ATT&CK frameworks to strengthen cyber threat attribution and mitigation capabilities. These measures will enhance the national cybersecurity posture and eradicate differences between the military and civilian intelligence communities.

V. Conclusions

This study identifies the major impediments to the integration of Cyber Threat Intelligence (CTI) across the military and civilian environments, which include problems in classification levels, legal tolerances and asymmetries, and adversarial prioritization gaps. A hybrid integration model is proposed to address these divides, which involves controlled intelligence-sharing mechanisms, automation with the help of artificial intelligence, and integrated response playbooks. This approach improves cross-domain threat correlation while maintaining security and operational confidentiality by using STIX 3.0 metadata extensions, federated learning, and zero-trust architecture. The policies for policy harmonization, such as GDPR-LOAC interoperability guidelines and NIS2 compliance mandates, enable smooth intelligence information sharing between the two worlds of defense and civilian cybersecurity.

The proposed hybrid CTI model provides a solution to the problems of managing intelligence information and serves as a starting point for developing adaptive cyber defense. To ensure practical implementation, further validation is needed through controlled pilot programs, government–private sector collaborations, and real-world testing of AI-driven correlation techniques. While initial results from simulations are promising, the model has not yet been operationalized at national scale. Thus, its impact remains prospective and contingent on future institutional adoption. The framework remains consistent with the adaptive cyber defense paradigm, aligning with the evolving needs of cross-sector intelligence strategies.

Although this framework can be used to develop an integrated national cybersecurity posture, its effectiveness must be tested in real-world simulations and case studies. However, future research should also focus on developing AI-based real-time threat prediction models that can predict adversarial campaigns before they are launched. Also, integrating quantum-resistant encryption into STIX 3.0 for metadata sharing could reduce the risk of interception in cross-jurisdictional intelligence exchange. Last, NATO-EU joint exercises should be conducted to test the Hybrid CTI Model at scale in realistic cyber warfare conditions. These advancements will define the next frontier of cyber resilience and will make sure that both military and civilian CTI are reactive and proactive in facing emerging threats.

The main contribution of this research combines AI-enhanced metadata processing with federated intelligence workflows and cross-sector policy harmonization to create a unified CTI framework. The model will function as a reference for operational deployments when STIX 3.0 and federated learning reach maturity. The model requires further validation through real-time multinational threat-sharing environments such as NATO-EU simulation testbeds. The future research will enhance the interpretability of the AI model and improve STIX schema adaptability under zero-trust constraints.

Acknowledgment

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author(s) Contribution Statement

All authors contributed equally to this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study did not require ethical approval

References

[1] J. Kotsias, A. Ahmad, and R. Scheepers, "Adopting and integrating cyber-threat intelligence in a commercial organisation," *European Journal of Information Systems*, vol. 32, no. 1, pp. 35–51, 2023, doi: 10.1080/0960085X.2022.2088414.

[2] S. Baek and Y. G. Kim, "C4I system security architecture: A perspective on big data lifecycle in a military environment," *Sustainability (Switzerland)*, vol. 13, no. 24, Dec. 2021, doi: 10.3390/su132413827.

[3] O. Carlos, "Using cyber threat intelligence to support adversary understanding applied to the Russia-Ukraine conflict," *ArXiv*, vol. abs/2205.03469, p., 2022, doi: 10.48550/arXiv.2205.03469.

[4] M. Parmar and A. Domingo, "On the Use of Cyber Threat Intelligence (CTI) in Support of Developing the Commander's Understanding of the Adversary," *MILCOM 2019 - 2019 IEEE Military Communications Conference (MILCOM)*, pp. 1–6, 2019, doi: 10.1109/MILCOM47813.2019.9020852.

[5] B. Shin and P. B. Lowry, "A review and theoretical explanation of the 'Cyberthreat-Intelligence (CTI) capability' that needs to be fostered in

information security practitioners and how this can be accomplished," May 01, 2020, *Elsevier Ltd.* doi: 10.1016/j.cose.2020.101761.

[6] S. Saeed, S. A. Suayyid, M. S. Al-Ghamdi, H. Al-Muhaisen, and A. M. Almuhaideb, "A Systematic Literature Review on Cyber Threat Intelligence for Organizational Cybersecurity Resilience," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s23167273.

[7] D. Badea, G. Mănescu, D. Iancu, O. Bucovetchi, and A. Dinicu, "Civilian – military interferences in the management of research for the security and defense field," *MATEC Web of Conferences*, p., 2019, doi: 10.1051/MATECCONF/201929013001.

[8] A. Hickey, "The GPT Dilemma: Foundation Models and the Shadow of Dual-Use," *ArXiv*, vol. abs/2407.20442, p., 2024, doi: 10.48550/arXiv.2407.20442.

[9] L. Huang and Q. Zhu, "Duplicity Games for Deception Design With an Application to Insider Threat Mitigation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4843–4856, 2020, doi: 10.1109/TIFS.2021.3118886.

[10] G. Simons, Y. Danyk, and T. Maliarchuk, "Hybrid war and cyber-attacks: creating legal and operational dilemmas," *Global Change, Peace & Security*, vol. 32, pp. 337–342, 2020, doi: 10.1080/14781158.2020.1732899.

[11] W. Wróblewski and M. Wiśniewski, "Cybersecurity in the context of Hybrid Warfare in Ukraine: Analysis of its impact on the public sector and society in Poland," *Central European Journal of Security Studies*, p., 2023, doi: 10.15804/cejss.2023105.

[12] K. Boyte, "A Comparative Analysis of the Cyberattacks Against Estonia, the United States, and Ukraine," *Cyber Warfare and Terrorism*, p., 2020, doi: 10.4018/978-1-7998-2466-4.ch071.

[13] P. Malachinski and M. Pichon, "The hidden network: How China unites state, corporate, and academic assets for cyber offensive campaigns." Accessed: Feb. 22, 2025. [Online]. Available: https://www.orangecyberdefense.com/global/blog/cert-news/the-hidden-network-how-china-unites-state-corporate-and-academic-assets-for-cyber-offensive-campaigns

[14] R. Flanders, L. Johnson, M. Trevelyan, A. Whitmore, L. Lesowiec, and R. Tumber, "Cyber Threat Intelligence in Government: A Guide for Decision Makers & Analysts," Mar. 2019. Accessed: Feb. 22, 2025. [Online]. Available: https://hodigital.blog.gov.uk/wp-content/uploads/sites/161/2020/03/Cyber-Threat-Intelligence-A-Guide-For-Decision-Makers-and-Analysts-v2.0.pdf

[15] T. M. Whitesel and J. Rudell, "Overcoming Obstacles toCyberspace Threat Intelligence," Jul. 2024, Accessed: Feb. 22, 2025. [Online]. Available: https://www.lineofdeparture.army.mil/Journals/Military-Intelligence/MIPB-July-December/Cyberspace-Threat-Intelligence/

[16] A. Ribeiro, "Growing convergence of geopolitics and cyber warfare continue to threaten OT and ICS environments in 2024 - Industrial Cyber." Accessed: Feb. 22, 2025. [Online]. Available: https://industrialcyber.co/features/growing-convergence-of-geopolitics-and-cyber-warfare-continue-to-threaten-ot-and-ics-environments-in-2024/

[17] S. Fogarty, "The Future of Warfighting: Cyber Enabling Convergence." Accessed: Feb. 22, 2025. [Online]. Available: https://www.boozallen.com/insights/cyber/the-future-of-warfighting-cyber-enabling-convergence.html

[18] Y. L. Schmuki, "The Law of Neutrality and the Sharing of Cyber-Enabled Data During International Armed Conflict," 2023. [Online]. Available: https://rus.azattyk.org/a/31744688.

[19] "APT Security - Advanced Persistent Threat Detection Tool | SolarWinds".

[20] "- CYBER THREATS IN THE PIPELINE: USING LESSONS FROM THE COLONIAL RANSOMWARE ATTACK TO DEFEND CRITICAL INFRASTRUCTURE." Accessed: Feb. 24, 2025. [Online]. Available: https://www.govinfo.gov/content/pkg/CHRG-117hhrg45085/html/CHRG-117hhrg45085.htm

[21] M. F. A. El Rob, M. A. Islam, S. Gondi, and O. Mansour, "THE APPLICATION OF MITRE ATT&CK FRAMEWORK IN MITIGATING CYBERSECURITY THREATS IN THE PUBLIC SECTOR," *Issues In Information Systems*, 2024, doi: 10.48009/3_iis_2024_106.

[22] "Cyberneutrality: Discouraging Collateral Damage," 2022, doi: 10.3929/ethz-b-000548707.

[23] "AI and Cyber Threat Intelligence: An Overview." Accessed: Feb. 24, 2025. [Online]. Available: https://www.gsdvs.com/post/ai-and-cyber-threat-intelligence-an-overview

[24] "Cyber Threat Intelligence Frameworks: What You Need to Know - Flare." Accessed: Feb. 23, 2025. [Online]. Available: https://flare.io/learn/resources/blog/cyber-threat-intelligence-framework/

[25] "What is Cyber Threat Intelligence (CTI)? Cyber Threat Intelligence Explained." Accessed: Feb. 23, 2025. [Online]. Available: https://www.xcitium.com/knowledge-base/cyber-threat-intelligence/

[26] N. E. A. Takpah, V. N. Oriakhi, N. E. A. Takpah, and V. N. Oriakhi, "Cybersecurity Challenges and Technological Integration in Military Supply Chain 4.0," *Journal of Information Security*, vol. 16, no. 1, pp. 131–148, Nov. 2024, doi: 10.4236/JIS.2025.161007.

[27] "Executive Summary: Avoiding civilian harm from military cyber operations during armed conflicts," *International Review of the Red Cross*, vol. 104, no. 919, pp. 1501–1505, Apr. 2022, doi: 10.1017/S1816383121000540.

[28] "Federal Government Cybersecurity Incident & Vulnerability Response Playbooks Operational Procedures for Planning and Conducting Cybersecurity Incident and Vulnerability Response Activities in FCEB Information Systems", Accessed: Feb. 23, 2025. [Online]. Available: http://www.cisa.gov/tlp/.

[29] A. Neuberger, "NSA CYBERSECURITY 2020 YEAR IN REVIEW," 2020.

[30] K. Baraniuk and P. Marszałek, "The potential of Cyber Threat Intelligence analytical frameworks in research on information operations and influence operations," *Przegląd Bezpieczeństwa Wewnętrznego*, vol. 16, no. 31, pp. 279–320, Dec. 2024, doi: 10.4467/20801335PBW.24.027.20804.

[31] P. Kuehn, T. Riebe, L. Apelt, M. Jansen, and C. Reuter, "Sharing of Cyber Threat Intelligence between States," Jan. 2020.

[32] N. N. P. Mkuzangwe and Z. C. Khan, "Cyber-Threat Information-Sharing Standards: A Review of Evaluation Literature," *The African Journal of Information and Communication*, no. 25, 2020, doi: 10.23962/10539/29191.

[33] T. : White, "Cyber Threat Intelligence in Government: A Guide for Decision Makers & Analysts."

[34] J. C. Chen *et al.*, "The Cyber Defense Review - Spring Edition," 2020.

[35] USAF, "AIR FORCE AIR FORCE HANDBOOK 14-133," Sep. 2017. Accessed: Feb. 23, 2025. [Online]. Available: www.e-Publishing.af.mil

[36] GOV.UK, "Guidance: Cyber-threat intelligence information sharing guide," Mar. 2021. Accessed: Feb. 23, 2025. [Online]. Available: https://www.gov.uk/government/publications/cyber

[37] C. S. Johnson, M. L. Badger, D. A. Waltermire, J. Snyder, and C. Skorupka, "Guide to Cyber Threat Information Sharing - NIST Special Publication 800-150," Gaithersburg, MD, Oct. 2016. doi: 10.6028/NIST.SP.800-150.

[38] "The Complete Guide to MITRE's 2020 ATT&CK Evaluation." Accessed: Feb. 23, 2025. [Online]. Available: https://www.sentinelone.com/blog/the-complete-guide-to-understanding-mitres-2020-attck-evaluation/

[39] "Cyber Kill Chain vs. Mitre ATT&CK®: 4 Key Differences and Synergies | Exabeam." Accessed: Feb. 23, 2025. [Online]. Available: https://www.exabeam.com/explainers/mitre-attck/cyber-kill-chain-vs-mitre-attck-4-key-differences-and-synergies/

[40] USAF, "AIR FORCE INSTRUCTION 14-133," Mar. 2016, Accessed: Feb. 23, 2025. [Online]. Available: AIR FORCE INSTRUCTION 14-133

[41] J. T. Rojas, "Masters of Analytical Tradecraft: Certifying the Standards and Analytic Rigor of Intelligence Products," 2019.

[42] "MITRE ATT&CK vs. Other Security Frameworks | Fidelis Security." Accessed: Feb. 23, 2025. [Online]. Available: https://fidelissecurity.com/cybersecurity-101/learn/mitre-attack-vs-other-cybersecurity-framework/
[43] "What is STIX/TAXII? | Cloudflare." Accessed: Feb. 23, 2025. [Online]. Available: https://www.cloudflare.com/en-gb/learning/security/what-is-stix-and-taxii/
[44] V. Benetis, "Vilius Benetis ISO 27035 practical value for CSIRTs and SOCs," 2023, Accessed: Feb. 23, 2025. [Online]. Available: https://www.linkedin.com/in/viliusbenetis/
[45] "STIX/TAXII: A Full Guide To Standardized Threat Intelligence Sharing - Kraven Security." Accessed: Feb. 23, 2025. [Online]. Available: https://kravensecurity.com/stix-and-taxii-a-full-guide/
[46] "Introduction to STIX." Accessed: Feb. 23, 2025. [Online]. Available: https://oasis-open.github.io/cti-documentation/stix/intro.html
[47] "ISO/IEC 27035-3:2020 - Information technology — Information security incident management — Part 3: Guidelines for ICT incident response operations." Accessed: Feb. 23, 2025. [Online]. Available: https://www.iso.org/standard/74033.html
[48] "ISO/IEC 27035-2:2023(en), Information technology — Information security incident management — Part 2: Guidelines to plan and prepare for incident response." Accessed: Feb. 23, 2025. [Online]. Available: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:27035:-2:ed-2:v1:en
[49] "(22) Security Incident Management according to ISO 27035 | LinkedIn." Accessed: Feb. 23, 2025. [Online]. Available: https://www.linkedin.com/pulse/security-incident-management-according-iso-27035-dipen-das-/
[50] "ISO/IEC 27035 infosec incident management." Accessed: Feb. 23, 2025. [Online]. Available: https://www.iso27001security.com/html/27035.html
[51] "Understanding MITRE's 2020 ATT&CK Evaluation." Accessed: Feb. 23, 2025. [Online]. Available: https://xmcyber.com/blog/understanding-mitres-2020-attck-evaluation/
[52] "CyCraft Classroom: MITRE ATT&CK vs. Cyber Kill Chain vs. Diamond Model | CyCraft." Accessed: Feb. 23, 2025. [Online]. Available: https://www.cycraft.com/en/post/mitre20200701
[53] AirForce, "AIR FORCE DOCTRINE PUBLICATION 3-12 CYBERSPACE OPERATIONS," 2023, Accessed: Feb. 23, 2025. [Online]. Available: https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-12/3-12-AFDP-CYBERSPACE-OPS.pdf
[54] "Cyber Threat Information Sharing (CTIS) - Shared Cybersecurity Services (SCS) | CISA." Accessed: Feb. 23, 2025. [Online]. Available: https://www.cisa.gov/resources-tools/services/cyber-threat-information-sharing-ctis-shared-cybersecurity-services-scs
[55] D. of Defense, "Department of Defense Zero Trust Overlays Office of the Chief Information Officer CLEARED For Open Publication Department of Defense OFFICE OF PREPUBLICATION AND SECURITY REVIEW," 2024.
[56] R. A. Bitzinger, "Civil-Military Integration and ary Integration and Chinese Military Modernization," 2004, Accessed: Feb. 23, 2025. [Online]. Available: https://apcss.org/Publications/APSSS/Civil-MilitaryIntegration.pdf
[57] J. Yu, Y. Lu, Y. Zhang, Y. Xie, M. Cheng, and G. Yang, "A Unified Model for Chinese Cyber Threat Intelligence Flat Entity and Nested Entity Recognition," *Electronics (Switzerland)*, vol. 13, no. 21, Nov. 2024, doi: 10.3390/electronics13214329.
[58] A. de Melo e Silva, J. J. C. Gondim, R. de Oliveira Albuquerque, and L. J. G. Villalba, "A methodology to evaluate standards and platforms within cyber threat intelligence," *Future Internet*, vol. 12, no. 6, Jun. 2020, doi: 10.3390/fi12060108.
[59] "Latest misp-stix Release: Enhanced Support for Analyst Data." Accessed: Feb. 26, 2025. [Online]. Available: https://www.misp-project.org/2025/02/07/MISP_Support_for_Analyst_Data_converter_from_STIX2.html/?utm_source=chatgpt.com
[60] OASIS, "STIX Best Practices Guide Version 1.0.0," 2022. [Online]. Available: https://docs.oasis-open.org/cti/stix-bp/v1.0.0/cn01/stix-bp-v1.0.0-cn01.docx
[61] A. Ramsdale, S. Shiaeles, and N. Kolokotronis, "A comparative analysis of cyber-threat intelligence sources, formats and languages," *Electronics (Switzerland)*, vol. 9, no. 5, May 2020, doi: 10.3390/electronics9050824.
[62] L. Alevizos and M. Dekker, "Towards an AI-Enhanced Cyber Threat Intelligence Processing Pipeline," Mar. 2024.
[63] R. Fieblinger, M. T. Alam, and N. Rastogi, "Actionable Cyber Threat Intelligence using Knowledge Graphs and Large Language Models," Jun. 2024, [Online]. Available: http://arxiv.org/abs/2407.02528

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.665

Design and Preliminary Evaluation of an Immersive 3D Stereoscopic Simulation Game for Historical Education: The Hindenburg Disaster

Khairil Nazrel Bin Khairil Khusnin, Muhammad Fayyadh Bin Muhamad Rashidi, Nurazlin Zainal Azmi*
Department of Information Systems, Kulliyyah of Information & Communication Technology,
International Islamic University Malaysia, Gombak, Malaysia

*Corresponding author: nurazlinazmi@iium.edu.my
(Received: 25th November 2025; Accepted: 31st December, 2025; Published on-line: 30th January, 2026)

Abstract— History education often relies on static text and images, offering limited opportunities for experiential learning about complex historical events. This study addresses this gap by designing and examining an immersive 3D stereoscopic simulation game centered on the 1937 Hindenburg disaster. The objectives were: (i) to model a historically informed 3D replica of the Hindenburg environment, (ii) to develop an interactive gameplay experience that situates players within the unfolding event, and (iii) to conduct a pilot evaluation of usability and perceived educational value. The game was developed using Blender for asset creation and Unreal Engine 5 for implementation, following an iterative pipeline of pre-production, production, and post-production. A toggleable stereoscopic mode was integrated to enhance depth perception and immersion. The pilot evaluation was conducted with four participants using functional testing and user-acceptance feedback. Results indicated that users found the application easy to navigate, immersive, and supportive of understanding the sequence and context of the disaster, while also identifying areas for improvement such as clearer guidance and expanded interaction features. These findings provide preliminary evidence that stereoscopic serious games can serve as promising supplementary tools for historical learning and motivate future refinement and larger-scale empirical evaluation.

Keywords— serious games, stereoscopic 3D, historical simulation, immersive learning, game-based learning, pilot evaluation.

I. Introduction

History education often relies on static text, images, and teacher-centered delivery, which limits students' ability to visualize events, empathize with historical actors, and understand causal relationships. Interactive digital environments and serious games, however, have shown potential to transform historical learning from passive "learning about" to active "learning through experience," allowing learners to explore events, contexts, and consequences dynamically. The Hindenburg disaster of 1937 – a catastrophic accident that claimed 35 lives and ended the era of hydrogen-filled passenger airships – remains an important yet comparatively under-represented topic in public awareness and educational media when compared, for example, to the Titanic [1].

Prior work has demonstrated that simulation and game-based learning can improve motivation, immersion, and conceptual understanding in history education. Studies on historical strategy and role-playing games report increased engagement, critical thinking, and perspective-taking when learners interact with reconstructed historical environments rather than simply reading about them. Meanwhile, modern simulation games such as Microsoft Flight Simulator and Stormworks illustrate how realism, interactivity, and physics-based systems can meaningfully support experiential learning. However, few existing studies explore the use of stereoscopic 3D environments specifically for historical reenactment, particularly those combining narrative decision-making, immersive perspectives, and historically grounded reconstruction.

This study addresses that gap by proposing a 3D stereoscopic simulation game centered on the Hindenburg disaster. The primary aim is to recreate the historical event through an immersive stereoscopic view, enabling users to explore the airship environment and understand the sequence of events leading to the tragedy. The specific objectives of this work are:

1) to design a 3D replica of the Hindenburg airship based on historical references

https://doi.org/10.31436/ijpcc.v12i1.665

2) to develop an interactive gameplay experience that situates users within the disaster scenario; and:

3) to conduct usability-oriented user testing to examine navigation, immersion, and perceived educational value.

The main contributions of this work are threefold. First, it presents a practical design framework for integrating stereoscopic 3D technology into a serious historical game. Second, it demonstrates a prototype that translates historical narrative elements into interactive first-person gameplay. Third, it provides preliminary user-testing insights regarding usability, immersion, and learning perception, informing future development of stereoscopic educational simulations.

The remainder of this paper is organized as follows. Section II reviews related work on simulation games, historical learning, and stereoscopic environments. Section III describes the methodology and development approach, including asset modeling and gameplay design. Section IV details implementation procedure. Section V presents results and user feedback. Section VI concludes the paper and outlines directions for future work.

II. Literature Review

A. Game-Based Learning and Historical Understanding

Existing research consistently shows that game-based learning can shift history learning from passive recall to active meaning-making. Strategy and role-playing games allow learners to experiment with multiperspectivity, causality, and "what-if" scenarios, encouraging them to reason about historical events as dynamic systems rather than fixed narratives [2]-[3]. Studies further report increases in motivation, immersion, and historical empathy when learners interact with simulated historical environments rather than relying solely on texts or lectures [4]-[5].

B. Immersion, Interactivity, and Engagement

Simulation games such as Microsoft Flight Simulator and Stormworks: Build and Rescue demonstrate how realism, physics-based mechanics, and open-ended problem-solving can support experiential learning. Meanwhile, titles such as Portal (in stereoscopic configurations) illustrate the value of depth perception and spatial interaction in puzzle-based environments. Collectively, these works suggest that immersion and meaningful interaction are key to sustaining engagement and supporting deeper cognitive processing.

Table I maps key features observed in reference games to the design of our proposed system. Microsoft Flight Simulator contributes principles of realism and environmental authenticity [6]; Stormworks informs interactivity through physics-driven tasks [7]; Portal demonstrates depth-driven immersion and puzzle-based engagement [8]; and Titanic: Fall of a Legend highlights historically grounded storytelling [9]. These elements were selectively adapted not as direct replications, but as design heuristics guiding how realism, interactivity, narrative structure, and stereoscopic depth could be integrated into a unified educational simulation.

TABLE I
Game Features Adaptation Matrix

Feature / Game	Graphics	Interactivity	Storytelling	Stereoscopic
Microsoft Flight Simulator	✓	✓		
Stormworks		✓		
Portal (3D)	✓	✓	✓	✓
Titanic: Fall of a Legend	✓	✓	✓	

C. Historical Simulation and Narrative Experience

Historical exploration games, including Titanic: Fall of a Legend, show how reconstructed environments and narrative progression can help players connect emotionally with past events. However, many such systems rely on monoscopic visuals and largely linear storytelling, offering limited embodied experience or decision-based exploration.

D. Identified Gap and Design Rationale

While prior studies support the educational value of serious games, few works explicitly combine:

(i) historically grounded environments,
(ii) interactive narrative branching, and
(iii) stereoscopic 3D immersion [3]-[5], [10].

This gap shapes the design rationale of the Hindenburg simulation, which integrates first-person perspective, decision-based interaction, and stereoscopic depth to encourage users to reason about events while experiencing them from within the scenario. The review therefore establishes the foundation for our design choices: realism and interactivity to promote engagement, narrative framing to support understanding, and stereoscopic rendering to amplify immersion and presence.

III. Methodology

This project followed a structured three-phase game-development pipeline consisting of pre-production, production, and post-production activities (Fig. 1). This structure ensured systematic planning, technically grounded implementation, and iterative refinement aligned with the project objectives.

https://doi.org/10.31436/ijpcc.v12i1.665

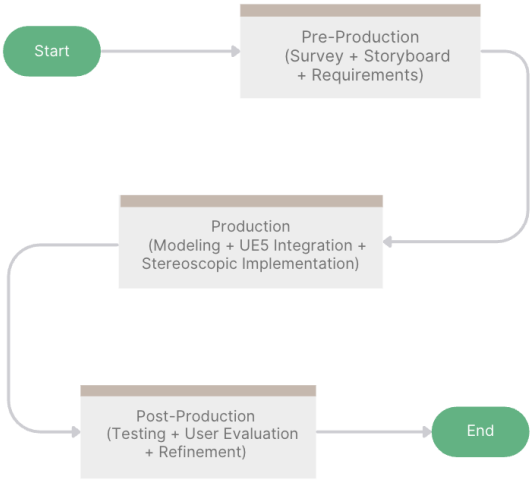


Fig. 1 Overall development methodology

A. Pre-Production

The pre-production phase focused on understanding user expectations and defining design requirements. A short online survey (n = 22) was distributed to undergraduate students via Google Forms. The survey consisted of multiple-choice questions covering:

i. prior experience with simulation and historical games,
ii. interest level in history-focused gameplay,
iii. perceived usefulness of immersive 3D environments for learning, and
iv. willingness to use stereoscopic 3D features.

Results indicated strong interest in historical simulations (95.5%) and positive attitudes toward stereoscopic viewing (72.7%), supporting the decision to proceed with an immersive, narrative-driven design. These findings were used to refine gameplay scope, interaction complexity, and visual presentation.

Narrative elements, character roles, and scene layouts were storyboarded using Adobe Fresco. Assets, environments, and interaction flows were defined at this stage to ensure coherence between educational intent and gameplay elements.

B. Production

During the production phase, all 3D models – including characters, interior furnishings, and the Zeppelin structure – were created in Blender and textured for visual realism. Characters included four roles (Passenger, Ship Crew, Engineer, and Bystander) designed according to 1930s references [11]. Models were rigged and exported to Unreal Engine 5.

Within Unreal Engine, one complete storyline from the passenger perspective was implemented. Core mechanics included first-person navigation, object interaction, triggered dialogue, and scripted event sequences. A stereoscopic system was developed using blueprint scripts enabling toggling between normal and stereoscopic mode, adjusting field-of-view, and saving player preferences.

C. Post-Production

The post-production phase focused on testing and evaluation. Functional testing ensured correct navigation, interaction behavior, and stability across scenes. User acceptance testing was conducted with four participants from mixed technical backgrounds. Participants were asked to complete a guided gameplay session and provide feedback on ease of navigation, clarity of content, immersion, and perceived educational value.

Observations and questionnaire responses were analyzed descriptively. Feedback highlighted strengths in realism and immersion, while suggesting clearer tutorials and extended interactive content.

D. Narrative Structure

The system was designed to support dual perspectives: a passenger attempting to escape the disaster and a bystander assisting in rescue coordination. Although only the passenger storyline was fully implemented due to scope constraints, both perspectives were planned to encourage critical reflection on different lived experiences of the same historical event.

IV. Implementation

A. Application Architecture and Development Framework

The system was implemented as an immersive educational simulation structured around three tightly integrated layers: content creation, interaction and narrative control, and stereoscopic presentation. This architecture ensures that historical fidelity, player agency, and perceptual immersion work together to support experiential learning.

Blender was used to construct historically grounded three-dimensional assets, including the airship interior, environmental props, and character models (Fig. 2 – Fig. 4). These assets form the spatial and visual context that allows users to explore the Hindenburg as a lived environment rather than a static reconstruction. Unreal Engine 5 served as the primary runtime platform, managing real-time rendering, interaction logic, and stereoscopic visualization. Adobe After Effects was employed to create cinematic sequences and animated overlays that introduce historical context and support narrative flow.

The use of this three-tool pipeline was intentional: Blender provides modeling accuracy and flexibility, Unreal Engine enables interactive and immersive simulation, and After Effects allows controlled narrative framing. Together,

https://doi.org/10.31436/ijpcc.v12i1.665

they support both technical realism and pedagogical clarity, which are essential for serious games in historical education.

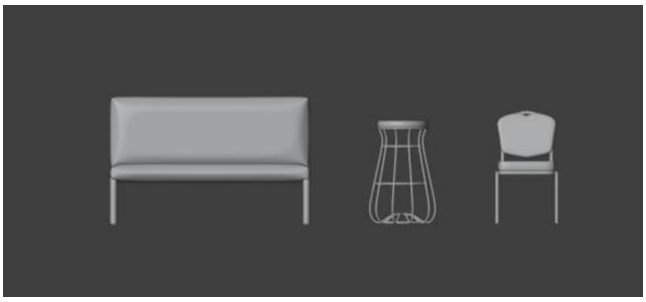


Fig. 2 Chair models

Fig. 3 The passenger model

Fig. 4 Zeppelin airship model

B. *Interactive and Narrative Design*

Player interaction was designed to promote agency, engagement, and causal reasoning, which are core to experiential learning. Within Unreal Engine, blueprints were used to implement movement, object interaction, triggered events, and environmental navigation. These mechanics allow players to actively explore the airship, respond to hazards, and progress through the unfolding disaster scenario.

The game was structured around a first-person perspective to enhance embodiment and situational awareness. Although the full dual-perspective design (passenger and bystander) was planned, the passenger storyline was fully implemented and used for evaluation. This perspective places players inside the airship, where they must navigate physical space and make decisions under time pressure, reinforcing the cognitive and emotional dimensions of the historical event.

C. *Stereoscopic Rendering and Usability Design*

A key feature of the system is its toggleable stereoscopic 3D mode, which enhances depth perception and spatial realism. Unreal Engine renders dual viewpoints corresponding to the left and right eyes, producing a three-dimensional visual experience when viewed with compatible 3D glasses.

To balance immersion with comfort and accessibility, several usability-oriented features were implemented:

- Adjustable field-of-view (FOV) allows users to modify perspective to reduce eye strain and improve visual comfort (Fig. 5).
- Toggleable stereoscopic mode enables users to switch between stereoscopic and standard rendering depending on preference or hardware capability (Fig. 6).
- Persistent settings allow stereoscopic preferences to be saved and loaded across sessions, preventing repetitive reconfiguration.

These features ensure that immersion does not come at the expense of usability, which is especially important in educational contexts where cognitive load must be carefully managed.

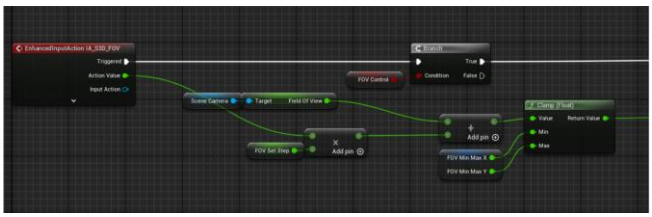


Fig. 5 Blueprint for the FOV slider adjustment

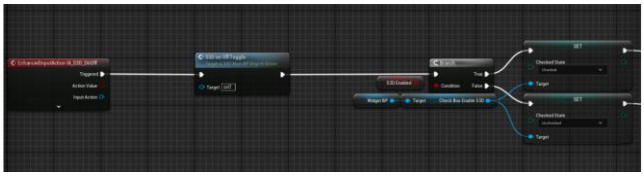


Fig. 6 On and off blueprint for the stereoscopic option

D. *Media Integration and Scene Composition*

To support narrative comprehension and historical context, 2D and 3D media elements were integrated into the gameplay experience. Adobe After Effects was used to create animated sequences such as the Hindenburg introduction, location reveal, and newspaper highlights (Fig. 7 – Fig. 8). These scenes provide historical framing before

https://doi.org/10.31436/ijpcc.v12i1.665

and during gameplay, guiding players' understanding of the event.

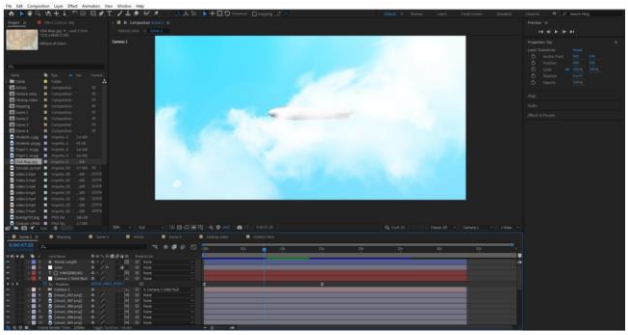


Fig. 7 Introduction to Hindenburg composition

Fig. 8 The Hindenburg disaster newspaper highlight [12]

Fig. 9 Passenger lounge

Fig. 10 Smoking room

Within Unreal Engine, 3D environments – including passenger lounges, smoking rooms, and airship interiors – were assembled with lighting, textures, and physics effects to simulate the atmosphere of the 1930s Zeppelin (Fig. 9 – Fig. 10). These environments serve not merely as visual backdrops but as interactive learning spaces in which players observe, navigate, and react to unfolding events.

E. Audio and Multimodal Integration

Audio was treated as a critical component of immersion and learning. Ambient sounds, environmental effects, and era-appropriate music were layered to create a believable soundscape. Narration was generated using AI-based text-to-speech tools (ElevenLabs and TTSMaker) and synchronized with cutscenes and gameplay to reinforce storytelling and guide attention.

The coordinated integration of audio, visual, and interactive elements creates a multimodal learning environment in which users process historical information through sight, sound, and action rather than through text alone.

V. Testing and Evaluation

This evaluation was conducted as a pilot study to assess system functionality, usability, and perceived educational value prior to any large-scale deployment. The objective was not to establish statistically generalizable learning outcomes, but to determine whether the system operates as intended, provides a coherent user experience, and supports immersive engagement with historical content.

The testing activities were structured into four complementary components.

A. Functional Validation

Core interface elements, including navigation controls, menu systems, and character selection features, were tested to ensure reliable and intuitive operation. Particular attention was given to primary interaction elements such as the Play button, avatar selection screen, and in-game prompts to confirm that users could progress through the application without confusion or error.

B. Multimedia Synchronization

The application's audiovisual components were evaluated to ensure smooth playback and correct synchronization between video, audio, and stereoscopic rendering. This included verification of cutscene timing, narration alignment, and responsiveness of stereoscopic adjustments, ensuring that multimedia elements did not disrupt immersion or comprehension.

https://doi.org/10.31436/ijpcc.v12i1.665

C. *Gameplay Mechanics*

Player movement, object interaction, and environmental triggers were tested to confirm accurate system response to user input. These tests ensured that navigation within the airship, interaction with key objects, and progression through scripted events functioned consistently and as designed.

D. *Immersive Experience and Visual Fidelity*

Immersion was evaluated by examining the stability and consistency of stereoscopic 3D rendering across different scenes and gameplay states. Loading accuracy, depth perception, and persistence of user-defined stereoscopic settings were verified to ensure that the immersive experience remained coherent throughout gameplay.

E. *Pilot User Evaluation*

User acceptance testing was conducted with four participants from mixed backgrounds (three students and one engineer). Participants completed a guided gameplay session followed by structured feedback on navigation, content clarity, immersion, and overall experience. A test plan covering navigation, multimedia playback, gameplay mechanics, and stereoscopic functionality was executed, with all core system functions achieving a 100% pass rate (Fig. 11).

	A	B	C	D	E	F	G
5	Test Case	Step	Test Steps	Expected Result	Actual Result	Pass/Fail	Note
6	Functional Test Cases						
7		1	Click on the "PLAY" button.	Player navigates to the opening video.		Pass	Home Screen
8	Test Case ID: TC_001 Test Title : Verifying the button functionalities in Home Screen	2	Click on the "OPTION" button.	Option menu pop up.		Pass	Home Screen
9		3	Player change the setting accordingly to their preferences and click on the "APPLY" button.	The game will update to the new setting.		Passs	Home Screen
10		4	Player click on the "BACK" button.	Player will be redirected to the main menu.		Pass	Home Screen
11		5	Click on the "EXIT" button.	Game exits successfully.		Pass	Home Screen
12	Test Case ID: TC_002 Test Title : Testing the Opening Video	1	Opening video played automatically	The opening video played with sound.		Pass	Opening Video
13		2	Click on the "Skip/Play" button.	The opening video is skipped and directed to player selection menu.		Pass	Opening Video
14	Test Case ID: TC_003 Test Title : Verifying the Character Selection Screen	1	Verify character selection screen loads.	Character selection screen loads with 2 options.		Pass	Character Selection
15		2	Click on Passenger Experience.	Player will be teleported to passenger level.		Pass	Character Selection
16		3	Click on Bystander Experience.	Player will be teleported to bystander level.		Pass	Character Selection Screen
17	Test Case ID: TC_004 Test Title : Testing Character Movement with Keyboard and Mouse Control	1	Press W to move the character forward.	The character moves forward when W is pressed.		Pass	In Game
18		2	Press S to move the character backward.	The character moves backward when S is pressed.		Pass	In Game
19		3	Press A to move the character left.	The character moves left when A is pressed.		Pass	In Game
20		4	Press D to move the character right.	The character moves right when D is pressed.		Pass	In Game
21		5	Press spacebar to make the character. jump	The character jumps when spacebar is pressed.		Pass	In Game
22		6	Press Ctrl to crouch the character.	The character crouches when Ctrl is pressed.		Pass	In Game
23		5	Move the mouse left, right, up, and down to look around.	The camera view changes smoothly when the mouse is moved.		Pass	In Game
24		6	Combine keyboard and mouse movements simultaneously.	The character moves and looks around simultaneously without lag or unexpected behavior.		Pass	In Game
25	Test Case ID: TC_005 Test Title : Testing 3D Stereoscopic Mode	1	3D stereoscopic settings automatically popup on each level.	Player will be able to change the setting accordingly.		Pass	in Game
26		2	Click on "Save" button.	Player can save the stereoscopic setting.		Pass	In Game
27		3	Click on "Load" button.	Player load the stereoscopic setting from previous save.		Pass	In Game
28		4	Click on small circle button on the top right.	The 3D stereoscopic settings popup disappear.		Pass	In Game

Fig. 11 Screenshot of selected functional test cases

Although the sample size was limited, the pilot provided valuable insight into usability and perceived learning value. Participants reported high levels of immersion and generally strong understanding of the historical content, while also identifying areas for refinement. Suggested enhancements included clearer onboarding and tutorials, richer gameplay mechanics for experienced users, and expanded stereoscopic features, as summarized in Table II.

TABLE III
SUMMARY OF PARTICIPANT FEEDBACK

Participant Background	Ease of Navigation	Understanding of Content	Key Suggestion
BIT Student (Novice Gamer)	Excellent	Good	None
Engineer (Hardcore Gamer)	Excellent	Excellent	Add advanced gameplay mechanics
BIT Student (Novice Gamer)	Good	Good	Simpler tutorial needed
BIT Student (Gamer)	Excellent	Excellent	Expand stereoscopic features

VI. CONCLUSION

This study presented the design, implementation, and pilot evaluation of a 3D stereoscopic simulation game for historical education, using the Hindenburg disaster as a case study. The results of the pilot study indicate that immersive, interactive environments have strong potential to support historical understanding by allowing learners to experience events from within a reconstructed context rather than through static representations alone. Participants reported high levels of engagement, clear navigation, and meaningful interaction with the historical content, suggesting that the system provides a viable foundation for experiential history learning.

Rather than making definitive claims about learning effectiveness, this work contributes a proof-of-concept demonstrating how stereoscopic rendering, narrative-driven gameplay, and first-person interaction can be integrated into a serious game for history education. The findings from the pilot study offer early evidence that such systems are usable, immersive, and educationally promising, while also identifying practical areas for refinement.

Despite its successful implementation of a fully functional passenger-level experience and stereoscopic system, the project was constrained by hardware limitations and the complexity of Unreal Engine 5, which restricted the completion of the planned bystander perspective and limited the depth of visual effects.

Future work will therefore focus on:

- completing the bystander perspective to support multi-viewpoint historical reasoning;

- refining stereoscopic rendering for greater visual comfort and depth perception;
- optimizing performance for a wider range of devices, including mobile platforms; and
- replacing placeholder assets with custom-designed models to improve visual coherence and historical authenticity.

Overall, this pilot study establishes a solid technical and pedagogical foundation for future large-scale evaluation of stereoscopic serious games as tools for immersive historical learning.

ACKNOWLEDGMENT

First and foremost, we extend our deepest gratitude to Allah, whose endless blessings and guidance have allowed us to complete this project successfully.

We would like to express our heartfelt appreciation to our supervisor, Dr. Nurazlin Zainal Azmi, for her invaluable guidance, constructive feedback, and unwavering support throughout the process. Her encouragement and insights have been instrumental in shaping this work.

Special thanks are due to all our lecturers, for their dedication, knowledge, and mentorship, which have greatly contributed to our academic journey. Lastly, we would like to express our sincere gratitude to all our dear friends for their support, kindness, and companionship, which made this journey memorable and fulfilling.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] D. Grossman, "The Hindenburg Disaster," Airships.net, Jun. 28, 2019. [Online]. Available: https://www.airships.net/hindenburg/disaster/ [Accessed: Jan. 7, 2026].

[2] L. Radetich and E. Jakubowicz, "Using Video Games for Teaching History. Experiences and Challenges," Athens J. Hist., vol. X, no. Y, pp. 1–14, 2014.

[3] A. Hellerstedt and P. Mozelius, "Game-based Learning for History: Student Perceptions and Preferences," in Proc. Irish Conf. Game-Based Learn., Cork, Ireland, 2024, pp. 1–7.

[4] B. D. Redder, "Revitalisation of History through Historical Games in the Digital Era: An Opening Provocation into Teaching History through Multimodality," in Revitalising Higher Education: Insights from Te Puna Aurei LearnFest 2022, T. Bowell, N. Pepperell, A. Richardson, and M.-T. Corino, Eds. Cardiff, U.K.: Cardiff Univ. Press, 2024, pp. 71–79.

[5] G. P. Kusuma, L. K. P. Suryapranata, E. K. Wigati, and Y. Utomo, "Enhancing Historical Learning Using Role-Playing Game on Mobile Platform," Procedia Comput. Sci., vol. 179, pp. 886–893, 2021.

[6] Asobo Studio, "Microsoft Flight Simulator," 2020. [Online]. Available: https://www.asobostudio.com/games/microsoft-flight-simulator. [Accessed: Jan. 9, 2026].

[7] Geometa, "Stormworks: Build & Rescue," 2018. [Online]. Available: https://geometa.co.uk/stormworks. [Accessed: Jan. 9, 2026].

[8] N. Schneider, "Portal Stereoscopic 3D Review," MTBS3D – Meant To Be Seen 3D, Jun. 7, 2010. [Online]. Available: https://www.mtbs3d.com/articles/game-reviews/11439-portal-stereoscopic-3d-review.html. [Accessed: Jan. 9, 2026].

[9] Interactive Gaming Studios, "Titanic: Fall of a Legend" [Video Game]. Steam, 2022. [Online]. Available: https://store.steampowered.com/app/1835200/Titanic_Fall_Of_A_Legend/. [Accessed: Jan. 9, 2026].

[10] G. Inan Kaya and E. Sar Ijbilen, "Educational use of games: A mobile serious game for history education," in Edu World 2016 7th International Conference, 2017, pp. 1001–1008

[11] Schneider, S. R. (2021). What Men REALLY Wore in the 1930s. Gentleman's Gazette.

[12] The Brick & Maple, "Hindenburg," The Brick & Maple, accessed on Dec. 27, 2025. [Online]. Available: https://thebrickandmaple.com/tag/hindenburg/.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.639

Comparative Evaluation of Lightweight CNN and YOLOv8 Models for Brain Tumor Detection in Resource-Constrained Settings

[1] *Ahsiah Ismail, [2]Mohd Yamani Idna Idris
[1]Department of Computer Science, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia
[2]Department of Computer System and Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

*Corresponding author: ahsiah@iium.edu.my
(Received: 27th November 2025; Accepted: 12th January 2026; Published on-line: 30th January, 2026)

Abstract—Brain tumor detection is essential for timely diagnosis, early intervention, and effective treatment planning. With advancements in artificial intelligence (AI), deep learning methods have emerged as powerful tools in medical imaging, offering automated, consistent, and efficient detection of brain abnormalities. However, achieving clinically reliable performance requires models that can accurately differentiate between tumor and non-tumor cases. This paper investigates and compares the performance of two deep learning models which are a lightweight Convolutional Neural Network (CNN) and the You Only Look Once (YOLO) YOLOv8 model for brain tumor classification in resource-constrained setting. Both models were trained and evaluated using the BR35H dataset, which comprises 3,000 MRI scans categorized into tumor and non-tumor classes. The performance of the models are evaluated using accuracy, precision, recall, F1-score as well as inference time supplemented by confusion matrix, ROC analysis and Grad-CAM visualizations to assess class-wise prediction performance. The experimental results indicate that YOLOv8 demonstrated high predictions across both tumor and non-tumor categories. YOLOv8 outperformed the CNN, achieving an accuracy of 0.998, precision of 0.997, recall of 1.00, and an F1-score of 0.998. However, only a minimal difference was observed in the inference time per image between YOLOv8 and the CNN, with YOLOv8 being slower by just 10.6 ms. Finally, the results demonstrate YOLOv8's robustness and reliability for early tumor detection, a critical factor in preventing diagnostic delays. The findings further highlight YOLOv8's suitability for integration into clinical decision-support systems, particularly in resource-constrained environments where accurate and fast automated diagnosis can significantly enhance patient care.

Keywords— deep learning, CNN, YOLO, Brain tumor, detection

I. Introduction

Brain tumors remain a critical global health concern, and early diagnosis is essential for improving treatment outcomes and patient survival. Magnetic Resonance Imaging (MRI) is the most widely used modality for detecting brain abnormalities due to its high contrast resolution and non-invasive nature. MRI is the standard non-invasive imaging modality for identifying brain tumors due to its superior soft tissue contrast and high-resolution capabilities, which allow clinicians to visualize tumor morphology, size and location. However, in many clinical settings, particularly in developing countries, the number of trained radiologists or experts is insufficient to meet the increasing demand for diagnostic imaging[1][2]. This shortage contributes to a heavy workload and prolongs the detection process, as manual interpretation of MRI scans is inherently time-consuming, potentially causing delays in intervention and treatment planning. Brain tumors are among the most life-threatening neurological disorders worldwide, often requiring prompt and accurate diagnosis to guide effective treatment and improve patient outcomes[3][4][5]. Consequently, brain tumor detection and classification using MRI scans are critical tasks in medical diagnostics, significantly influencing patient outcomes through timely and precise diagnosis.

Advancements in deep learning, particularly Convolutional Neural Networks (CNNs) and You Only Look Once (YOLO) frameworks, have revolutionized the field of medical image analysis, offering promising solutions for the automated detection and classification of brain tumors[6][7][8][9][10]. CNNs are the deep learning models designed to process and analyze visual data. The model excels in tasks such as object detection, image segmentation, and classification by automatically learning hierarchical features from the input images. CNNs have been widely used in brain tumor detection due to their ability to accurately identify and segment tumor regions in MRI

https://doi.org/10.31436/ijpcc.v12i1.639

scans[7][11]. On the other hand, YOLO is an object detection framework designed to perform both localization and classification within a single unified architecture, enabling the model not only to identify the presence of a tumor but also to precisely localize its region within an MRI slice. Its deeper and optimized architecture enhances feature representation, improving sensitivity to subtle tumor boundaries and small-scale abnormalities.

Given the urgent need for accurate and timely tumor detection to improve treatment outcomes and survival rates, this research investigates the detection performance of these two CNN and YOLO models for brain tumor classification. Experiments were conducted using the BR35H brain tumor MRI dataset to evaluate both a lightweight CNN and the YOLOv8 model. YOLOv8 was selected for its advanced feature extraction and robust detection capabilities, while the lightweight CNN serves as a baseline due to its simplicity and controlled architectural design. By comparing these two models, the research provides a focused analysis of performance, robustness and clinical applicability, particularly in resource-constrained clinical environments. Limiting the evaluation to these two models allows for a focused and practical investigation. By analyzing their strengths and limitations under identical experimental conditions, this research provides meaningful insights into selecting suitable AI architectures for clinical deployment. The findings contribute to the development of efficient, accurate, and explainable computer-aided diagnosis (CAD) systems for brain tumor detection. The contributions of this paper are as follows:

- Conducts a focused evaluation in which a lightweight CNN serves as a conventional, low-complexity feature-extraction classifier, while YOLOv8 represents a detection-oriented, multi-scale architecture.
- Provides a comparative evaluation of CNN and YOLOv8 for binary brain tumor detection using MRI scans, highlighting their practical applicability in clinical and resource-constrained settings, as well as their suitability for deployment in diverse healthcare environments.
- Analyzes detection performance of both models across tumor and non-tumor classes to quantify diagnostic accuracy, while employing explainable AI techniques to evaluate model interpretability.

The remainder of this paper is structured as follows: Section II presents related work on brain tumor detection. Section III discusses the methodology employed in this research. The experimental setup is described in Section IV, followed by results and discussion in Section V. Finally, conclusions and future work are presented in Section VI.

II. Related work

Deep learning approaches have been widely applied to brain tumor detection and classification using MRI images. The application of deep learning techniques for brain tumor classification has evolved significantly over the past few years. Early research for brain tumor classification focused primarily on CNN-based architectures. CNN-based methods have been widely used for brain tumor classification due to their strong feature extraction capabilities. Sajjad et. al proposed a multi-grade brain tumor classification framework using CNNs combined with image enhancement techniques, reporting strong performance across glioma and meningioma detection tasks[12]. Similarly, Deepak et. al employed transfer learning on pre-trained CNN, demonstrating that CNN architectures can effectively capture hierarchical tumor features from MRI scans[13]. O the other hand, Daniel Reyes et al. compared various CNN architectures including VGG, ResNet, EfficientNet, and ConvNeXt for brain tumor classification [14]. In their work, the CNN model able to achieve promising results with the best model reaching 98.7% accuracy on datasets containing over 3000 images of gliomas, meningiomas, and pituitary tumors. The research demonstrated that ResNet, MobileNet, and EfficientNet were the most accurate networks, with MobileNet and EfficientNet showing superior performance in terms of computational complexity. CNNs have been the backbone of many automated brain tumor diagnosis systems due to their ability to extract hierarchical spatial features and achieve high accuracy on medical images. CNNs have been widely used for brain tumor detection due to their robust feature extraction capabilities. The standard CNN classification models also able to provide a probabilistic output without directly indicating the tumor's spatial location. Various studies have demonstrated the effectiveness of CNNs in classifying and detecting brain tumors from MRI scans and have shown promising results in terms of accuracy, precision, and recall metrics particularly for limited labeled MRI data[15]. While CNNs are computationally efficient and suitable for resource-constrained environments, their relatively shallow feature extraction pipelines can limit performance on complex or heterogeneous tumor appearances.

CNNs have been extensively used for brain tumor detection due to their ability to automatically extract hierarchical features from medical images. Various CNN architectures, such as EfficientNet, ResNet, and VGG-16, have demonstrated high accuracy in tumor detection tasks[16][17][18]. EfficientNet able to achieved a promising accuracy by optimizing depth, width, and resolution

simultaneously, making it both accurate and computationally efficient[19]. However, CNNs often struggle with capturing long-range dependencies and contextual information, which are crucial for accurate tumor classification[20]. Transformers, particularly Vision Transformers (ViTs), have been introduced to address the limitations of CNNs. ViTs use the self-attention mechanisms to capture both local and global features, significantly improving the model's ability to understand complex spatial relationships in medical images[21]. ViTs model able to achieve high accuracy in classification task. Jia et al. proposed a transformer-augmented deep learning model for tumor detection under cystoscopy, demonstrating improved feature representation and classification accuracy[22]. Despite their effectiveness, ViTs can be computationally intensive and require large datasets for training[23].

Hybrid models have also been developed typically using CNNs for initial feature extraction and Transformers for capturing long-range dependencies. Jaffar et. al proposed a hybrid model combining ResNet50 and a Transformer encoder demonstrated superior performance in classifying brain tumors, achieving an accuracy of 99.2%[20]. While Sankari et. al. integrated CNNs with ViTs and achieved a precision of 98.7%, outperforming both standalone CNN and Transformer models[24]. Hybrid models have consistently shown higher accuracy compared to standalone CNN or Transformer models. For instance, a CNN-ViT hybrid achieved an accuracy of 98%, surpassing the standalone models[25]. By combining local feature extraction of CNNs with the global context modeling of Transformers, hybrid models provide a more comprehensive feature representation[26]. These models have demonstrated better generalization to unseen data, making them more reliable for clinical applications[26]. While these approaches achieve strong performance, Hybrid models can be computationally demanding, requiring significant processing power and memory[23][24]. In addition, large and diverse datasets are often required to train these models effectively, which can be a limitation in medical imaging where data is scarce[23]. Despite their accuracy, integrating these models into existing clinical workflows poses challenges especially related to computational complexity and data requirements, which may limit practical deployment in clinical settings particularly in resource-constrained environments.

In contrast, the YOLO models balance accuracy and efficiency where the models reduce computational complexity while maintaining high detection accuracy[27]. YOLO models represents a class of single-stage object detectors designed to prioritize speed while maintaining competitive accuracy. This architecture simultaneously predicts bounding boxes and class probabilities, making it well suited for real-time analysis where detection latency is an important performance factor[28]. The detection architectures of a YOLO model have been progressively developed. Their model architectures have marked a significant advancement in many classifications task including brain tumor detection and classification. In YOLOv1 model, a unified detection framework is introduced in the model where it enabled real-time object detection[28]. On the other hand, the YOLOv2 and YOLOv3 model improved feature extraction, multi-scale detection, and training stability[29]. YOLOv4 model further enhanced accuracy using advanced backbone networks and data augmentation strategies[30]. While YOLOv5 and YOLOv6 models continued the trend with improved lightweight architectures for deployment efficiency. YOLOv5 model was adopted by Fayez Ghufran et al. for brain tumor identification and classification task. They compared the performance of YOLOv5 model with Faster R-CNN[31].In their work, they proposed 9-layer CNN and able to achieved an accuracy of 98.21%, highlighting the potential of YOLO-based model to support real-time tumor detection with lower computational requirements compared to earlier models. The YOLO models continued to evolve, with each version introducing architectural improvements to enhance detection performance. The YOLOv7 model incorporated extended efficient layers and model re-parameterization strategies to maximize detection performance[32]. While in YOLOv8, significant architectural improvements over previous YOLO variants being introduced, including decoupled heads for classification and localization, a more efficient CSPDarknet-based backbone, and enhanced training optimization strategies[33]. In YOLOv8 model, it also includes a dedicated classification mode (YOLOv8-cls), enabling the framework to operate not only as a detector but also as a high-performance image classifier. This dual capability makes the YOLOv8 model particularly well-suited for medical image analysis tasks that demand both accuracy and efficiency. Its optimized inference pipeline, reduced computational complexity, and superior generalization performance make it an appealing choice for brain tumor classification compared with earlier YOLO versions. Aniket Prabhu Taradale et al. specifically focused on YOLOv8 for brain tumor segmentation and classification, achieving promising performance metrics including 95% F1-Score, 96.20% precision, 93.6% recall, and 97.2% mAP50[34]. Their work emphasized YOLO's real-time detection capabilities, representing a significant improvement over segmentation method. On the other hand, Zougari et al. reported a detection accuracy of 0.85 using YOLOv8 on a dataset of 1101 MRI images[35]. Similarly Verma et al. also reported that YOLOv8 demonstrated superior performance in brain tumor

https://doi.org/10.31436/ijpcc.v12i1.639

classification, successfully distinguishing glioma, meningioma, and pituitary tumors with an accuracy of 99.12%[36]. YOLOv8 model fine-tuned for brain tumor detection, achieves competitive accuracy and efficiency. Its balanced performance between precision and computational efficiency makes it well-suited for deployment in clinical environments, where both reliability and processing speed are critical[37].

Despite YOLOv8 provides advanced performance, CNN models remain essential as it offers a simpler architectural design, allow direct control over feature extraction layers, and are widely adopted in medical imaging. These characteristics make CNNs an ideal baseline model for comparative evaluation, allowing a clear assessment of performance improvements offered by more advanced architectures such as YOLOv8. Both CNN and YOLO-based models have shown strong potential in brain tumor detection from MRI images. CNN particularly effective at hierarchical feature extraction and often achieve high classification accuracy, although they tend to be computationally intensive. In contrast, YOLOv8 offers real-time detection capabilities and highly efficient computation, attributes that are critical for clinical environments requiring rapid decision support. YOLOv8 able to maintains competitive and often superior accuracy across various classification task[9][36].

Although CNN and YOLO-based models have shown strong performance in brain tumor detection and classification, there are limited studies that directly compare these two approaches under a unified experimental setup. This research conducts a direct comparison of a lightweight CNN and YOLOv8 using the BR35H dataset for binary tumor classification, explicitly considering resource-constrained environments and using identical training hyperparameters and evaluation protocols. Performance is assessed not only through accuracy, precision, recall, F1-score including inference time but also with confusion matrix analysis, ROC curves, and Grad-CAM visualizations, providing both quantitative and interpretable insights. These enables a clear understanding of the models' relative performance, suitability for deployment in low-resource settings, and reliability for clinical decision support, particularly in minimizing false negatives that could delay diagnosis and treatment. By explicitly controlling dataset, task type, model version, and evaluation setup, this reseaerch provides an insight for future research in clinically relevant, resource-constrained brain tumor detection and classification.

III. METHODOLOGY

This research implements an experimental workflow for brain tumor detection using two deep learning models which are CNN and YOLOv8 classification model. The workflow is illustrated in Fig. 1. which outlines the sequential stages of dataset preparation, model development, hyperparameter optimization and performance evaluation. Each of these steps will be discussed in the following subsections.

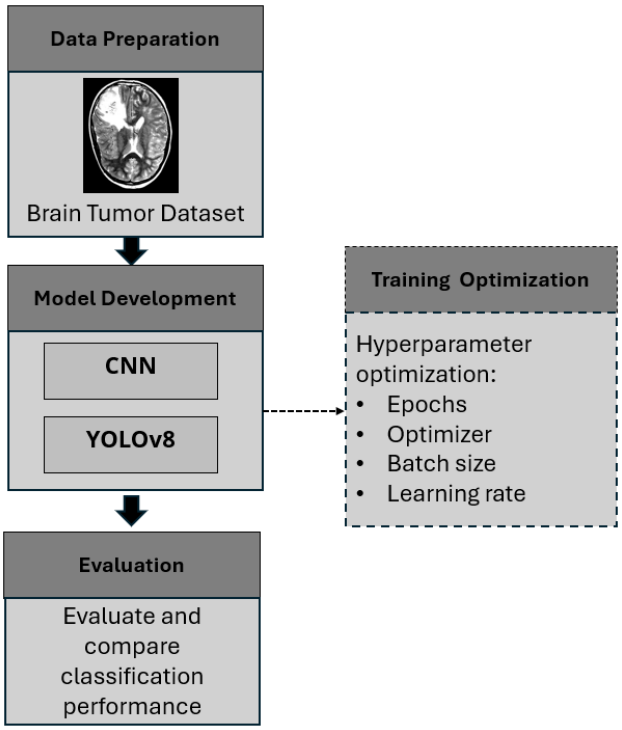


Fig. 1. Workflow of brain tumor detection using two deep learning models: CNN and YOLOv8.

A. *Data Preparation*

In this research, the BR35H dataset[38], *comprising 3,000 MRI brain images evenly divided into tumor and no-tumor classes, was employed due to its suitability for binary brain tumor detection. Its balanced class distribution minimizes bias from class imbalance, ensuring fair and reliable training and evaluation of both CNN-based classifiers and detection-oriented models such as YOLOv8. In addition, BR35H's standardized structure and adoption in prior research enable clinically meaningful and unbiased performance comparisons across different deep learning architectures.*

B. *Model Development*

To identify the most suitable model for brain tumor classification, two deep learning models namely, CNN and YOLOv8 were developed. Each model is described in detail in the following subsections.

I) Convolutional Neural Network (CNN): The CNN model architecture employs a convolutional neural

https://doi.org/10.31436/ijpcc.v12i1.639

network as both the feature extractor and classifier, as illustrated Fig. 2. The CNN processes each MRI scan through a sequence of structured layers, including the input layer, convolutional layers, pooling layers, and fully connected layers. Each layer performs a specific computational transformation that converts the MRI image into a high-dimensional feature representation suitable for tumor classification. Three convolutional layers are used in the CNN architecture, each with progressively increasing depth to capture more complex spatial patterns at successive levels of abstraction. The first convolutional layer is responsible for detecting low-level features such as edges, contours, and basic intensity gradients within the MRI scan. The second convolutional layer extracts mid-level visual patterns, including textural differences and structural variations commonly associated with tumor regions. The third convolutional layer captures high-level, more abstract features, such as the overall shape, size, and spatial relationships of tumor regions, which are critical for accurate classification and localization. Each convolutional layer utilizes a set of learnable kernels that slide across the MRI image to detect localized spatial features. These filters generate feature maps by computing weighted dot products, enabling the network to encode spatial irregularities, lesion boundaries, and tumor shapes. Following each convolutional layer, a corresponding max-pooling layer is applied to reduce spatial resolution, retain dominant features, and decrease computational complexity. The pooling layers also help introduce translation invariance, enabling the network to detect tumor features regardless of their precise position within the MRI. After feature extraction, the flattened output is passed into fully connected layers that learn global feature interactions. The final classification is performed, which outputs the probability distribution across the two classes which are tumor and no tumor.

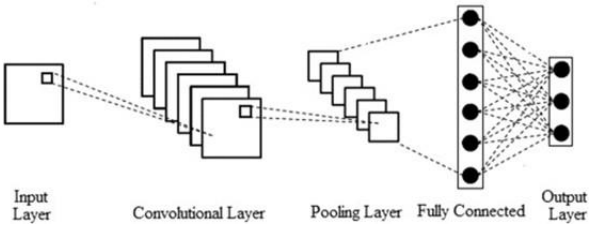


Fig. 2. CNN model Architecture[39]

2) *YOLOv8*: The YOLOv8 architecture consists of three main components which are the backbone, neck, and detection head. Each playing a critical role in feature extraction, multi-scale feature aggregation, and final tumor prediction. The backbone extracts hierarchical features from MRI images through a series of convolutional and C2f blocks. The stem layer uses a 3×3 convolutional kernel to reduce computational cost, while the C2f blocks efficiently process features at multiple pyramid levels (P1–P5). Low-level features (P1) capture fine details, while higher-level features (P5) encode semantic information, allowing detection of tumors of varying sizes and shapes. The neck aggregates multi-scale features from the backbone to improve detection accuracy. Using a feature pyramid approach, it concatenates features across pyramid levels (P1–P5) without requiring identical channel dimensions, reducing parameter count and tensor size while maintaining robust multi-scale tumor detection. The SPPF module is also applied to capture contextual information across varying receptive fields. The detection head receives the fused features from the neck and predicts tumor centers, bounding box dimensions, and class probabilities in an end-to-end manner. This eliminates the need for anchor boxes, which are commonly required in CNN-based models. The architectural structure of YOLOv8 is illustrated in Fig. 3

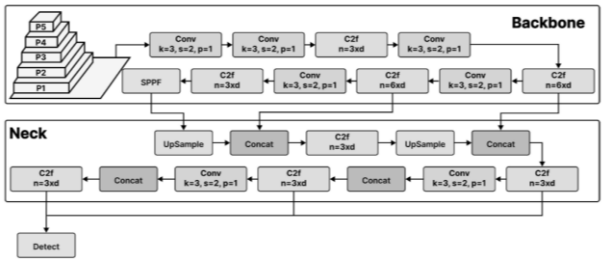


Fig. 3. YOLOv8 architecture structure[40]

Unlike CNNs, which are effective for feature extraction and classification but may struggle with tumors of irregular shapes or varying sizes, YOLOv8 performs end-to-end detection, simultaneously localizing and classifying tumors. This capability leads to improved detection accuracy. YOLOv8 directly predicts object centers using a center-based detection strategy, enabling precise localization of irregularly shaped tumors and enhancing generalization across diverse tumor morphologies. These features make YOLOv8

https://doi.org/10.31436/ijpcc.v12i1.639

particularly suitable for brain tumor detection tasks where both accurate classification and precise localization are critical. In addition, YOLOv8 incorporates the Spatial Pyramid Pooling Fast (SPPF) module to capture features at multiple scales, which is important for detecting tumors of varying sizes while maintaining computational efficiency.

IV. EXPERIMENTAL SETUP

To evaluate the performance of both models for brain tumor classification, a series of experiments were conducted using the BR35H MRI dataset and implemented in Python. Both models were trained under identical training hyperparameters to ensure a fair comparison, as summarize in Table I. Adam optimizer with a learning rate of 0.001, a batch size of 16, and 30 epochs are used in the experiments. Prior to training, all MRI images from the BR35H dataset were preprocesses, including resizing to 224×224 pixels. The 224×224 pixels image size was chosen to balance computational efficiency and the preservation of critical anatomical details, as larger images significantly increase training time and memory requirements, while smaller images may lose important tumor features that are essential for accurate detection. The batch size of 16 was selected to optimize memory usage while maintaining stable gradient updates during training. The models were trained for 30 epochs, providing sufficient iterations for convergence and adequate learning while minimizing the risk of overfitting. The selection of image size, number of epochs, and batch size in these experiments follows the settings reported in previous studies[41][42][43][44][45][46][47][48][49]. All the experiments were performed in a CPU-based environment, demonstrating the feasibility of the approach in resource-constrained settings.

Table I.

Hyperparameter settings for CNN and YOLOv8 models training

Model / Hyperparameter	CNN	YOLOv8
Optimizer	Adam	Adam
Activation Function	ReLU (Rectified Linear Unit)	SiLU (Sigmoid Linear Unit / Swish)
Kernel size	3×3	3×3
Number of Conv Layers	3	225
Loss Function	CrossEntropyLoss	CrossEntropyLoss
Image Size (pixels)	224×224	224×224
Learning rate	0.001	0.001
Epoch Size	30	30
Batch Size	16	16

The BR35H brain tumor dataset, consisting of two classes which is tumor and non-tumor is used for evaluation, with details in Table II and sample images shown in in Fig. 4.

Table II

BR35H Dataset Distribution

Classes	Number of Images
Tumor	1500
Non-Tumor	1500
Total	3000

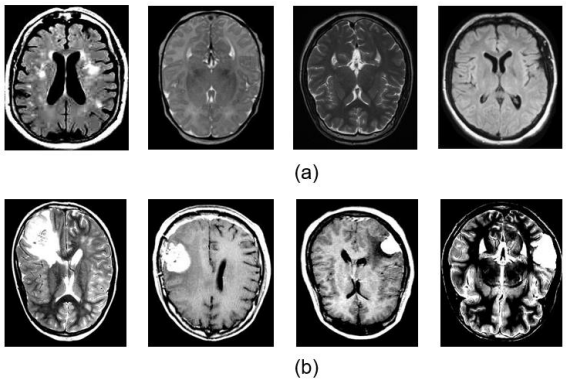


Fig. 4. Example images in BR35H dataset (a) non-tumor (b) tumor

For model evaluation, an 80:20 train-test split was applied to the dataset for training and testing. This ratio is adopted in this research following prior studies, as most deep learning-based works in the literature use similar splits for training and evaluation [50][51][52][53][54][55]. Following these ratios, the training set comprises 2,400 images, while the remaining 600 images are used for testing, as detailed in Table III.

Table III

Dataset Distribution for Training and Testing

Dataset Split	Training Set (80%)	Testing Set (20%)
Non-Tumor	1200	300
Tumor	1200	300
Total	2400	600

V. RESULTS AND DISCUSSION

In this research, the performance of the brain tumor classification models was evaluated using four performance metrics which are accuracy, precision, recall and F1-score including inference time. Accuracy represents the overall

https://doi.org/10.31436/ijpcc.v12i1.639

proportion of correct predictions across both tumor and non-tumor cases. Precision reflects the proportion of correctly identified tumor cases among all instances predicted as tumor, while recall measures the model's ability to detect all positive cases. The F1-score provides a mean of precision and recall, offering a balanced performance indicator. The inference time per image was recorded in each experiment to evaluate computational efficiency. The formulas for calculating each of the evaluation metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = 2x \frac{(Precision\ x\ Recall)}{Precision + Recall} \quad (4)$$

where TP, FP, TN and FN are defined in Table IV.

Table IV.
Terminology and derivations

Terminology	Derivations
TP (True Positive)	Tumor image correctly identified as tumor
TN (True Negatives)	Non-tumor image correctly identified as non-tumor
FP (False positives)	Non-tumor image incorrectly identified as tumor
FN (False negatives)	Tumor image incorrectly identified as non-tumor

A comparison of the performance of CNN and YOLOv8 in classifying brain tumor MRI images using the BR35H dataset are presented in Table V.

Table V

Comparative results (accuracy, precision, recall and inference time per image) for CNN and yolov8

Models	Accuracy	Precision	Recall	F1-score	Inference Time per Image (ms)
CNN	0.980	0.983	0.977	0.980	4.5
YOLOv8	0.998	0.997	1.00	0.998	15.1

Overall, the results demonstrate that both models achieved high accuracy, precision, recall, and F1-score, confirming their effectiveness for binary brain tumor detection. However, slightly lower performance was observed with the CNN model, which achieved an accuracy of 0.980, a precision of 0.983, a recall of 0.977, and an F1-score of 0.980. While CNN performed well, YOLOv8 demonstrated superior performance across all metrics, achieving an accuracy of 0.998, a precision of 0.997, recall of 1.00, and an F1-score of 0.998. Although the CNN results were slightly lower than YOLOv8, the promising results shown in the CNN methods indicate that the CNN architecture is capable of learning discriminative features from MRI images and provides reliable classification performance. The slight discrepancy between precision and recall shows that CNN occasionally misclassified a small number of tumor or non-tumor samples, which is expected given its relatively shallow architecture compared to modern deep learning models. In contrast, the value of 1.00 in recall implies that YOLOv8 successfully identified all tumor samples in the test set without missing any cases. This is very important aspect especially in medical diagnosis where false negatives carry significant clinical risks. The enhanced performance is due to YOLOv8's advanced architecture, which incorporates optimized convolutional blocks, improved feature extraction modules, and more efficient gradient flow, enabling the model to capture complex patterns in MRI images more effectively than the CNN. While both models performed strongly, the difference in accuracy and recall highlights YOLOv8's model is more robustness, generalizable and sensitive in detection. These results shows that YOLOv8 is more suitable for deployment in real-time clinical decision-support systems, where high reliability and low error rates are essential. Overall, YOLOv8 outperformed the CNN in accuracy, precision, recall and F1-Score, demonstrating its effectiveness as a modern deep learning model for brain tumor classification. Although YOLOv8 has a slightly higher inference time of 15.1 ms per image, which is only 10.6 ms longer than the CNN, it achieves superior performance and enhanced interpretability. The trade-off between speed and accuracy shows that YOLOv8 is well-suited for real-time clinical decision-support systems, and its inference time remains feasible for deployment in resource-constrained environments, making it a viable option for clinical application.

The confusion matrix analysis also further confirms that misclassification rates were extremely low for both classes especially for YOLOv8 as shown in Fig.5. Fig.5 (a) presented CNN results while Fig.5 (b) presented YOLOv8. CNN results in Fig.5 (a) show that the model correctly classified 295 non-tumor images and 293 tumor images. The CNN model generated 5 false positives, where non-tumor cases were incorrectly predicted as tumor, and 7 false negatives, where tumor cases were incorrectly classified as non-tumor. These misclassifications indicate that although the CNN achieves high accuracy, its sensitivity to subtle tumor features remains slightly

https://doi.org/10.31436/ijpcc.v12i1.639

limited, particularly in cases where tumor margins appear blurred or low contrast. In contrast, the confusion matrix of YOLOv8 as shown in Fig.5 (b) demonstrates substantially stronger performance. The model correctly predicted all 300 non-tumor cases and 299 tumor cases, resulting in zero false positives and only one false negative. This reflects YOLOv8's superior capacity to distinguish between tumor and non-tumor patterns, benefiting from deeper feature extraction, multi-scale representation learning, and an optimized detection head. The near-perfect separation between the two classes further explains its significantly higher precision, recall and F1-score compared to the CNN model.

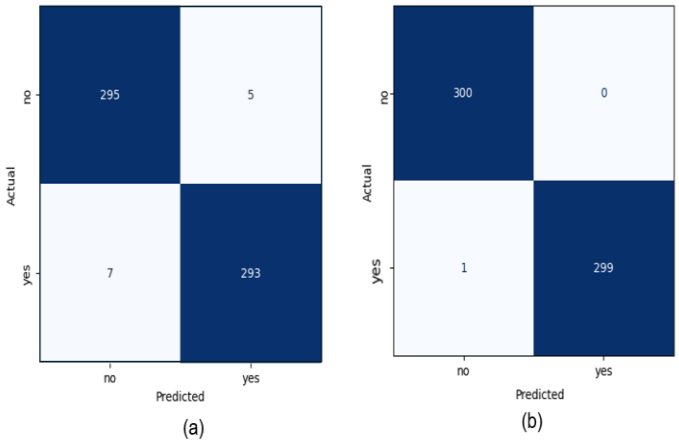


Fig. 5. Comparison confusion matric (a) CNN and (b) yolov8 model

Overall, the results clearly demonstrate that YOLOv8 outperforms the CNN in both recall and precision. This superior performance indicates that YOLOv8 is more suitable for deployment in real-time or clinical decision-support settings, particularly where minimizing false negatives is crucial. Detecting tumor cases accurately is essential, as missed detections can delay diagnosis and treatment, potentially worsening patient outcomes. YOLOv8's ability to significantly reduce false negatives, therefore makes it highly reliable for early and clinically sensitive tumor detection.

To further evaluate the comparative performance of the CNN and YOLOv8 models, The Receiver Operating Characteristic (ROC) curves for the CNN and YOLOv8 models are presented in Fig. 6(a) and Fig. 6(b), respectively.

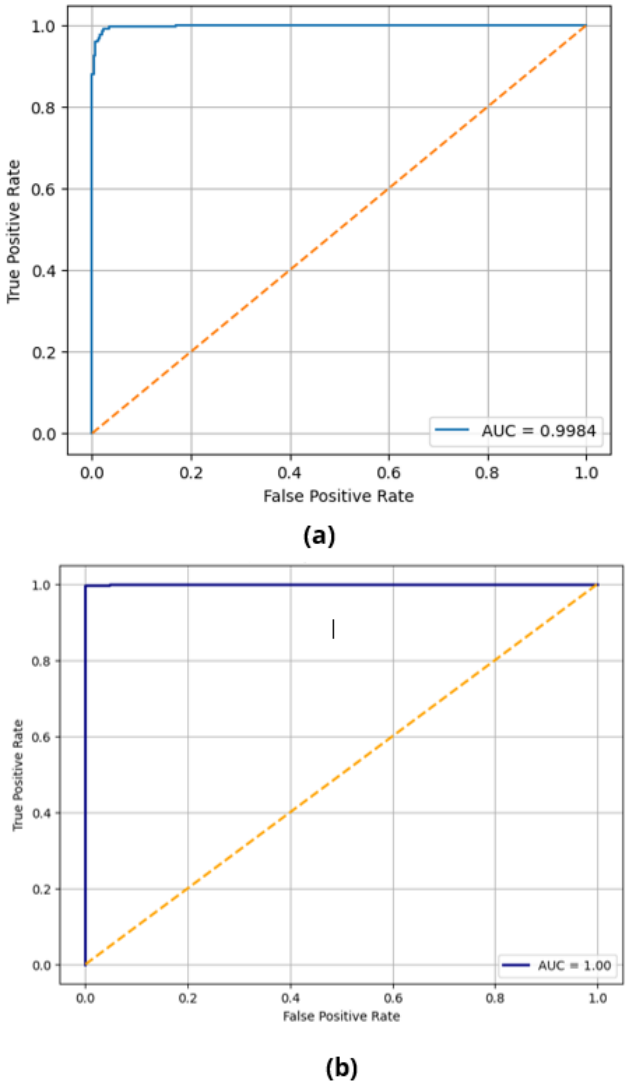


Fig. 6. Comparison ROC Curve (a) CNN and (b) yolov8 model

These ROC curves provide a threshold-independent evaluation of the classification performance of both models on the dataset tested. The curve illustrates the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across varying decision thresholds, providing a threshold-independent assessment of classification performance. As shown in Fig. 6 (a), the CNN model achieves an Area Under the Curve (AUC) of 0.9984, indicating excellent discriminative capability between tumor and non-tumor classes. The curve rises steeply toward the upper-left corner, reflecting a high true positive rate with a relatively low false positive rate across most thresholds. This shows that the CNN model is effective in correctly identifying tumor cases (high TP) while maintaining a low rate of misclassifying non-tumor images as tumor (low FP). While the ROC curve for the

https://doi.org/10.31436/ijpcc.v12i1.639

YOLOv8 model as presented in Fig. 6 (b) demonstrates an AUC of 1.00, indicating near-perfect separation between tumor and non-tumor classes on the dataset tested. The curve closely follows the left and top borders of the ROC space, shows that YOLOv8 consistently achieves a very high true positive rate with minimal false positives. This near-perfect performance is due to the characteristics of the BR35H dataset, which is relatively clean, balanced, and well-curated, with clearly defined tumor and non-tumor cases. This demonstrates strong robustness across decision thresholds, reducing the likelihood of false negatives (FN), which is particularly critical in clinical settings where missed tumor detections can delay diagnosis and treatment.

Overall, the ROC analysis highlights that both models exhibit strong classification capability; however, YOLOv8 demonstrates superior threshold stability and robustness. Rather than relying solely on point estimates such as accuracy, the ROC results confirm that YOLOv8 maintains reliable performance across a wide range of operating points, supporting its suitability for automated brain tumor detection, especially in clinical and resource-constrained environments where consistent decision-support systems is essential.

While the ROC curve analysis confirms the strong discriminative capability of both models across varying decision thresholds, such performance metrics alone are insufficient to establish clinical trustworthiness. In medical imaging applications, it is equally important to understand how and why a model arrives at its predictions, particularly in brain tumor detection where erroneous or poorly justified decisions may have serious clinical consequences. Therefore, we further investigate the internal decision-making behavior of the CNN and YOLOv8 models using Gradient-weighted Class Activation Mapping (Grad-CAM). This explainable artificial intelligence (XAI) approach enables visualization of the spatial regions within MRI scans that most strongly influence model predictions, thereby facilitating assessment of whether the models attend to clinically meaningful tumor regions or rely on indirect contextual cues or imaging artifacts. Such interpretability analysis is essential not only for validating model reliability under real clinical variability but also for supporting safe human-in-the-loop deployment, where automated predictions are intended to assist rather than replace expert radiological judgment. Fig. 7 (a) and Fig. 7 (b) illustrates representative Grad-CAM visualizations for the CNN and YOLOv8 models, providing insight into how each architecture using spatial information within brain MRI scans to support its predictions. In these visualizations, warmer colors indicate regions that contribute most strongly to the model's decision, allowing direct assessment of whether the learned attention aligns with the visible tumor anatomy.

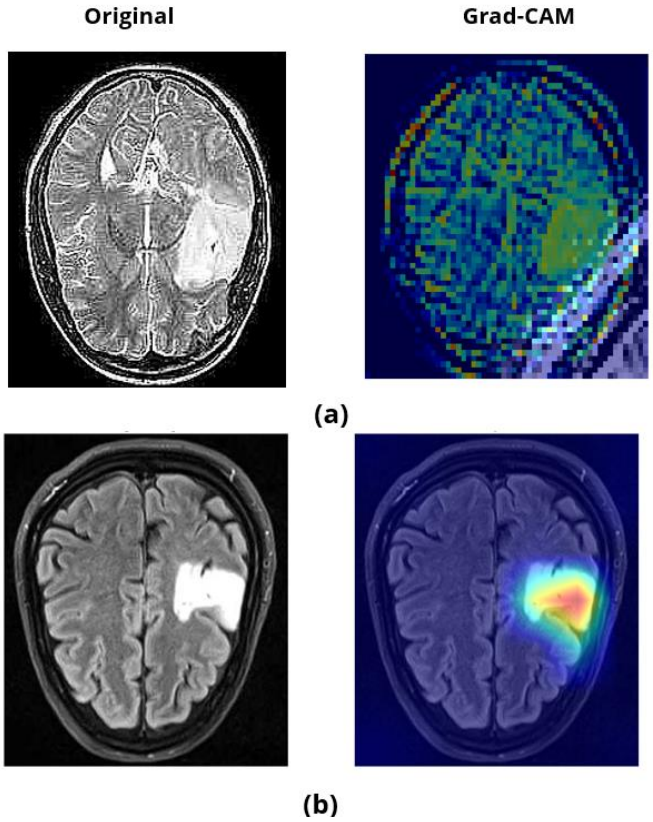


Fig. 7. Grad-CAM visualizations for brain tumor detection (a) CNN and (b) YOLOv8

As shown in Fig. 7 (a) the CNN model exhibits a relatively diffuse and fragmented activation pattern. Although the original MRI clearly demonstrates a prominent hyperintense lesion consistent with a brain tumor, the corresponding Grad-CAM heatmap highlights multiple regions of moderate importance distributed across both tumorous and non-tumorous areas. The absence of a sharply localized focus suggests that the CNN may be relying on a combination of global image characteristics, contextual anatomical cues, or correlated features rather than selectively attending to the pathological region itself. While this behaviour does not preclude high classification performance, it raises concerns regarding interpretability and robustness, particularly in real clinical environments where MRI data are inherently heterogeneous. Variations in scanner vendors, magnetic field strengths, acquisition protocols (e.g., T1-, T2-, or FLAIR-weighted imaging) and the presence of noise or motion artifacts may alter these contextual cues, potentially limiting generalizability

when models depend on indirect or dataset-specific patterns.

In contrast, Fig. 7 (b) demonstrates that the YOLOv8 model produces a markedly more focused and clinically intuitive activation map. The Grad-CAM heatmap reveals intense, well-localized attention precisely over the tumor region, with minimal activation in surrounding healthy tissue. This spatial correspondence indicates that YOLOv8 has learned to prioritize salient pathological features. Such localization capability is valuable in real-world MRI settings, demonstrating the model's robustness to variations in imaging protocols, scanner specifications, and common artifacts. By focusing on the actual tumor regions, the model is less likely to be misled by irrelevant patterns in the data, making its predictions more reliable across institutions and patient groups.

From a clinical and ethical perspective, these interpretability findings are critical. High-performing models that lack transparent and anatomically aligned reasoning pose significant challenges for clinical trust. Reliance on such "black-box" systems particularly in brain tumor diagnostics can lead to delayed diagnoses or inappropriate clinical decisions. In contrast, the clearer localization exhibited by YOLOv8 facilitates safer integration into clinical workflows by supporting human-in-the-loop deployment, allowing radiologists to verify that AI predictions focus on meaningful tumor regions and exercise informed oversight. This is essential for ensuring patient safety, maintaining clinical accountability and mitigating the risks associated with fully automated decision-making.

VI. CONCLUSION AND FUTURE WORK

This paper presented a comparative analysis of two deep learning approaches namely, CNN and YOLOv8 for brain tumor classification using the BR35H MRI dataset in resource-constrained settings. Both models demonstrated strong detection capability, where YOLOv8 obtain a superior performance with 0.998 accuracy, 0.997 precision, 1.00 recall, and 0.998 F1-score. Its superior performance reflects its enhanced feature extraction capability and efficient architectural design, which appears more robust in distinguishing between tumor and non-tumor classes. The confusion matrix analysis further confirmed these observations, as YOLOv8 produced fewer misclassifications than the CNN model. The YOLOv8 model exhibited a substantial reduction in false negatives, an essential performance criterion for early and reliable tumor detection. This is particularly important from a clinical and patient-care perspective, as missed tumor cases can lead to delayed diagnosis, postponed treatment initiation, and potentially poorer health outcomes. By minimizing false negatives, YOLOv8 helps ensure that patients receive timely medical attention, which is critical for improving prognosis and overall survival rates. These results further indicate that YOLOv8 is a more reliable candidate for real-world deployment, particularly in diagnostic support systems where minimizing missed detections is crucial.

Despite the promising, near-perfect results obtained on the tested dataset, these findings may not fully reflect performance in real-world clinical settings, where MRI data are typically more heterogeneous and may include variations arising from different scanners, imaging protocols, noise, artifacts or rare tumor variants. Several limitations therefore remain to be addressed in future work. While this study focused on binary classification, extending the evaluation to multi-class tumor types such as meningioma, glioma, and pituitary tumors would enable a more clinically meaningful and comprehensive assessment. Although YOLOv8's ability to minimize false negatives (FN) highlights its potential as a reliable tool for early and clinically sensitive tumor detection, ethical and safety considerations must be carefully addressed to mitigate the risks associated with over-reliance on automated diagnosis. In particular, future research should emphasize the integration of human-in-the-loop frameworks, where AI systems function as decision-support tools that complement and assist expert radiologists. Further validation using diverse, multi-institutional datasets, cross-dataset evaluation, and experiments with multiple random seeds or k-fold cross-validation would provide a more thorough assessment of model consistency and reliability, thereby strengthening the generalizability and robustness of both CNN and YOLOv8 models. In addition, future studies could explore more advanced architectures, such as Vision Transformers (ViT) or hybrid CNN Transformer models in higher-resource environments, as well as lightweight adaptations of these architectures for deployment in resource-constrained clinical settings.

ACKNOWLEDGMENT

Special thanks to Kulliyyah of Information and Communication Technology, International Islamic University Malaysia and Faculty of Computer Science and Information Technology, University of Malaya for providing support and resources.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] N. M. Ramli and N. R. Mohd Zain, "The Growing Problem of Radiologist Shortage: Malaysia's Perspective," *Korean J. Radiol.*, vol. 24, no. 10, pp. 936–937, 2023, doi: 10.3348/kjr.2023.0742.

[2] S. Fong *et al.*, "Education Research : Training of Neurologists in South East Asian Countries," vol. 0, no. 1, pp. 1–11, 2025, doi: 10.1212/NE9.0000000000200201.

[3] O. J. P. Odong, M. Abo-Zahhad, and M. Abdelwahab, "An integrated deep convolutional neural networks framework for the automatic segmentation and grading of glioma tumors using multimodal MRI scans," *Discov. Appl. Sci.*, vol. 7, no. 8, 2025, doi: 10.1007/s42452-025-06968-5.

[4] L. H. da Cruz, R. R. Batista, and C. C. Rangel, "MR imaging evaluation of posttreatment changes in brain neoplasms," in *Functional Imaging in Oncology: Clinical Applications - Volume 2*, 2014, pp. 603–640. doi: 10.1007/978-3-642-40582-2_2.

[5] A. K. Bhagat and D. Vekariya, "Computational Intelligence approach to improve the Classification accuracy of Brain Tumor Detection," in *Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022*, 2022, pp. 659–664. doi: 10.1109/IC3I56241.2022.10073470.

[6] M. P. Jeyaraj, M. S. Kumar, P. Jeyaraj M, and S. Kumar M, "Automated Brain Tumor Segmentation using Hybrid YOLO and SAM," *Curr. Med. IMAGING*, vol. 21, 2025, doi: 10.2174/0115734056392711250718201911 WE - Science Citation Index Expanded (SCI-EXPANDED).

[7] A. Ashar, V. Sabhani, and D. P. Bavirisetti, "Classification of Magnetic Resonance Imaging (MRI) Scans for Brain Tumor Using Improved EfficientNet Architecture and Transfer Learning," in *2022 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022*, 2022. doi: 10.1109/ICDSAAI55433.2022.10028839.

[8] S. Margasagayan and N. Nageswaran, "Investigations of Deep Learning Algorithms for Identification of Brain Tumors," SRM Valliammai Engineering College, Department of Medical Electronics, Kattankulathur, India: Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCCSMD63546.2024.11015215.

[9] S. N. Appe, G. Arulselvi, and G. N. Balaji, "Enhancing brain cancer detection and localization using Yolov8 object detection: A deep learning approach," in *Machine Learning and Generative AI in Smart Healthcare*, 2024, pp. 261–280. doi: 10.4018/979-8-3693-3719-6.ch013.

[10] S. Roy, S. Sen, R. Mehera, R. K. Pal, and S. K. Bandyopadhyay, "Brain Tumor Detection: A Comparative Study Among Fast Object Detection Methods," in *Lecture Notes in Networks and Systems*, 2021, pp. 179–196. doi: 10.1007/978-981-16-4294-4_12.

[11] M. Z. Khan, F. U. Zaman, P. Bhatti, A. A. Khan, S. A. Sultan, and S. Dua Bhatti, "A Comparative Analysis of Deep Learning Architectures for Efficient Brain Tumor Detection," in *2024 International Conference on Sustainable Technology and Engineering, i-COSTE 2024*, 2024. doi: 10.1109/i-COSTE63786.2024.11024836.

[12] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019.

[13] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Comput. Biol. Med.*, vol. 111, p. 103345, 2019, doi: 10.1016/j.compbiomed.2019.103345.

[14] D. Reyes and J. Sánchez, "Performance of convolutional neural networks for the classification of brain tumors using magnetic resonance imaging," *Heliyon*, vol. 10, no. 3, p. e25468, 2024, doi: 10.1016/j.heliyon.2024.e25468.

[15] F. Muftic, M. Kadunic, A. Musinbegovic, A. A. Almisreb, and H. Ja'afar, "Deep learning for magnetic resonance imaging brain tumor detection: evaluating ResNet, EfficientNet, and VGG-19," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 6, pp. 6360–6372, 2024, doi: 10.11591/ijece.v14i6.pp6360-6372.

[16] S. Kulkarni, P. A. Bhosale, and S. Das, "Using CNN for brain tumor diagnosis: An overview," in *Future of AI in Biomedicine and Biotechnology*, 2024, pp. 104–124. doi: 10.4018/979-8-3693-3629-8.ch006.

[17] P. K. Parhi, S. Mohanty, A. Mohanty, and P. K. Patra, "Deep Learning Techniques to Detect Brain Tumors Using the EfficientNet-B0 CNN Architecture," in *Artificial Intelligence in Oncology: Cancer Diagnosis and Treatment, Medical Imaging, and Personalized Medicine*, 2025, pp. 115–126. doi: 10.1007/978-3-031-94302-7_8.

[18] P. Kataria, A. Dogra, M. Gupta, T. Sharma, and B. Goyal, "Trends in DNN Model based Classification and Segmentation of Brain Tumor Detection," *Open Neuroimag. J.*, vol. 16, 2023, doi: 10.2174/18744400-v16-230405-2022-3.

[19] N. Verma and V. K. Bohat, "EfficientNet-based deep learning approach for early detection of brain tumors," *Qual. Quant.*, 2025, doi: 10.1007/s11135-025-02499-8.

[20] A. Y. Jaffar, "Combining Local and Global Feature Extraction for Brain Tumor Classification: A Vision Transformer and iResNet Hybrid Model," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 5, pp. 17011–17018, 2024, doi: 10.48084/etasr.8271.

[21] G. N. Sundar, D. Narmadha, N. A. Jerry, S. K. Thangavel, S. K. Shanmugam, and A. A. Ajibesin, "Brain Tumor Detection and Classification using Vision Transformer (ViT)," in *3rd International Conference on Automation, Computing and Renewable Systems, ICACRS 2024 - Proceedings*, 2024, pp. 562–567. doi: 10.1109/ICACRS62842.2024.10841703.

[22] X. Jia, E. Shkolyar, M. A. Laurie, O. Eminaga, J. C. Liao, and L. Xing, "Tumor detection under cystoscopy with transformer-augmented deep learning algorithm," *Phys. Med. Biol.*, vol. 68, no. 16, 2023, doi: 10.1088/1361-6560/ace499.

[23] S. S. Mahanty, D. Muduli, A. Kumari, and S. K. Sharma, "Pretrained DeIT for Brain Tumor Classification: A Fine-Tuning Approach with Label Smoothing," in *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*, 2024. doi: 10.1109/ICCCNT61001.2024.10725957.

[24] C. Sankari, V. Jamuna, and A. R. Kavitha, "Hierarchical multi-scale vision transformer model for accurate detection and classification of brain tumors in MRI-based medical imaging," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-23100-0.

[25] S. Jraba, M. Elleuch, H. Ltifi, and M. Kherallah, "Enhanced Brain Tumor Detection Using Integrated CNN-ViT Framework: A Novel Approach for High-Precision Medical Imaging Analysis," in *10th 2024 International Conference on Control, Decision and Information Technologies, CoDIT 2024*, 2024, pp. 2120–2125. doi: 10.1109/CoDIT62066.2024.10708482.

[26] M. N. I. Shanto, M. T. Mubtasim, S. V Rakshit, and M. A. Aman Ullah, "Enhanced Classification of Brain Tumors from MRI Scans Using a Hybrid CNN-Transformer Model," M. S. Shah, Ed., Kennesaw State University, Department of Computer Science, Kennesaw, United States: Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/QPAIN66474.2025.11171896.

[27] H. Lin, J. Wang, Y. Guo, and L. Fang, "A Lightweight Brain Tumor Detection Network with Dynamic Inverted Residuals and Synergistic Attention," in *2025 4th International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology, AIoTC*

2025, 2025, pp. 185–191. doi: 10.1109/AIoTC66747.2025.11198723.
[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
[29] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv Prepr. arXiv2004.10934*, 2020.
[31] F. Ghufran, A. Kumar, A. Kumar, A. Kaushik, and S. Saxena, "Detection of Brain Tumors in MRI Images through Machine Learning, Deep Learning, Faster R-CNN & YOLOv5," V. Sharma and J. Sinha, Eds., GNIT Group of Institutions, Department of Computer Science and Engineering, Noida, India: Institute of Electrical and Electronics Engineers Inc., 2024, pp. 754–759. doi: 10.1109/ICAC2N63387.2024.10895900.
[32] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475.
[33] "Explore Ultralytics YOLOv8." Accessed: Nov. 16, 2025. [Online]. Available: https://docs.ultralytics.com/models/yolov8/
[34] A. P. Taradale and S. L. Kattimani, "Effective Segmentation and Classification of Brain Tumour MRI Images Using YOLOv8 Technique," in *2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA)*, IEEE, 2024, pp. 1–6.
[35] O. Zougari *et al.*, "Deep Learning-Based Brain Tumor Detection Using YOLOv8 on MRI Images," in *Lecture Notes in Networks and Systems*, 2026, pp. 389–401. doi: 10.1007/978-3-032-01967-7_37.
[36] S. Verma, A. Choudhury, P. Narad, and A. Sengupta, "Optimizing Brain Tumor Classification Using YOLOv8: A Comparative Study with CNN Architectures on MRI Scans," in *International Conference on Signal Processing and Communication, ICSC*, 2025, pp. 500–505. doi: 10.1109/ICSC64553.2025.10968576.
[37] K. Dhivya, U. Surya, and A. Devi, "Empowered Brain Tumor Detection Using Deep Learning Methodology," in *International Conference on Advancements in Power, Communication and Intelligent Systems, APCI 2024*, 2024. doi: 10.1109/APCI61480.2024.10616431.
[38] "Br35H: Brain Tumor Detection 2020, Kaggle, 2020." [Online]. Available: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
[39] M. Reza Keyvanpour and M. B. Shirzad, *Machine learning techniques for agricultural image recognition*. INC, 2022. doi: 10.1016/B978-0-323-90550-3.00011-4.
[40] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverria, "Yolov5 vs. yolov8: Performance benchmarking in wildfire and smoke detection scenarios," *J. Image Graph.*, vol. 12, no. 2, pp. 127–136, 2024.
[41] D. Garg, "A Deep Learning Approach for Face Detection using YOLO," *2018 IEEE Punecon*, pp. 1–4.
[42] S. Arora and M. Sharma, "Deep Learning for Brain Tumor Classification from MRI Images," in *Proceedings of the IEEE International Conference Image Information Processing*, The NorthCap University, Department of Computer Science and Engineering, Gurugram, India: Institute of Electrical and Electronics Engineers Inc., 2021, pp. 409–412. doi: 10.1109/ICIIP53038.2021.9702609.
[43] S. Krishnapriya and Y. Karuna, "A deep learning model for the localization and extraction of brain tumors from MR images using YOLOv7 and grab cut algorithm," *Front. Oncol.*, vol. 14, 2024, doi: 10.3389/fonc.2024.1347363 WE - Science Citation Index Expanded (SCI-EXPANDED).
[44] S. Iftikhar, N. Anjum, A. B. Siddiqui, M. Ur Rehman, and N. Ramzan, "Explainable CNN for brain tumor detection and classification through XAI based key features identification," *Brain Informatics*, vol. 12, no. 1, 2025, doi: 10.1186/s40708-025-00257-y.
[45] M. S. Mou *et al.*, "Brain Tumor Detection on MRI Images Using a Combination of CNN and Ensemble Learning Approach," M. B. Islam, M. E. Hamid, and S. Biswas, Eds., Rabindra Maitree University, Department of Computer Science and Engineering, Kushtia, Bangladesh: Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICRPSET64863.2024.10955888.
[46] P. Venkateswarlu Reddy *et al.*, "Implementation of Latest Deep Learning Techniques for Brain Tumor Identification from MRI Images," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2023, pp. 1166–1171. doi: 10.1109/ICCES57224.2023.10192620.
[47] H. Mzoughi, I. Njeh, M. BenSlima, N. Farhat, and C. Mhiri, "Vision transformers (ViT) and deep convolutional neural network (D-CNN)-based models for MRI brain primary tumors images multi-classification supported by explainable artificial intelligence (XAI)," *Vis. Comput.*, vol. 41, no. 4, pp. 2123–2142, 2025, doi: 10.1007/s00371-024-03524-x.
[48] S. Sarker, "Transfer Learning and Explainable AI for Brain Tumor Classification: A Study Using MRI Data from Bangladesh," *2024 6th Int. Conf. Sustain. Technol. Ind. 5.0, STI 2024*, vol. 0, pp. 1–6, 2024, doi: 10.1109/STI64222.2024.10951092.
[49] A. M. J. Zubair Rahman *et al.*, "Advanced AI-driven approach for enhanced brain tumor detection from MRI images utilizing EfficientNetB2 with equalization and homomorphic filtering," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12911-024-02519-x.
[50] R. Deebika and M. Sangeetha, "Optimized ResNet-50 Deep Learning Model for Highly Accurate Plant Disease Classification and Detection Efficiency," in *Proceedings of 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2025*, 2025, pp. 47–52. doi: 10.1109/ICICV64824.2025.11085739.
[51] W. H. M. Isa, M. A. Abdullah, M. A. M. Razman, A. P. P. A. Majeed, and I. M. Khairuddin, "Deep Learning Algorithms for Recognition of Badminton Strokes: A Study Using SDNN, RNN, and RNN-GRU Models with Off-Court Video Capture," in *Lecture Notes in Networks and Systems*, 2024, pp. 53–60. doi: 10.1007/978-981-99-8498-5_5.
[52] S. John and M. G. Jibukumar, "Lightweight Sequential CNN for Alzheimer's Disease Progress Classification," in *2024 Asian Conference on Intelligent Technologies, ACOIT 2024*, 2024. doi: 10.1109/ACOIT62457.2024.10939666.
[53] İ. Ceran, M. Kaya, Y. Kaçmaz, A. E. Ergün, T. Çelikten, and A. Onan, "A Deep Learning Approach to Sentiment Classification: Insights from Product Review Analysis," in *Lecture Notes in Networks and Systems*, 2025, pp. 37–44. doi: 10.1007/978-3-031-97992-7_5.
[54] V. S. Threshika, P. Naiba Sree, P. R. Pravin, and S. Madhavi, "A Deep Learning Framework for the Classification of Indian Heritage Using Curated Datasets and CNN Models," in *Proceedings - 3rd International Conference on Advancement in Computation and Computer Technologies, InCACCT 2025*, 2025, pp. 951–954. doi: 10.1109/InCACCT65424.2025.11011360.
[55] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," 2018.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.651

Cross-Media Fake Content Detection via Independent Deep Learning Classifiers

Iqbal Najihah binti Samsul Kamal, Anna Safiya binti Samsudin, Raini binti Hassan*
Department of Computer Science, International Islamic University Malaysia, Gombak, Malaysia.

*Corresponding author: hrai@iium.edu.my
(Received: 15th December 2025; Accepted: 13th January, 2026; Published on-line: 30th January, 2026)

Abstract — The rapid advancement of generative models has enabled the creation of highly realistic fake multimedia content, including altered images, deepfake videos, and synthetic audio. These forgeries undermine information integrity and pose significant societal risks, especially by encouraging misinformation, digital fraud and impersonation. As these threats directly affect public trust and institutional transparency, they challenge the goals outlined in SDG 16: Peace, Justice, and Strong Institutions, which focuses on reducing corruption, preserving information integrity, and ensuring accountable, trustworthy systems. To address these issues, this paper proposes a deep learning–based system that classifies multimedia content across three modalities, which are image, video, and audio. Unlike conventional multimodal fusion approaches that necessitate paired data inputs, this paper introduces a novel routing-based unification architecture. The suggested framework makes use of a content-adaptive routing mechanism that treats each modality independently. Using a dual-backbone Swin Transformer and EfficientNet for images, Video Swin Transformer for video, and Wav2Vec 2.0 for audio, the system automatically determines the type of input file and sends it to the relevant specialized deep learning classifier. This design allows for a versatile, single-entry-point forensic tool that maintains high accuracy by leveraging domain-specific experts without the computational overhead of processing multiple streams concurrently. Experimental results demonstrate strong performance across individual modalities, with the audio model achieving 96.95% accuracy and the image model showing robust precision despite challenges posed by high quality generative forgeries.

Keywords— Deep Learning, Multimedia Forensics, Swin Transformer, Wav2Vec 2.0, Machine Learning, Data Science.

I. Introduction

The rapid development of generative models and artificial intelligence (AI) has drastically changed how multimedia content is created and altered. While these technologies have driven innovation in visual effects, digital media production, and human-computer interaction, they have also promoted the emergence of deepfakes, a highly realistic synthetic images, videos, and audio. Reliable detection methods are desperately needed because such content has the potential to disseminate false information, impersonate people, and weaken public trust in digital communication.

Deepfake techniques, including GANs, encoder-decoder models, and diffusion models, produce synthetic media that closely resembles real content, making manual detection more challenging [1]. These deepfakes have been used in identity fraud, political manipulation, disinformation campaigns, and various forms of social engineering, raising concerns for people, organizations, and public institutions. Consequently, the research community has prioritized developing automated systems that can distinguish between real and altered multimedia content across various modalities.

Although high-performance detection algorithms are available, there is still a big usability and system integration gap. Currently, many cutting-edge detection models are extremely specialized and made to handle a single modality, such as independently analysing only images, only video or only audio. Forensic analysts and regular users, who frequently need different software tools or platforms to verify various file types, are left with a fragmented landscape as a result. For instance, to verify a suspicious news report, it might be necessary to use one tool to look at the headline image and another environment to examine an audio clip that goes with it. This lack of unification slows down the reaction to disinformation campaigns and causes friction in the verification process.

This paper supports a unified "Cross-Media Fake Content Detection" framework using separate deep learning classifiers to address this fragmentation. This method emphasizes adaptability and architectural independence by developing core innovation of routing architecture that serves as a single interface for various media formats. The system cleverly directs the input to the most competent independent deep learning model by examining its structure.

https://doi.org/10.31436/ijpcc.v12i1.651

To ensure the resilience of these independent classifiers, this paper makes use of a variety of modality-specific benchmark datasets. For images, the IMD2020 dataset offers a balanced set of real and manipulated samples involving inpainting and real-world forgeries [2], while the CASIA 2.0 dataset offers traditional image modifications like splicing, and copy-move editing. In the video domain, the DeeperForensic1.0 (v2) dataset, which contains complex face-swapping manipulations, serves as a high-quality benchmark for deepfake detection [3]. For audio, the ASVspoof dataset provides standardized real and spoofed speech samples, including synthetized voice-converted and synthetized audio [4]. Furthermore, the models of images, video, and audio are trained on distinct repositories to ensures that the specific artifacts unique to each medium are learned accurately. This data-driven strategy ensures that the system is accurate in its detection capabilities across various content types and unified in its interface.

In the end, this methodology guarantees that a user can confirm a suspicious file in a single streamlined environment, regardless of whether it is an image, voice recording, or video clip. This project intends to provide a detection tool that is both accurate and practically deployable for real-world scenarios where the format of the incoming threat is unpredictable by combining specialized, independent classifiers under a single "Cross-Media" umbrella. It is crucial to note that the developed framework is not a fully deployable commercial forensic tool, but rather a research-oriented prototype evaluated under controlled conditions to show the viability of cross-media routing.

II. RELATED WORK

This paper builds on several works that remain relevant in today's multimedia fake news landscape. Transformer-based models have become a strong foundation due to their pretraining on large, modern datasets which reduce the need for traditional handcrafted feature engineering.

For image forgery detection, this paper adopts a hybrid approach using Swin Transformer and EfficientNet, chosen for their ability to capture both global and fine-grained details. Prior work has tended tothis stufy lean toward one side of this spectrum, which creates a clear comparative context for the present aproach. B. Singh et al. (2022) [9] relied on EfficientNet-B0 within a multimodal setting, pairing it with a text encoder for credibility analysis. Their reliance on EfficientNet provided strong local feature extraction, but their fusion design predated transformer-based attention mechanisms. Compared with that framework, the current study benefits from Swin Transformer's hierarchical global reasoning, providing a broader contextual understanding that EfficientNet alone could not capture in Singh et al.'s setup.

Similarly, Almsrahad et al. (2024) [10] used EfficientNet-B0, though they focused on ELA-processed images from CASIA. Their results highlight EfficientNet's usefulness for low-level forensic patterns, yet their dependence on ELA artifacts limits robustness to modern social-media imagery. In contrast, this paper avoids hand-crafted preprocessing and instead integrates EfficientNet with Swin Transformer to balance low-level artifact detection with higher-level semantic consistency, addressing the brittleness seen in ELA-driven pipelines.

More recent work has leaned toward transformer-only designs. Gong et al. (2024) [11] applied Swin Transformer to video frames, introducing consistency-loss mechanisms to strengthen temporal generalization. Their focus on temporal cues, however, leaves open the question of how Swin could be paired with CNN-based forensic extractors. Mishra et al. (2023) [12] further showed that Swin outperforms many CNN baselines in robustness, but their evaluation—like other Swin-centric studies—prioritizes transformer capacity over hybrid feature diversity.

Across these studies, the pattern is clear where EfficientNet-based approaches excel at localized artifacts but struggle with global context, while Swin-based approaches capture global structure yet often overlook lo-level forensic detail. By combining both, this paper positions itself between the two extremes, aiming to inherit the strengths of each and mitigate their individual weaknesses.

For audio forgery detection, this paper uses a hybrid of wav2vec2, BiLSTM, and an attention mechanism to balance high-level speech representations with temporal modelling. Prior work by J. M. Martin-Donas et al. [13] established wav2vec2 as a strong front-end feature extractor for audio deepfake detection. While their model combined wav2vec2 with downstream classifiers, architectures integrating BiLSTM with attention were not explored, leaving a gap in modelling longer temporal dependencies as well as localized acoustic cues.

Samia et al. (2024) [14] explored a hybrid architecture of CNN, BiLSTM, and Multi-Head Attention, showing that combining temporal modeling with attention significantly boosts reliability. A key limitation, however, is their use of CNN-based spectral features, which restricts the model to handcrafted inputs. We address this by employing wav2vec2 to work directly with raw waveform representations. This approach leverages learned speech embeddings rather than static spectral cues. Ultimately, by coupling wav2vec2 with BiLSTM and attention, our model captures both global patterns and local anomalies more effectively.

For video forgery detection, this paper employs Video Swin Transformer as a standalone backbone, prioritizing its capacity to learn complex spatio-temporal patterns directly

https://doi.org/10.31436/ijpcc.v12i1.651

from raw video clips. This produces a more robust representation of motion-based manipulations compared to models that focus only on spatial cues. In contrast, Khalid et al. (2023) [15] used a Swin Y-Net Transformer, where the Y-Net design fused multi-scale features through parallel Swin branches. Their model effectively captured both local and global forgery signals, yet the limited dataset in their study introduced overfitting, especially for specific manipulation types. This restricts the generalizability that our Video Swin implementation aims to preserve through more balanced and diverse training data.

Deressa Zhou et al. (2023) [16] explored a different angle by combining ConvNeXt, Swin Transformer, and AE/VAE components to detect visual artifacts and latent inconsistencies. Their hybrid design improved generalization on unseen deepfake datasets thanks to the latent reconstruction loss. However, their approach remained frame-level and lacked a dedicated temporal modeling head, meaning it could not fully exploit motion cues. The method also depended heavily on precise face extraction; performance degraded noticeably when evaluated on full-frame inputs. In contrast, the current study avoids such dependency by using Video Swin's native spatio-temporal processing, reducing reliance on face cropping and allowing the model to handle a wider range of video structures.

Broadly, this paper unifies image, audio, and video detection under a single framework to address the fragmentation in forensic tooling. Although cross-modal analysis has been studied in the past, these studies frequently suffer from dependency on paired inputs. In 2022, Zhou et al. [17] applied CLIP to align image and text features, improving fake-news detection on datasets such as Weibo, PolitiFact, and GossipCop. Such fusion-dependent architectures work well for news articles that contain both, but they fall short when analysing isolated media files (such as a standalone audio recording or video clip) in the absence of related text.

Two years later, Ma et al. (2024) [18] proposed an event-aware multi-view fusion framework combining text, image, and additional signals. Their model reduced ambiguity in mismatched news content by emphasizing event structure, which is beneficial for real-world news contexts. Nevertheless, the system is computationally demanding and less useful for general-purpose forensics where the context is unknown due to its reliance on event-level consistency.

This supports the credibility of our study, since we focus on a 'content-agnostic' system. Unlike these rigid fusion architectures, our study suggests a Cross-Media Routing Framework. Our method does not require simultaneous data inputs by treating each modality with a specialized, independent deep learning classifier. This gives the system a degree of flexibility that strict multimodal fusion models don't, ensuring that it works whether the user submits a single image, a voice recording, or a video file.

III. METHODOLGY

The procedure for developing the cross-media fake content detection framework is described in this section. The intelligent routing mechanism that unifies them comes after dataset preparation, preprocessing pipelines, model architecture, training methods, and evaluation metrics for each independent classifier.

A. *Fake Image Detection*

- *Dataset preparation*

A final custom dataset of 28,000 images was created by randomly selecting 14,000 samples per class using a fixed seed to ensure class balance. A tuple (image_path, label) was used to store each entry, with label 0 denoting real and label 1 denoting fake. To ensure strong generalization and avoid information leakage, the dataset was divided into 70% training, 20% validation, and 10% testing after being shuffled using sklearn.utils.shuffle.

- *Preprocessing*

Two preprocessing pipelines were designed. The training pipeline included extensive augmentation to improve robustness against a variety of manipulation techniques. The transformations included RandomResizedCrop, RandomHorizontalFlip, RandomRotation (±10°), ColorJitter, RandomPerspective, GaussianBlur, RandomErasing, additive noise via a Lambda transform, and ImageNet normalization. These augmentations aid in exposing the model to generative artifacts and texture irregularities that are commonly found in manufactured media [5].

The evaluation pipeline only used to resize to 224x224 pixels, tensor conversion, and ImageNet normalization to ensure consistent and unbiased testing conditions.

- *Model Architecture*

A dual-backbone architecture was employed to take advantage of complementary visual representations. The first backbone, a Swin Transformer, offers hierarchical global-local modelling for the purpose of detecting subtle deepfake artifacts. The second backbone, EfficientNet-B3, uses a compound scaling design to capture fine-grained texture irregularities. Both backbones were kept completely frozen throughout training to minimize overfitting and training time.

https://doi.org/10.31436/ijpcc.v12i1.651

Let F_{img} and F_{eff} indicates the embeddings generated by EfficientNet-B3 and the Swin Transformer. The definition of the fused representation is:

$$F = [F_{\text{img}} ; F_{\text{eff}}],$$

where [;] denotes vector concatenation. This fused feature vector is subsequently transformed into a binary real-fake prediction by a multi-layer classifier.

- *Training*

The model was optimized using a cross-entropy objective with regularization and a cosine-annealing learning-rate schedule. To increase training efficiency, automatic mixed precision was employed. Based on validation performance, early stopping with a patience of five epochs was used to avoid overfitting.

- *Evaluation*

Performance was evaluated on the held-out test set using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrix, and a thorough per-class classification report. These metrics are in line with accepted methods in research on deepfake detection

B. Fake Video Detection

- *Dataset preparation*

For the video-based fake multimedia detection experiment, this paper utilized the DeeperForensic1.0 dataset, a large-scale benchmark for face manipulation detection. The dataset consists of high-quality real videos featuring 100 professional actors and their corresponding AI-generated videos, created using an end-to end face swapping framework []. A curated video dataset was created by sampling 2,00 real and 2,00 fake videos using a fixed seed to maintain class balance. A tuple (video_path, label) was used to store each dataset entry, with 0 denoting real and 1 denoting fake. To maintain the class distribution, the dataset was divided into 80% training and 20% validation.

- *Preprocessing*

Videos were preprocessed by uniformly sampling 8 frames per video, resizing frames to 224x224 pixels, and normalizing them using ImageNet statistics. During training, frame-level augmentation such as RandomResizedCrop and RandomHorizontalFlip were applied to improve robustness against varying visual content. For evaluation, only resizing and normalization were applied to maintain consistency.

- *Model architecture*

The foundation for video feature extraction was a Swin 3D Transformer (Tiny) that had been pretrained. To convert the extracted embeddings to binary predictions, a linear layer was used in place of the original classification head. Let F_vst represent the embeddings generated by the Swin3D backbone. The final forecast is calculated as follows:

$$y = softmax(FC(F_{\text{vst}}))$$

Where FC is the fully connected classification layer.

- *Training*

The model was trained using cross-entropy loss and the Adam optimizer with a learning rate of 1×10^{-4}. Because of memory limitations, the batch size was set to 5. The model with the lowest validation loss was saved as the last checkpoint, and early stopping was implemented based on validation loss. To maintain consistent input shapes, video padding and frame extraction were carefully handled during the ten epochs of training.

- *Evaluation*

The model was evaluated on the validation set using confusion matrix to assess per-class performance. The approach guarantees that the model retains generalization to unseen samples while learning discriminative temporal and spatial patterns suggestive of manipulated videos.

C. Fake Audio Detection

- *Dataset preparation*

The audio modality was developed using the ASVspoof 2019 Logical Access (LA) corpus, which contains of bonafide human speech and spoofed utterances generated through various text-to-speech and voice conversion systems [4]. To ensure a controlled and balanced training set, all 2,580 bonafide samples were kept and matched with 2,580 randomly chosen spoofed samples using a fixed seed. For the development and evaluation subsets, stratified sampling was then used to reduce both subsets while maintaining the initial class imbalance, yielding in 410 bonafide and 3,590 spoof files for development and 413 bonafide and 3,587 spoof files for evaluation. A clean, organized dataset appropriate for representation learning and subsequent classification is created by pairing each audio file with its matching label (0=bonafide, 1=spoof).

- *Preprocessing*

The pretrained Wav2Vec 2.0 model, which offers self-supervised embeddings that capture phonetic, spectral, and prosodic cues pertinent to spoof detection, was used to transform all audio files into fixed-length feature representations. The extracted embeddings were stored as PyTorch tensors to ensure consistent input dimensionality and prevent repeated computation. Since audio duration varies among utterances, sequences were only padded at the batch level during loading, allowing the model to process

https://doi.org/10.31436/ijpcc.v12i1.651

variable-length speech while maintaining temporal patterns.

- *Model Architecture*

The classifier consists of a bidirectional LSTM and an attention mechanism that highlights the most informative temporal frames for differentiating between bonafide and spoofed speech. The bidirectional design allows the model to capture long-rage temporal dependencies, while the attention layer generated a weighted representation that concentrates on segments with spoof-related artifacts. The aggregated representation is mapped to a binary output (bonafide vs. spood) by a fully connected layer.

- *Training*

The model was trained using cross-entropy loss and the Adam optimizer with a fixed learning rate. Training proceeded for a limited number of epochs, and the final model was chosen based on the lowest validation loss to reduce overfitting. Since the dataset contains class imbalance in the development and evaluation sets, metrics were tracked across both classes to guarantee stable generalization.

- *Evaluation*

Accuracy, precision, recall, F1-score, and confusion matrices were used to evaluate the model's performance, enabling a comprehensive understanding of both bonafide and spoof classes. This evaluation framework provides insight into false-accept and false-reject tendencies, which are crucial in anti-spoofing applications. The modular design also ensures that the audio classifier can be easily incorporated into the entire multimodal late-fusion pipeline.

D. *Cross-Media Routing and System Integration*

To operationalize the separate classifiers into a single, coherent framework, a unified inference class was created using PyTorch. By acting as an intelligent router, this system shields the user from the intricacies of the underlying model.

- *System Initialization and Resource Management*

All three pre-trained model architectures are loaded into GPU memory (cuda) by the system upon instantiation. The specific weights for each classifier are loaded from independent .pth checkpoints. By maintaining these as separate files, the system allows for the individual updating of a particular modality without necessitating a full system retraining.

- *Intelligent Input Routing*

A routing mechanism based on file extension is used in the core logic. The system examines the extension when a file path is passed to the predict() function to identify the proper processing stream. If an unspoorted format is detected, the system raises an error, enduring processing stability.

- *Dynamic Preprocessing*

The inference pipeline places more emphasis on consistency than training pipelines, which heavily rely on augmentation. To ensure that input tensors match the dimensions required by the corresponding backbones, the system uses OpenCV (cv2) to sample fixed video frames, torchaudio to normalize audio sampling rates, and PIL to resize images.

- *Unified Output Standardization*

The system creates a probability distribution by passing the raw model logits through a Softmax layer, regardless of the modality employed. Three essential metrics are included in the final output, which are the modality employed, the prediction label, and a confidence score.

IV. RESULTS

A. *Fake Image Detection*

The experimental results for fake image detection is presented in Table. 1. Similarly, Figure 1 shows that the model performed well on the held-out test set consisting of 2,800 images. Overall, the model achieved an accuracy of 73.4%, precision 72.0%, recall of 76.3%, and an F1-score of 74.1%, proving a balanced performance in detecting real and fake images.

TABLE I
RESULT OF FAKE IMAGE DETECTION ON THE EVALUATION SET.

Metric	**Score (%)**
Accuracy	73.4
Precision	72.0
Recall	76.3
F1-score	74.1

According to per-class results (Table 2) and confusion matrix (Figure 1), the model classified 990 as real (70.5%) out of 1,405 real images, while 415 were misclassified as fake. Conversely, 1,065 images were accurately detected as fake (76.3%) among 1,395 fake images, whereas 330 images were incorrectly labelled as real. This illustrates the model's comparatively better ability to identify phony images, probably because of unique generation artifacts that are still present in contemporary synthetic image pipelines.

TABLE II
CLASSIFICATION REPORT (PER-CLASS) FOR FAKE IMAGE DETECTION ON THE EVALUATION SET

	Precision (%)	**Recall (%)**	**F1-score (%)**
Real (0)	75.0	70.5	72.7
Fake (1)	72.0	76.3	74.1
Accuracy	73.4	73.4	73.4

https://doi.org/10.31436/ijpcc.v12i1.651

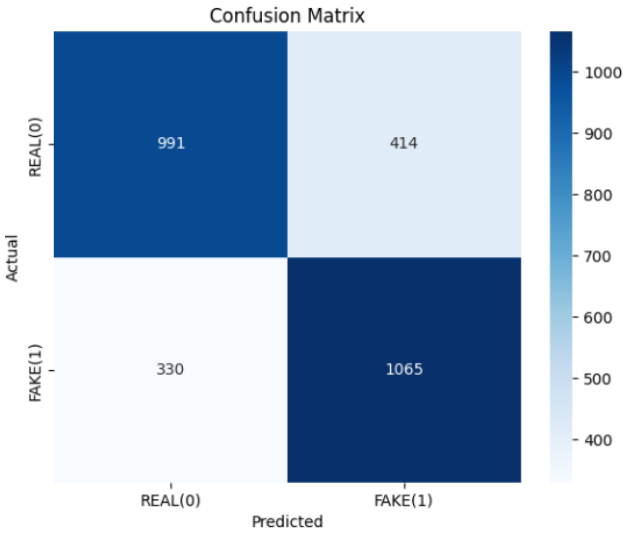


Fig. 1 Confusion Matrix for fake image detection

The model's discriminative ability is further supported by the ROC curve in Figure 2, which shows strong separability between the real and fake classes with a ROC-AUC of 0.824. The high AUC implies that the features successfully improve the model's capacity to discern minute cues present in altered images.

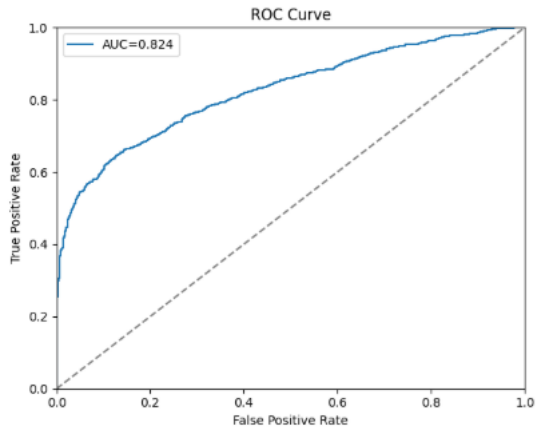


Fig. 2 ROC Curve for fake image detection

B. *Fake Video Detection*

Based on Table 3, an independent test set of 800 video samples, which have 400 real and 400 fake, was used to assess the video deepfake detection model. The model achieved 100% accuracy, precision, recall, and F1-score.

TABLE III
RESULT OF FAKE VIDEO DETECTION ON THE EVALUATION SET

Metric	**Score (%)**
Accuracy	100
Precision	100
Recall	100
F1-score	100

The per-class classification report shows perfect performance, with 100% precision, recall, and F1-score for both real and fake videos, yielding an overall evaluation accuracy of 100% (see Table 4).

TABLE IV
CLASSIFICATION REPORT (PER-CLASS) FOR FAKE VIDEO DETECTION ON THE EVALUATION SET.

	Precision (%)	**Recall (%)**	**F1-score (%)**
Real (0)	100	100	100
Fake (1)	100	100	100
Accuracy			100

According to Figure 3, a flawless classification pattern can be seen. All 400 of real videos were correctly classified as real (100%), with zero instances mistakenly identified as fake. Similarly, the model achieved a perfect score for fake videos (100%), correctly identifying each of the 400 instances with no false negatives. The absence of both false negative and false positive shows that the model maintains maximum specificity and sensitivity. Furthermore, the balanced precision and recall across classes proves that the model is unbiased toward either the real or fake class.

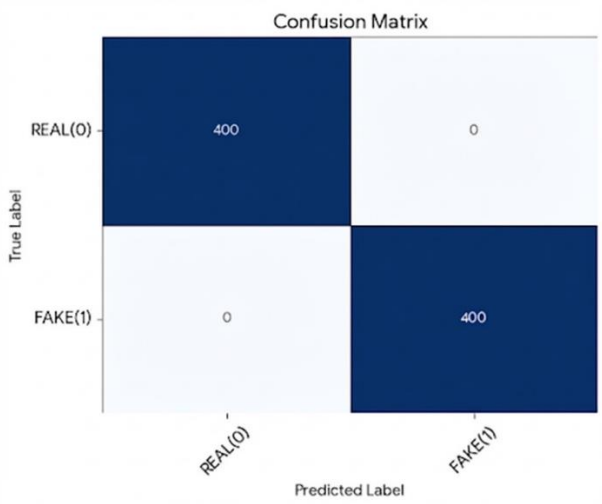


Fig. 3 Confusion Matrix for fake video detection

The ROC curve in Figure 4 shows perfect class separability with a ROC-AUC of 1.000, further supports the model's discriminative ability. This implies that complex spatiotemporal anomalies and synthesis artifacts present in phony videos are successfully captures by the Video Swin Transformer, enabling a clear differentiation between real and fake content.

https://doi.org/10.31436/ijpcc.v12i1.651

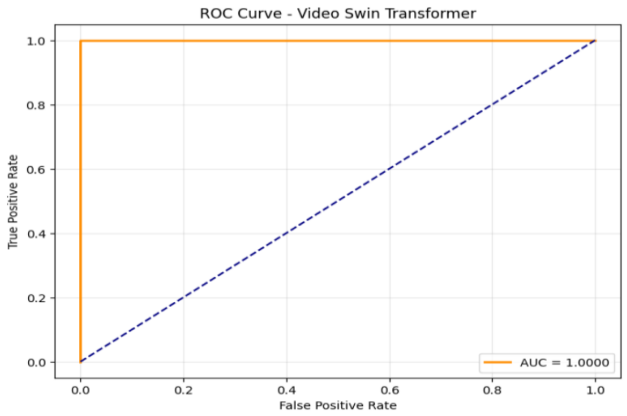


Fig. 4 ROC Curve for fake video detection

C. *Fake Audio Detection*

Fake audio detection achieved strong performance on the evaluation set. The model recorded an accuracy of 92.2%, demonstrating reliable overall classification. High precision of 99.4% indicates minimal false positives, while a recall of 91.8% reflects effective identification of fake audio samples (see Table 5). The F1-score of 95.5% confirms a balanced and robust detection capability, highlighting the model's effectiveness in distinguishing authentic and manipulated audio content under realistic evaluation conditions.

TABLE V
FINAL EVALUATION METRICS FOR FAKE AUDIO DETECTION ON THE EVALUATION SET.

Metric	**Score (%)**
Accuracy	92.2
Precision	99.4
Recall	91.8
F1-score	95.5

The model successfully identified 392 real samples (94.9%) out of 413, while 21 were incorrectly classified as fake, according to the per-class performance displayed in Tale 6 and the confusion matrix in Figure 5. On the other hand, out of 3,587 fake samples, the model correctly identified 3,294 (91.8%) of them, with 293 being mistakenly classified as real. This finding shows that the model is strong in detecting fake audio, as reflected in the very high precision, indicating that when the model predicts an audio clip as fake, it is almost always correct. Because of the inherent variability in human speech, real audio is still more difficult to model, as indicated by the comparatively lower recall for the real class.

TABLE IV
CLASSIFICATION REPORT (PER-CLASS) FOR FAKE AUDIO DETECTION ON THE EVALUATION SET.

	Precision (%)	**Recall (%)**	**F1-score (%)**
Real (0)	57.0	95.0	71.0
Fake (1)	99.0	92.0	95.0
Accuracy			92.0

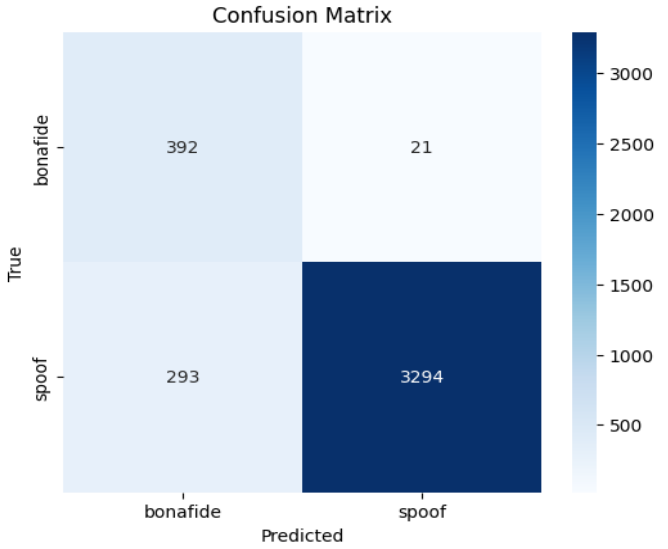


Fig. 5 Confusion matrix for fake audio detection.

The ROC curve in Figure 6, which shows a high level of class separability with a ROC-AUC of 0.972, further supports the model's discriminative ability. This suggests that the extracted Wav2Vec 2.0 embeddings effectively capture subtle acoustic inconsistencies and synthesis artifacts found in fake audio when combined with the BiLSTM and attention mechanism.

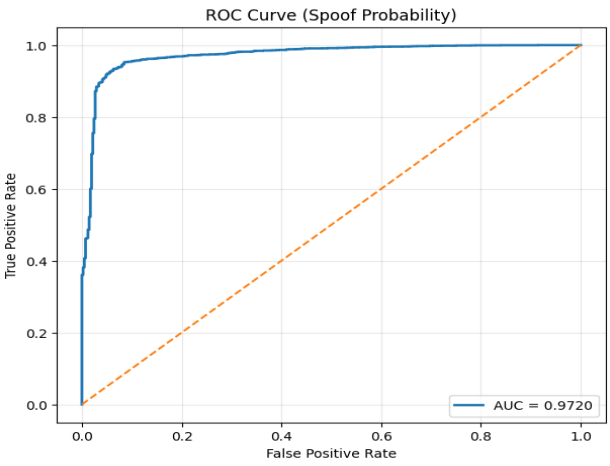


Fig. 6 ROC Curve for fake audio detection

D. System Integration and Cross-Media Verification

The fully integrated class was tested on a set of seven random samples that included a variety of media formats in order to verify the efficacy of the suggested routing architecture. To test robustness, the test set contained real-

https://doi.org/10.31436/ijpcc.v12i1.651

world media and benchmark samples from IMD2020, CASIA, DeeperForensic1.0, and ASVspoof.

```
{'modality_used': 'Image', 'prediction': 'FAKE', 'confidence': '99.81%'}
{'modality_used': 'Image', 'prediction': 'REAL', 'confidence': '71.61%'}
{'modality_used': 'Video', 'prediction': 'FAKE', 'confidence': '100.00%'}
{'modality_used': 'Video', 'prediction': 'REAL', 'confidence': '100.00%'}
{'modality_used': 'Video', 'prediction': 'FAKE', 'confidence': '99.92%'}
{'modality_used': 'Audio', 'prediction': 'REAL', 'confidence': '99.88%'}
{'modality_used': 'Audio', 'prediction': 'FAKE', 'confidence': '99.84%'}
```

Fig. 7 Sample predictions of multimodal detection

Based on Figure 7, the system successfully routed input files to the appropriate modality. This demonstrates that the predict() function's logic is dependable for mixed-media workflows. Across all modalities, the integrated system showed high levels of confidence. While synthetic content was consistently detected with confidence scores exceeding 99%, the lowest confidence recorded for a real image was 71.61%.

V. DISCUSSION

Cross-media fake news detection matters because misinformation spreads quickly and can destabilize communities. False content often carries emotional charge, creates confusion, and fuels misleading narratives that people may unknowingly amplify. Systems that can detect manipulated or misleading content across multiple modalities help reduce that risk and support healthier information ecosystems.

This paper employs a Content-Adaptive Routing Framework to tackle the problem of various multimedia forgeries. Our system operates as a unified forensic interface, in contrast to inflexible multimodal systems that require the integration of disparate data stream which often failing when a user provides only one file type. The input format (image, video, or audio) is dynamically identified by the system's routing logic, which then sends it to a specialized "expert" deep learning classifier. The system can successfully verify isolated media files without relying on paired data (e.g., requiring audio to accompany video) in part to this strategy's high availability and robustness.

Compared to monolithic fusion models, the suggested routing architecture has several engineering advantages. Because resources are only allocated to the appropriate model for a given input (for example, the heavy Video Swin model is never loaded into memory when analysing a simple JPEG), it is producing a computationally efficient system. Additionally, the design is very modular; future enhancements to the audio component, for instance, can be incorporated without requiring a full retraining of the image or video subsystems by simply updating the AudioModel class weights. The framework is a workable, scalable solution for real-world multimedia verification because of its flexibility.

A system like ours could be applied during elections, integrated into newsroom verification pipelines, or used in social-media monitoring to flag suspicious content before it gains traction.

Despite the strengths, there are limitations. Our project relies on publicly available datasets, which are relatively small and may not capture the full diversity of real-world social-media content. A larger, more varied dataset would improve generalization. Computational cost is another constraint: multimodal deep learning requires significant processing power, which can make experimentation slower and deployment more expensive.

VI. CONCLUSION

Cross-media using images, audio, and video provides a valuable technological approach for helping users avoid becoming victims of false information. This paper achieved strong performance across all three modalities, particularly in detecting *fake* content, which is typically more challenging. However, several limitations were encountered. The datasets were sourced from publicly available repositories, which may not be as current or diverse as datasets from private domains. This limits real-world representativeness. In addition, computational constraints due to budget limitations restricted the scale and complexity of the experiments. Future work can focus on reducing domain shift between controlled, lab-based datasets and real-world multimedia. Enhancing generalization in this way will help the model produce more robust results and better align with real-time, real-world data.

ACKNOWLEDGMENT

In the name of Allah, the Most Merciful and Gracious. We are grateful to Allah (SWT) for giving us the knowledge and perseverance to finish this paper. For their steadfast support, we are grateful to our parents, instructors, peers, and supervisor. May this endeavour gain Allah's blessings and benefit society. Finally, the authors would like to thank the Department of Computer Science at IIUM for supporting this paper.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

I.N. Samsul Kamal and A.S. Samsudin contributed to the design and implementation R. Hassan provided supervision, validated the methodology, and reviewed the final manuscript.

DATA AVAILABILITY STATEMENT

All datasets utilized in this paper are sourced from publicly available repositories. The IMD2020, CASIA 2.0, DeeperForensic1.0 and ASVspoof 2019 datasets can be accessed via their respective citations.

ETHICS STATEMENT

This paper utilized exclusively publicly available benchmark datasets. No private data was collected, and no human subjects or animals were involved in the experimentation process.

REFERENCES

[1]. L. Verdoliva, "Media Forensics and DeepFakes: an overview," *arXiv preprint, arXiv:2001.06564,* Jan. 2020. [Online]. Available: https://arxiv.org/pdf/2001.06564

[2]. A. Novozámský, B. Mahdian, and S. Saic, "IMD2020: A Large-Scale Annotated Dataset Tailored for Detecting Manipulated Images," in *2020 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, Snowmass Village, CO, USA, Mar. 2020, pp. 71-80, doi: 10.1109/WACVW50321.2020.9096940

[3]. *EndlessSora/DeeperForensics-1.0: [CVPR 2020] A Large-Scale Dataset for Real-World Face Forgery Detection*. (2025). GitHub. https://github.com/EndlessSora/DeeperForensics-1.0

[4]. H. Delgado, N. Evans, T. Kinnunen, K. A. Lee, X. Liu, A. Nautsch, J. Patino, M. Sahidullah, M. Todisco, X. Wang, and J. Yamagishi, "ASVspoof 2021: Automatic speaker verification spoofing and countermeasures challenge evaluation plan," *arXiv preprint arXiv:2109.00535*, Sept. 2021. [Online]. Available: https://arxiv.org/abs/2109.00535.

[5]. B. Dolhansky *et al.*, "The DeepFake Detection Challenge (DFDC) dataset," *arXiv preprint arXiv:2006.07397*, Jun. 2020. [Online]. Available: https://arxiv.org/abs/2006.07397

[6]. Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *arXiv preprint arXiv:2103.14030*, 2021. [Online]. Available: http://arxiv.org/abs/2103.14030

[7]. *EndlessSora/DeeperForensics-1.0: [CVPR 2020] A Large-Scale Dataset for Real-World Face Forgery Detection*. (2025). GitHub. https://github.com/EndlessSora/DeeperForensics-1.0

[8]. Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2021.

[9]. B. Singh and D. K. Sharma, "Predicting image credibility in fake news over social media using multi-modal approach," *Neural Computing and Applications*, vol. 34, no. 24, pp. 21503–21517, 2021. https://doi.org/10.1007/s00521-021-06086-4

[10]. Y. Almsrahad and N. M. Charkari, "Image Fake News Detection using Efficient NetB0 Model," *Journal of Information Systems and Telecommunication (JIST)*, vol. 12, no. 45, pp. 41–48, 2024. https://doi.org/10.61186/jist.40976.12.45.41

[11]. L. Y. Gong, X. J. Li, and P. H. J. Chong, "Swin-Fake: A Consistency Learning Transformer-Based Deepfake Video Detector," *Electronics*, vol. 13, no. 15, p. 3045, 2024. https://doi.org/10.3390/electronics13153045

[12]. S. R. Mishra, H. Mohapatra, S. A. Edalatpanah, and M. K. Gourisaria, "Advanced deepfake detection leveraging swin transformer technology," *Engineering Review*, vol. 44, no. 4, pp. 45–56, 2024. https://doi.org/10.30765/er.2583

[13]. J. M. Martín-Doñas and A. Álvarez, "The Vicomtech Audio Deepfake Detection System based on Wav2Vec2 for the 2022 ADD Challenge," *arXiv preprint arXiv:2203.01573*, 2022. [Online]. Available: https://arxiv.org/abs/2203.01573

[14]. S. Dilbar, M. A. Qureshi, S. K. Noon, and A. Mannan, "AudioFakeNet: A Model for Reliable Speaker Verification in Deepfake Audio," *Algorithms*, vol. 18, no. 11, p. 716, 2025. https://doi.org/10.3390/a18110716

[15]. F. Khalid, M. H. Akbar, and S. Gul, "SWYNT: Swin Y-Net Transformers for Deepfake Detection," in *2023 International Conference on Robotics and Artificial Intelligence (ICRAI)*, 2023, pp. 1–6. https://doi.org/10.1109/icrai57502.2023.10089585

[16]. Wodajo, D., Atnafu, S., & Akhtar, Z. (n.d.). "Deepfake Video Detection Using Generative Convolutional Vision Transformer," *arXiv preprint arXiv:2307.07036*, 2023. [Online]. Available: https://arxiv.org/pdf/2307.07036

[17]. Y. Zhou, Q. Ying, Z. Qian, S. Li, and X. Zhang, "Multimodal Fake News Detection via CLIP-Guided Learning," *arXiv preprint arXiv:2205.14304*, 2022. [Online]. Available: https://arxiv.org/abs/2205.14304

[18]. Ma, Z., Luo, M., Guo, H., Zeng, Z., Hao, Y., & Zhao, X. (2024). "Event-Radar: Event-driven Multi-View Learning for Multimodal Fake News Detection," *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5809–5821. https://doi.org/10.18653/v1/2024.acl-long.316

[19]. Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, and H. Hu, "Video Swin Transformer," *arXiv preprint arXiv:2106.13230*, 2021. [Online]. Available: https://arxiv.org/abs/2106.13230

[20]. Y. Sun, X. Li, J. Wang, L. He, and X. Liu, "Audio Anti-Spoofing Based on Audio Feature Fusion," *Algorithms*, vol. 16, no. 7, p. 317, 2023. https://doi.org/10.3390/a16070317

[21]. M. Li and X.-P. Zhang, "Interpretable Temporal Class Activation Representation for Audio Spoofing Detection," in *Interspeech 2024*, 2024. [Online]. Available: https://arxiv.org/abs/2406.08825

[22]. X. Liu, W. Ge, X. Wang, and J. Yamagishi, "LENS-DF: Deepfake Detection and Temporal Localization for Long-Form Noisy Speech," in *IJCB 2025*, 2025. [Online]. Available: https://arxiv.org/abs/2507.16220

[23]. Sivaraman, D. K., Saif, M., MR, M. F., & Moosa, M. (2025). Enhanced Fake Image Localization in Social Media using Swin Transformer and EfficientNet Feature Fusion. *International Journal for Research in Applied Science and Engineering Technology*, *13*(4), 4052–4058. https://doi.org/10.22214/ijraset.2025.69194

[24]. S. A. Khan and D.-T. Dang-Nguyen, "Deepfake Detection: Analysing Model Generalisation Across Architectures, Datasets and Pre-Training Paradigms," *IEEE Access*, vol. 12, pp. 1880–1908, 2024. https://doi.org/10.1109/access.2023.3348450

[25]. A. Novozámský, B. Mahdian, and S. Saic, "IMD2020: A Large-Scale Annotated Dataset Tailored for Detecting Manipulated Images," in *2020 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2020, pp. 71–80. https://doi.org/10.1109/wacvw50321.2020.9096940

[26]. D. Goel, "CASIA 2.0 Image Tampering Detection Dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset

[27]. EndlessSora, "DeeperForensics-1.0," *GitHub repository*, 2025. [Online]. Available: https://github.com/EndlessSora/DeeperForensics-1.0/tree/master/dataset

[28]. D. Wan, M. Cai, S. Peng, W. Qin, and L. Li, "Deepfake Detection Algorithm Based on Dual-Branch Data Augmentation and Modified Attention Mechanism," *Applied Sciences*, vol. 13, no. 14, p. 8313, 2023. https://doi.org/10.3390/app13148313

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.644

Berita Debunked: Real-time Fake News Detection and Alert System

Ahmad Faisal Daniell bin Mohd Yusoff, Aiman Kamil bin Zainuddin, Raini binti Hassan*

Department of Computer Science, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia

*Corresponding author: hrai@iium.edu.my
(Received: 5th December 2025; Accepted: 22nd December, 2025; Published on-line: 30th January, 2026)

***Abstract*—** BeritaDebunked is an AI-driven near real-time fake news detection and alert system designed to combat misinformation in Malaysia, particularly on platforms such as WhatsApp. The system combines natural language processing and multimodal deep learning by using BERT for textual analysis and BLIP-2 for image–text evaluation. Deployed as a browser extension, it flags suspicious messages and allows continuous model updates through a scalable backend. Evaluation on the Fakeddit benchmark dataset demonstrates that the proposed hybrid architecture achieves an accuracy of (83.3%), with a precision of (82.6%) and an F1-score of (84.9)%. While unimodal text baselines achieved marginally lower raw accuracy (82.9%), the hybrid model demonstrates superior robustness in detecting multimodal context mismatches. The system demonstrates real-time capability with an average inference latency of 56.42 ms. By enabling timely detection and user-friendly alerts, BeritaDebunked aims to support digital literacy efforts, reduce the spread of misinformation, and contribute to Sustainable Development Goal 16 by strengthening information integrity.

Keywords— Hybrid, Fake news, SDG 16, BERT, BLIP-2, Multimodal deep learning, NLP

I. Introduction

Due to the widespread use of social media and messaging platforms, Malaysia is facing a challenge in controlling the rapid spread of fake news, particularly on WhatsApp, Facebook and Twitter. The process of manually verifying the authenticity of forwarded messages is difficult and unfeasible due to high volume and speed. Although initiatives by Malaysia Communication and Multimedia Commission (MCMC) Sebenarnya.my exist, the current approach remains slow, reactive and unable to prevent early public impact [1] [2].

The research also focuses on evaluating and comparing different machine learning models for fake news detection. The algorithms and results are presented and compared in a detailed yet concise manner using multiple evaluation metrics to identify the most reliable approach.

To address this problem, this project proposes the development of a real-time AI-driven fake news detection and alert system designed specifically for WhatsApp, leveraging natural language processing (NLP), machine learning, and deep learning techniques including models such as BERT for text classification and BLIP-2 for multimodal content analysis. The system prioritizes user privacy by analysing only message content and excludes personal metadata or private chat logs. It is built using Python, Flask/FastAPI, and a front-end browser extension, the platform ensures accessibility, scalability, and practical deployment for public use.

Despite its potential, developing such a system introduces several challenges, including privacy concerns due to WhatsApp's end-to-end encryption, compliance with Malaysia's Personal Data Protection Act (PDPA), and the need for fast, real-time performance supported by scalable infrastructure. Ethical issues such as algorithmic bias, transparency, and responsible alerting must also be addressed to prevent user distrust or over-reliance. The scarcity of labelled WhatsApp datasets further complicates model training, while risks of false positives, maintenance requirements, and evolving platform features pose additional hurdles.

Overall, the development of this AI-powered real-time detection system is critical for protecting information integrity, enhancing public digital literacy, and supporting organisations such as MCMC, MyCERT, fact-checkers, and media outlets in combating misinformation. This project also sets a technological precedent for misinformation detection on encrypted platforms, aligning with the goals of SDG 16 by promoting peace, justice, and strong institutions in the digital age.

This paper makes three main contributions: (1) a hybrid BERT+BLIP-2 multimodal model trained on Fakeddit for fake news detection; (2) an end-to-end architecture that integrates the model into a browser extension for real-time alerts; and (3) a comparative evaluation against unimodal baselines using standard metrics on a multimodal benchmark dataset.

https://doi.org/10.31436/ijpcc.v12i1.644

II. Literature Review

The proposed paper draws directly from the reviewed literature to establish an effective framework for real-time fake news detection and alerting across text and images. Insights from both unimodal and multimodal models have been adapted to address core limitations and gaps that are identified in prior studies, ensuring the system is methodologically robust, scalable and aligned with current research trends. *Evolution of Detection Models: From Unimodal to Multimodal*

Early research into fake news detection predominantly focused on unimodal approaches, utilizing machine learning and deep learning to analyze textual features. A few studies employing models like BERT, LSTM, and XGBoost have achieved exceptional accuracy rates. For instance, Cavus et al. [3] and Sharma et al. [4] demonstrated that semantic analysis can effectively identify false narratives, reaching accuracy levels up to 99.9%. However, these models face significant limitations in real-world applications. Unimodal systems are blind to visual context, rendering them ineffective against multimedia misinformation. Furthermore, approaches such as those by Rashad et al. [5] and Limbachia [6] rely on query-based inputs or domain-specific training (e.g., COVID-19 data), limiting their generalizability and scalability. To address these deficiencies, recent scholars have shifted toward multimodal architecture that processes both text and images.

To address these deficiencies, recent scholarship has shifted toward multimodal architectures that process both text and images. Advanced hybrid frameworks have emerged to tackle this complexity. Yan et al. [7] utilized BERT and BLIP-2 as feature extractors, integrating them through a 1D-CCNet attention mechanism and Heterogeneous Cross-Feature Fusion Method (HCFFM). This approach demonstrated the superior capability of BLIP-2 in capturing visual semantics compared to traditional CNNs. Similarly, Ojo et al. [8] employed a BiLSTM + VGG19 architecture, achieving 97.2% accuracy. While these systems demonstrate strong performance, current research is often hindered by high model complexity, small datasets, and class imbalances [9], [10]. Additionally, most existing multimodal models are restricted to image-text pairs and struggle with cross-domain generalization.

A. *Comparison of Existing Fake News Detection Tools*

Beyond academic models, several consumer-facing systems attempt to mitigate misinformation, though they rely largely on source-level credibility rather than real-time content analysis. NewsGuard [11] and Media Bias/Fact Check (MBFC) [12] operate primarily as browser extensions that rate the reliability of news domains. NewsGuard employs human analysts to grade sites based on journalistic criteria, while MBFC categorizes sources by political bias and factual reporting. While valuable for digital literacy, their source-level approach is a critical limitation. They cannot flag individual false articles hosted on generally credible sites, nor can they assess viral content on encrypted platforms. Furthermore, their reliance on human curation introduces subjectivity and scalability issues, with ratings often criticized for being US-centric or potentially biased.

In contrast, ClaimBuster [13] utilizes NLP to detect check-worthy factual claims in real-time. While it automates the detection process, surpassing the speed of human evaluators, it remains limited to textual content. It relies heavily on matching claims against existing fact-checking databases, meaning it often fails to detect novel misinformation or nuanced context-dependent falsehoods that involve imagery.

B. *Synthesis and Research Gap*

Literature reveals a distinct gap in current countermeasures. Unimodal models [3]- [6] lack of visual context while existing multimodal architectures, such as attention-heavy mechanisms proposed by Yan et al. [7], often prioritize architectural novelty over the latency requirements of real-time detection systems and struggles with generalization and deployment scalability. Furthermore, commercial tools either prioritize source reputation over content analysis [11]- [12] or ignore multimedia entirely [13].

Consequently, there is a critical need for a hybrid, real-time detection system capable of analyzing both text and images within encrypted environments. Building upon the robust feature extraction capabilities established by Yan et al. [7], the proposed system integrates BERT for deep semantic text analysis and BLIP-2 for visual reasoning. However, unlike prior complex fusion methods, this paper employs a direct concatenation and dense classification approach to balance high accuracy with privacy-preserving, real-time performance required for platforms like WhatsApp.

III. Methodology

This paper proposes a hybrid multimodal deep learning framework designed to detect misinformation by analysing both textual and visual components of a message. The system integrates state-of-the-art pre-trained models, BERT for text and BLIP-2 for images into a unified classification pipeline.

A. *Dataset*

The core dataset that will be used is Fakeddit, a publicly available dataset multimodal fake news dataset that includes both text and images, along with multi-class labels

https://doi.org/10.31436/ijpcc.v12i1.644

representing different levels of truthfulness [14]. The system utilizes this dataset to learn the patterns and features that distinguish real news from fake news across multiple modalities. Although the model is evaluated on the Fakeddit benchmark, which captures multimodal news content, WhatsApp messages in Malaysia may differ in style, language, and media usage. Therefore, the current results should be interpreted as an initial validation of the architecture, with future work focusing on collecting or adapting datasets that more closely reflect local WhatsApp communication patterns.

B. *Experimental Configuration and Reproducibility*

To ensure the reproducibility of our results, all experiments were conducted using a fixed random seed (seed=42) for both data splitting and model initialization. The dataset was partitioned into training (80%) and testing (20%) sets using stratified random sampling to preserve the class distribution of the original Fakeddit dataset. The hybrid model was implemented using PyTorch and the Hugging Face Transformers library. We utilized the AdamW optimizer and a batch size of 16 to fit within the memory constraints of a standard NVIDIA T4 GPU. No additional class balancing techniques (such as SMOTE or weighted loss) were applied.

C. *Proposed System Architecture*

System Analysis and Design Diagram are essential tools for meddling, understanding, and communicating the structure and behaviour of this system [15]. They will visually map the information to support specific goals and enhance cognitive processing during task performance. Effective diagrams such as system architectural diagrams are well known. The architectural diagrams provide a high-level overview of system components and their interactions, supporting communication, design and maintenance [16].

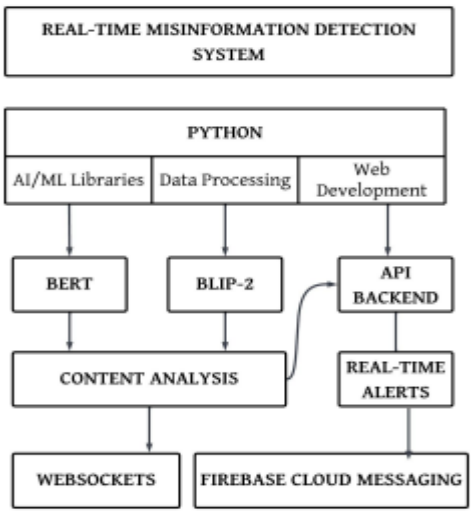


Fig. 1 System Architecture Design for BeritaDebunked

The architectural framework of the proposed system is illustrated in Figure 1. The core processing layer is built upon a Python-based backend that orchestrates the multimodal analysis. Incoming data from the browser extension is routed to the Content Analysis module, where the BERT and BLIP-2 models operate in parallel to extract linguistic and visual features, respectively. These features are fused to generate a credibility score, which is then transmitted via a high-performance FastAPI backend. To ensure real-time responsiveness, the system leverages WebSockets for low-latency communication, delivering immediate verification alerts to the user interface while asynchronously caching results in the Firebase cloud database for scalability.

IV. Results And Discussion

The developed system integrates BERT for text analysis and BLIP-2 for multimodal understanding within a Python-based framework. BERT enables nuanced detection of sentiment and bias in textual claims, while BLIP-2 analyses image-text alignment to identify inconsistencies characteristic of manipulated media. This dual-model approach addresses a critical gap in existing tools, which often rely on single-modality analysis.

While the full Fakeddit dataset contains over one million samples, this paper utilized a focused subset of 30,000 samples to balance training time with statistical confidence. Research by standard deep learning benchmarks indicates that validation sets exceeding 6,000 samples are sufficient to achieve model convergence and reliable performance estimates. Consequently, the results reported in this paper offer a high degree of confidence regarding the system's real-world applicability.

The dataset is split into training (80%) and testing (20%) subsets. This 80:20 ratio was selected as a standard convention in machine learning to maintain a balance between sufficient data for the model to learn complex multimodal feature representations and a large enough unseen validation set to rigorously test generalizability and prevent overfitting.

The qualitative analysis of the hybrid model reveals distinct behavioural advantages over unimodal approaches. While the BERT component successfully flags sensationalist text typical of 'clickbait' news, it struggles with posts where the text is neutral, but the accompanying image provides a misleading context. The integration of BLIP-2 addresses this semantic gap by generating image captions that are cross-referenced with the textual claims. This multimodal fusion allows the system to detect mismatch where the visual evidence contradicts the textual narrative a key indicator of sophisticated misinformation that text-only models often miss. This behaviour suggests that future improvements should focus on fine-tuning the cross-modal attention mechanisms rather than simply increasing dataset size.

https://doi.org/10.31436/ijpcc.v12i1.644

Performance benchmarking confirms the system's suitability for real-time deployment. On a standard T4 GPU environment, the model achieved an average inference latency of 56.42 ms per message with a throughput of 17.72 messages/second, well within the latency tolerance for instant messaging applications.

Table 1 reveals that the unimodal BERT model achieved an individual performance (82.9% accuracy), indicating strong textual cues in the dataset. However, the Hybrid model demonstrated competitive performance (83.3% accuracy), significantly outperforming the BLIP-2 baseline. While slightly higher than the unimodal text baseline, this trade-off is justified by the hybrid model's ability to detect multimodal context mismatches. This result highlights that while text remains the primary indicator of credibility in this dataset, the Hybrid architecture successfully integrates visual context with minimal loss in accuracy, providing a more holistic detection mechanism than text-only approaches.

TABLE I
RESULT COMPARISON OF MACHINE LEARNING MODELS

Modality	Machine Learning Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Unimodal	BERT	82.9	86.5	86.2	83.6
Multimodal	BLIP-2	60.7	60.6	70.9	65.4
	Hybrid (BERT+BLIP-2)	83.30	82.62	81.8	84.9

Figure 2 illustrates the Hybrid BERT+BLIP-2 model's training progression over 2 epochs. The validation accuracy stabilizes at approximately 83.3%, while the validation loss decreases to 0.39, indicating that the model successfully learned generalizable patterns without overfitting. The convergence of training and validation loss confirms the stability of the fine-tuning process on the 30,000-sample subset.

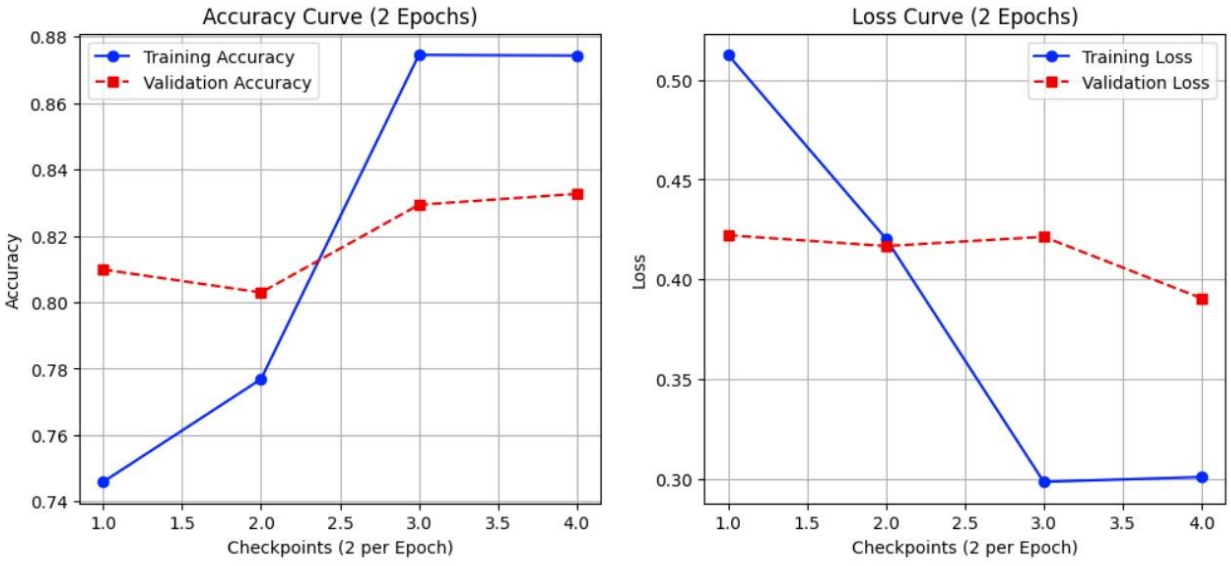


Fig. 2 Training and Validation Performance Curves for BERT+BLIP-2 model

It is important to note that these preliminary results are derived from a subset of the Fakeddit dataset due to computational resource constraints. While the current sample size is sufficient to validate the hybrid architecture's logic, scaling the training process to the full multi-terabyte dataset in future iterations is expected to further improve the model's precision and recall stability.

As detailed in

TABLE , the proposed hybrid model's accuracy (83.3%) presents a realistic performance baseline when balanced against the constraints of real-time deployment. While prior

https://doi.org/10.31436/ijpcc.v12i1.644

unimodal studies such as Cavus et al. [3] and Sharma et al. [4] reported accuracies exceeding 99%, these models were often trained on small, topic-specific datasets (e.g., COVID-19), which limits their ability to generalize to the broad-spectrum misinformation found on social media.

Similarly, in the multimodal domain, architectures like Yan et al. [7] achieved higher accuracy (92.5%) but relied on computationally intensive attention mechanisms (1D-CCNet) that are unsuitable for browser-based extensions. In contrast, the slightly lower accuracy observed in this paper reflects the trade-off required to achieve near real-time latency and privacy preservation. Unlike Ojo et al. [8] whose high-accuracy model operates offline, the proposed system successfully integrates verification into the user's workflow, prioritizing immediate impact and accessibility over raw metric maximization on a noisy benchmark like Fakeddit.

TABLE II.
COMPARATIVE ANALYSIS

Author, Year	Modality	Methodology	Dataset	Performance	Key Limitation
Cavus et. al. [3]	Unimodal (Text)	CRIPS-DM, BERT, MS Azure	COVID-19 News	Acc. up to 99.9%	Domain-specific training
Babar et. al. {18}		Hybrid N-Gram + LSTM	Social media	Acc. 96.5%	High computation cost
Sharma et. al. {4}		XGBoost, LSTM	News Dataset	Acc. up to 99.9%	Small dataset
Rashad et. al. [5]		TF-IDF Random Forest, Logistic Regression, LSTM	News Dataset	Acc. up to 99.8%	Query-based input
Limbachia [6]		Random Forest	News Dataset	Acc. 100%	Poor generalization
Yan et. al. (2024)	Multimodal	BERT + BLIP-2 (Model encoders text & images extractor) 1D-CCNet Attention Mechanism Heterogeneous Cross-Feature Fusion Method (HCFFM)	Multimodal News	Acc. 92.5–96.7%	Weak cross-domain support
Segura-Bedmar et. Al. [9]		CNN + BiLSTM	Multimodal News	Acc. 87%	Small dataset
Ojo et. al. [8]		BiLSTM + VGG19	Social media	Acc. 97.2%	No audio/video support
Saha [17]		DeBERTa + ConvNeXT	Multimodal News	Acc. 91.2%	Image-text only
Dellys et. al. [19]		ViLBERT + SVM	Multimodal News	Acc. 77%	Class imbalance
Proposed System		BERT + BLIP-2	Fakeddit	Acc. 83.3%	Computational constraint

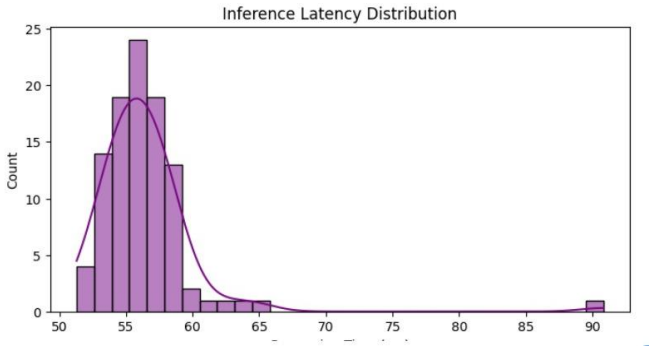


Fig. 1 Inference Latency Distribution and Throughput Analysis.

To validate the system's real-time capabilities, we conducted a latency stress test on a standard NVIDIA T4 GPU environment. As shown in Figure 3, the system achieved an average inference latency of 56.42 ms per message, with a 95th percentile lag of 60.11 ms. The system demonstrated a throughput of 17.72 messages per second.

These results contradict initial concerns regarding the computational overhead of the BLIP-2 component. With an average response time well below the 100ms threshold often cited for perceived instantaneity, Berita Debunked successfully meets the 'near real-time' requirement for interactive user verification workflows. Consequently, the system is characterized as offering prototype-level responsiveness suitable for user verification workflows, rather than high-frequency automated filtering.

For deployment, the backend architecture leverages asynchronous processing capabilities inherent in the FastAPI framework to handle concurrent requests efficiently.

Furthermore, the integration of Firebase Cloud Messaging ensures that alert delivery is decoupled from the heavy model inference process, preventing bottlenecks during high-traffic periods. Future work will focus on optimizing the BLIP-2 backbone to further reduce computational overhead.

V. Conclusions

To summarize, this paper successfully developed and validated a near real-time prototype, multimodal fake news detection system designed to combat misinformation on encrypted platforms. By integrating BERT for textual analysis and BLIP-2 for visual-semantic reasoning, the proposed hybrid model achieved a classification accuracy of 83.3% on a robust test set of 6,000 samples from the Fakeddit dataset. These results demonstrate that combining linguistic and visual features provides a more holistic verification mechanism than unimodal approaches, capable of identifying multimedia content in near real-time while maintaining user privacy through a browser-based extension architecture. However, this paper acknowledges certain limitations. The reliance on the Fakeddit benchmark as a proxy for WhatsApp messages introduces a domain shift, as the linguistic style of Reddit posts differs from private messaging patterns in Malaysia. Additionally, computational constraints necessitated the use of a 40,000-sample subset of the dataset, which, while statistically significant, does not capture the full variance of the complete dataset.

Lastly, the future work will focus on bridging this domain gap by fine-tuning the model on localized, anonymized Malaysian datasets to better capture regional dialects and specific forwarding behaviours. Furthermore, the project will prioritize the optimization and public deployment of the WhatsApp Web extension. This strategic focus acknowledges the current technical limitations of mobile operating systems in supporting real-time message interception, positioning the browser-based solution as the most viable path for immediate, scalable impact in combating misinformation.

Acknowledgment

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author(s) Contribution Statement

A.F.D. Mohd Yusoff contributed to the conceptualization, methodology, software development, and writing of the original draft. A.K. Zainuddin was responsible for data curation, validation, visualization, and reviewing and editing the manuscript. R. Hassan provided supervision and oversaw the project.

Data Availability Statement

The data that support the findings of this paper are openly available in the Fakeddit repository at [https://github.com/entitize/Fakeddit]. This dataset is a public benchmark for multimodal fake news detection. The source code and scraped samples used for the system demonstration are available from the corresponding author upon reasonable request.

Ethics Statement

This study did not require ethical approval.

References

[1]. A. Mat Isa, A. Z. H. Samsudin and M. R. Hendrawan, "Dissemination of Fake News and Information Disorder in Malaysia: A descriptive analysis," Environment-Behaviour Proceedings Journal, vol. 7, no. SI10, p. 53–58, 2022. doi: 10.21834/ebpj.v7isi10.4101.

[2]. "Kementerian Komunikasi," 19 May 2025. [Online]. Available: https://www.komunikasi.gov.my/awam/berita/23980-fake-news-spreaders-deserve-heavier-penalty. [Accessed Nov. 2025].

[3]. N. Cavus, M. Goksu and B. Oktekin, "Real-time fake news detection in online social networks: FANDC Cloud-based system," Scientific Reports, vol. 14, no. 1, 2024. doi: 10.1038/S41598-024-76102-9.

[4]. S. Sharma, M. Saraswat and A. K. Dubey, "Fake News Detection Using Deep Learning. Communications in Computer and Information Science," Communications in Computer and Information Science, vol. 1459, p. 249–259, 2021. doi: 10.1007/978-3-030-91305-2_19.

[5]. M. Rashad, N. Khalid, A. Hamza, S. Javed and K. B. Majeed, "A Semantic Fake News Detection System Using Machine Learning Classifier," Kashf Journal of Multidisciplinary Research, vol. 1, no. 12, p. 264–279, 2024. doi: 10.71146/KJMR171.

[6]. D. Limbachia, "Real-time Fake News Detection System Using AI," International Journal for Research in Applied Science & Engineering Technology, vol. 13, no. III, p. 560–565, 2025. doi: 10.22214/ijraset.2025.67294.

[7]. Y. Yan, H. Fu and F. Wu, "Multimodal Social Media Fake News Detection Based on 1D-CCNet Attention Mechanism," Electronics, vol. 13, no. 18, 2024.

[8]. A. O. Ojo, F. Najar, N. Zamzami, Z. T. Himdi and N. Bouguila, "SmoothDetector: A Smoothed Dirichlet Multimodal Approach for Combating Fake News on Social Media," IEEE Access, vol. 13, pp. 39289-39305, 2025. doi: 10.3390/electronics13183700.

[9]. I. Segura-Bedmar and S.-B. Alonso-Bartolome, "Multimodal Fake News Detection," Information, vol. 13, no. 6, p. 284, 2022. doi: 10.3390/INFO13060284.

[10]. H. Dellys, Mokeddem, Halimal and L. Sliman, "On the Integration of Social Context for Enhanced Fake News Detection Using Multimodal Fusion Attention Mechanism," AI, vol. 6, no. 4, 2025. doi: 10.3390/ai6040078.

[11]. "NewsGuard: Global Leader in Information Reliability.," [Online]. Available: https://www.newsguardtech.com/.

[12]. "Media Bias / Fact Check (MBFC)," [Online]. Available: https://chromewebstore.google.com/detail/media-bias-fact-

check/ganicjnkcddicfioohdaegodjodcbkkh?utm_source=item-share-cb.

[13]. "ClaimBuster: Automated Live Fact-checking," [Online]. Available: https://idir.uta.edu/claimbuster/.

[14]. K. Nakamura, S. Levy and W. Yang Wang, "r/Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection," LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings, 2020.

[15]. S. Kumari, "Visual Modeling: Unlocking Ideas and Enhancing Understanding: The Power of Visual Modeling," International Journal of Engineering & Technology, vol. 12, no. 2, pp. 20-25, 2023. doi: 10.14419/ijet.v12i2.32334.

[16]. M. Malinova and J. Mendling, "Cognitive Diagram Understanding and Task Performance in Systems Analysis and Design," MIS Quarterly, vol. 45, no. 4, pp. 2101-2158, 2021. doi: 10.25300/misq/2021/15262.

[17]. K. Saha, "DeBERTNeXT: A Multimodal Fake News Detection Framework," Lecture Notes in Computer Science, vol. 14074, p. 348–356, 2023. doi: 10.1007/978-3-031-36021-3_36.

[18]. M. Babar, A. Ahmad, M. U. Tariq and S. Kaleem, "Real-Time Fake News Detection Using Big Data Analytics and Deep Neural Network.," IEEE Transactions on Computational Social Systems, vol. 11, no. 4, p. 5189–5198, 2024. doi: 10.1109/TCSS.2023.3309704.

[19]. S. K. Hamed, M. J. Ab Aziz and M. R. Yaakub, "Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments.," Sensors, vol. 23, no. 4, 2023. doi: 10.3390/s23041748.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.642

A Conceptual Framework for a Lightweight AI System for Skin Disease Risk Prediction Using Epidemiological Data in Rural Bangladesh

Mohammad Raihanul Islam[1], Andi Fitriah binti Abdul Kadir[1], Syazwan Aizat Ismail[2]
[1]*Department of Computer Science, International Islamic University Malaysia,* Kuala Lumpur, 53100, Malaysia
[2]*National Poison Centre, University Sains Malaysia, Penang, Malaysia.*

*Corresponding author: raihanulmcse@gmail.com
(Received: 1st December 2025; Accepted: 9th January, 2026; Published on-line: 30th January, 2026)

Abstract— Skin disease remains a significant public health issue in rural Bangladesh, where limited access to dermatologists and inadequate diagnostic facilities often delay accurate assessment and treatment. To address these constraints, this conceptual paper presents a lightweight AI-based framework for predicting skin disease risks using structured epidemiological data gathered from hospital visits and interviews with patients and healthcare staff. The framework incorporates environmental, occupational, hygiene-related, and living-condition factors to model individual risk profiles. Preliminary experiments conducted on an existing dataset demonstrate that conventional machine learning algorithms, particularly K-Nearest Neighbors (KNN) and Random Forest, achieve strong predictive performance, with accuracy reaching up to 88% in train–test evaluations and 80% in 10-fold cross-validation. These results confirm the viability of achieving high diagnostic reliability without image-based tools, relying solely on patient and environmental attributes. The findings further support the practical feasibility of deploying the proposed model in resource-limited rural clinics to aid early risk identification and more efficient allocation of healthcare resources. Privacy protection is incorporated as a core component to ensure secure and ethical handling of patient information.

Keywords— Skin disease risk prediction, epidemiology, lightweight AI, rural healthcare, machine learning.

I. Introduction

Skin diseases are a major health concern in Bangladesh, especially in rural communities. Based on observations from local hospitals, clinics, health centers, and the patient's scenarios, these conditions are very common. People living in crowded areas, dirty environments, with poor water sources, low income, and poor sanitation are more affected compared to those living in developed areas. Similarly, frequent chemical exposure, irregular bathing and laundry practices, poor household environments, shared clothing, limited use of soap, flood-prone areas, and household pets also contribute to the increase in skin problems. In some areas near rivers and ponds, where there is no tube well or clean water facilities,-people are fully dependent on these water sources and often use them directly for bathing and washing. They also share clothing and bedding and face difficulties in accessing medical care when symptoms appear. In many communities, skin problems become not only a health issue but also an essential part of daily life. According to reports from community health camps, skin issues represent about 15–20% of all patient visits. Furthermore, it is estimated that more than 60% of people experience some form of skin problem during their lifetime [1], [2]. Skin conditions like scabies, tinea, vitiligo, urticaria, acne, ringworm, and impetigo, eczema are very common, particularly among children and senior citizens. Chronic conditions such as eczema, urticaria, tinea, and psoriasis frequently persist in adulthood, leading to long-term health and social challenges [3], [4].

A large number of families live in underserved areas in Bangladesh, typically in small and crowded houses. It is not unusual to find more than six people sleeping under one roof. When people live close proximity, skin problems can spread easily from one person to another. Children often share the same bed and wear each other's clothes, which also contribute to the transmission of disease. In addition, many people bathe in the same pond or canal. For this reason, contagious diseases such as scabies re-occur every year [3]. The key—factors are closely linked to the environment and unhygienic lifestyle of farming communities [5].

In addition, frequent floods and high humidity also create perfect conditions to create germs. Combined with poor drainage systems, stagnant water, and animal waste, diseases spread even more rapidly [6]. A persistent lack of

safe water and sanitation makes these problems more critical. Few households have access to safe piped water or toilets, and among poor families and those living in flood-prone areas, access is minimal [1] Knowledge about hygiene within community is also very limited, particularly among women, who are primarily responsible for household cleaning and related tasks. This gap, partially responsible for educational and cultural challenges, contributes to the continued spread of diseases.

The healthcare system in Bangladesh is gradually improving, but underserved areas such as villages, continue to face significant challenges. Usually, a single dermatology specialist is responsible for many patients. A report has shown that the majority of people living outside big cities do not have access to skin specialists, or proper patient records in hospital [7]. Villagers often seek hospital care only when conditions become severe. Approximately one-third of skin problems are treated without expert practitioners, relying instead on village pharmacies or traditional methods. However, misinformation, violations of medical guidelines and self-medication increase the risk of both contiguous and incontiguous diseases.

Advancements in healthcare, particularly those driven by artificial intelligence (AI), have shown significant progress in addressing gaps within the healthcare sectors. As a result, false information, weak health regulations, and self-medication increase the risk of long-term illness and drug resistance. New digital health tools, especially those leveraging AI, have the potential to mitigate these issues and improve access to appropriate care.

Within the domain of dermatology, the utilization of deep learning algorithms and image-based models has significantly improved the accuracy of disease detection, classification, and prioritization in settings with ample resources [8], [9]. Updated technologies are developed using large sets of high-quality skin images. Importantly, they can perform as well as, or even better than, experienced dermatologists. However, they rely on expensive technology, steady electricity, skilled personnel, and reliable internet access. Consequently, people in poor and undeserved areas of Bangladesh often cannot access these advanced solutions.

The most effective approach, therefore, is to leverage readily accessible epidemiological data variables such as age, gender, occupation, household structure, hygiene practices, water availability, and population density, and scrutinize these using interpretable, artificial intelligence frameworks. Supervised machine learning methodologies can yield actionable risk assessments based on structured data variables. This approach aligns with the World Health Organization's concept of social determinants of health which emphasizes that communities' health is shaped by their personal income, household environment, and social conditions. The primary goal of this framework is to make healthcare more accessible, user-friendly, build trust, and strengthen local skills enabling artificial intelligence to better support global health [11].

Although the sector continues to undergo reforms, these evident constraints in the real world have motivated the present research to design a model that is closely aligned with the community's actual conditions and capable of practical grassroots implementation.

II. Literature Review

Several epidemiology studies have documented that skin disease is a burden in Bangladesh. A clinic-based study by [2] reported that about 58% of patients with skin disease have fungal infections. In comparison, scabies and contagious diseases occur in more than 20% of cases, together with bacterial and viral infections. During the rainy season, disease and infections increase due to poor-quality water. The highest-risk populations are children and seniors [5].

Institutional field studies have reported that the prevalence of scabies is between 18% and 34% among students. This is mostly due to living in overcrowded houses and poor clothing hygiene [6]. Transmission is more rapid among family members, suggesting that social and environmental variables are important for skin issues [3].

Common chronic skin diseases like eczema, urticaria, and psoriasis cause both physical and psychological morbidity. These types of problems are connected with mental stress, depression, and reduced professional or educational performance [4]. In Bangladesh, chronic skin disease has been reported to affect approximately 10% to 20% of people. Severity worsens due to environmental and individual factors such as climate change, poor diet, family stress, and delays in obtaining a proper medical treatment [10].

Both infectious and non-infectious skin diseases can be identified within the Social Determinants of Health (SDH) model. Social and structural factors such as education, occupation, income, living environment, and access to healthcare play an important role in determining whether people are healthy or sick [12]. Other studies have shown that water facilities, hygiene, household gatherings, and contact with animals are strong predictors of skin disease, even after adjustments for age and gender [1].

In the past five years, many scholars have used artificial intelligence (AI) to predict disease risks with community and survey data. This approach is very effective in regions such as Africa and Asia, where it is still limited to image datasets and digital health records [13], [14]. For example, one study of [15] proposed a supervised classification method using health-related and climatic factors to predict skin problems utilizing KNN, SVM, and Random Forest. Their findings show

https://doi.org/10.31436/ijpcc.v12i1.642

that epidemiology-based prediction is feasible without using clinical image data. Most of those models used tabular data rather than clinical skin images. They are designed to be lightweight, interpretable, and suitable for use in public health programs.

AI-based analysis of community and survey data still necessitates consideration of fundamental ethical principles, such as informed permission, confidentiality, and appropriate data governance in public health programs, even though this work does not specifically address privacy-preserving strategies [15].

Several image-based deep learning systems have classified skin lesions with extremely high accuracy. For instance, [9] trained deep CNNs to achieve dermatologist-level performance on dermoscopic pictures, whereas [15], [8] employed hybrid models, ResNet, and DenseNet to get AUC values near 0.99. These methods rely on high-quality clinical photos and computing resources, which are challenging to implement in clinics located in rural Bangladesh.

Overall, the existing literature demonstrates that skin disease is prevalent in Bangladesh among children, students, and senior citizens and is heavily influenced by social and environmental variables. However, most AI-based literature on skin diseases either relies on relatively small, survey-based models or clinical skin images and well-recorded datasets, which are hard to gather in rural settings. Even though it is clear that factors such as water source, hygiene practices, household crowding, education, income, and animal contact are responsible for spreading the disease, there is still a lack of a lightweight, simple prediction model that uses the epidemiological data to identify the skin risk for rural communities. This study addressed the gap by creating and evaluating an epidemiological data model for skin disease risk prediction specifically adapted to the rural Bangladeshi context and appropriate for incorporation into community health initiatives.

TABLE I
SUMMARY OF RELATED STUDIES ON SKIN DISEASE RISK PREDICTION

Year	Author	Method Used	Accuracy	Research Gap	Epidemiology data	Dataset
2025	Abbas et al.	Transfer Learning (DL), Explainable AI	CNN - 98%, ResNet-50: 84% DenseNet-121: 89% accuracy	Fully image-based	Not used	Large image datasets
2025	Hoque et al.	Epidemiological survey analysis	Identified major predictors.	Not ML based;	Partially analysis	Field survey data (Bangladesh)
2025	Islam et al.	Cross-section asses	58% fungal infections, >20% scabies cases	No predictive modelling;	Epidemiological analysis only	Clinic-based records
2024	Hasan et al.	Risk factor analysis.	Scabies prevalence 18–34%	Lacks ML prediction;	Yes	Field survey (Madrasahs)
2024	Wan et al.	ML models on (EHR).	AUC ≈ 0.82–0.83 (high performance)	Fully digital EHR system;	No	Large-scale EHR dataset.
2024	Yusra et al.	Hybrid ML	99.26% skin-disease detection	ML-based diagnosis	No	Image datasets
2024	Panwar et al.	ML models	Used simple ML models	Limited dataset	Yes	Survey dataset
2024	Vayadande et al.	ML for risk prediction.	Effective for health surveys	Not dermatology-specific;	Yes	General health datasets
2022	Meena et al.	KNN, SVM, Random Forest.	97% (RF)	No privacy, hygienic also not rural	Partially	Hospital dataset
2022	Chouhan et al.	Economic impact study	Environment risk related	Not ML modelling	Epidemiological	Livestock dataset
2021	Samiul Huq et al.	Community based	Find major skin disease	No predictive	Yes	Clinical survey data

2017	Esteva et al.	CNN, Deep learning	Dermatologist-level accuracy	Not usable in low-resource settings;	No	ISIC image dataset

Many studies have used deep learning or ensemble machine learning techniques to detect skin diseases using infected skin images or well-recorded clinical datasets. These methods frequently rely on high-quality imaging, lab data, or urban lifestyle questionnaires, which are challenging to maintain in rural settings in Bangladesh. In the framework of lightweight, epidemiology-based risk prediction, Table I thus classifies the current literature into broad categories and identifies its primary shortcomings. [24] used ensemble models to obtain 97% accuracy for structured tabular data on the UCI dermatology dataset, however the features are specialized biopsy attributes rather than community epidemiology. Although they don't focus on rural skin conditions, epidemiology-based machine learning research such as [26] for parasitic infections and [19] for EHR-based melanoma risk demonstrate that survey data can support prediction. Although they did not concentrate on dermatology specifically, [13], [14] demonstrated lightweight ML on general health surveys.

TABLE II
LIMITATIONS OF EXISTING AI-BASED APPROACHES

Type	**Model**	**Limitations**	**Related work**
Image-based	CNN / transfer learning on dermoscopic images (ResNet-50, DenseNet-121, sequential CNN, etc.)	Requires dermatoscopes or high-quality clinical images, GPUs, and stable internet; not feasible for most rural Bangladeshi clinics	[20], [21], [22]
Image-based	Advanced deep models and ensembles (Xception, Inception-v3, Inception-ResNet-v2, MobileNet, multi-CNN	Optimised for large, multi-class image datasets; high computational cost; no integration of hygiene, socio-economic, or environmental variables.	[20], [21]
Image-based	Classical ML with texture features (GLCM, color statistics) + DT, SVM, KNN on ISIC / HAM10000	Depends on careful preprocessing (hair removal, segmentation, denoising) and good dermoscopic images; unsuitable where only tabular clinic data exist.	[19]
Image-based (mobile / app)	Mobile / app-based systems combining CNNs with ensemble and data-mining algorithms	Improves accessibility but still relies on smartphone cameras and connectivity; does not exploit routine epidemiological records from rural facilities.	[21], [23]
Tabular clinical (dermatology)	Ensemble data-mining on UCI dermatology dataset.	Uses biopsy and histopathology attributes collected in specialist hospitals;	[24]
lifestyle (skin)	ML on survey-based lifestyle and treatment data (LR, DT, RF, CatBoost, GBC, LightGBM) for chronic skin diseases	Focuses on symptom improvement in urban specialty clinics; does not model infectious vs non-infectious skin-disease risk in rural populations such as Bangladesh.	[25]
Epidemiology ML (non-skin infection)	ML-based risk-factor analysis using epidemiological survey data for intestinal parasitic infections	Shows how ML can use socio-demographic, environmental and haematological features for infection risk, but targets intestinal parasites (not skin diseases) and an Ethiopian context.	[26]
Epidemiology (skin prevalence & QoL)	Community prevalence and DLQI / CDLQI studies of skin disease in rural populations.	Quantify burden and quality-of-life impact but remain descriptive and do not propose ML-based risk-prediction tools for frontline workers.	[25], [28]
System-level (skin ML)	Reviews of ML/DL for skin-lesion recognition (traditional ML + many	Summarise image-centric pipelines and big datasets; provide little guidance on lightweight, interpretable, epidemiology-based models for low-resource settings.	[21], [29]

	CNN/UNet variants; multiple public image datasets)		

Table II highlighted that most current research is either image-based or depends on structured clinical datasets. Although they are descriptive rather than predictive, recent reviews and quality of life studies like [25], [4] further emphasize the social and psychological burden of skin disease. The primary image-based and epidemiology-based research mentioned above are summarized in Table I, and a summary of their technological limitations is given in Table II. However, only a small number of studies use epidemiological survey variables to create lightweight, comprehensible risk-prediction models appropriate for rural Bangladesh.

TABLE III
COMPARATIVE ANALYSIS OF THE PROPOSED EPIDEMIOLOGY-BASED FRAMEWORK WITH PREVIOUS SKIN DISEASE PREDICTION STUDIES

Study	**Data type**	**Methods**	**Best performance**	**Context vs this work**
Abbas et al. (2025)	Dermoscopic images	Sequential CNN, ResNet, DenseNet	98% accuracy, 99% AUC	High-resource, image-based; no rural epidemiology
Verma et al. (2019)	UCI dermatology dataset	Ensemble data-mining models	97% accuracy	Specialist clinic attributes; no hygiene/living-condition variables
Zafar et al. (2022)	Survey data (intestinal parasites)	LR, SVM, RF, XGBoost with SMOTE	AUC > 0.8	Epidemiology-based but not skin diseases; Ethiopian setting
Park et al. (2024)	Lifestyle data (chronic skin)	LR, RF, boosting models	High F1-scores	Urban specialty clinics; focuses on symptom control
This study	Epidemiological survey (skin risk)	KNN, RF, LR, NB	88% train–test, 80.2% 10-fold accuracy	Lightweight, interpretable model for rural Bangladesh

Table III demonstrates that most of the previous research either uses well recorded clinical datasets or dermoscopic pictures, which limits its applicability to rural settings with limited resources

However, the suggested framework, is more appropriate for rural Bangladeshi settings since it only uses lightweight models and epidemiological survey variables.

III. Research Methodology

A. THEORETICAL FRAMEWORK

These days, people in rural Bangladesh are aware of skin disease in rural Bangladesh, and they can understand that it happens for many reasons. As they are linked to social, economic, and environmental problems. According to the World Health Organization (WHO), this is explained through the Social Determinants of Health idea. This says that human health is not only about our body but also about the world around us [11]. Importantly, education, cleanliness, money, housing, and gender can also change a person's health. When the community does not have good conditions or low income, they are more likely to have skin problems [17].

The Social Determinants of Health (SDH) framework, which emphasizes that individual health outcomes are impacted not only by biological factors but also by education, income, family, water and sanitation, and larger environmental conditions, serves as the overall foundation for this study. In rural Bangladesh, where overpopulation, polluted water, and inadequate sanitary facilities significantly impact disease risk, these social and structural factors are particularly significant for skin diseases. Simultaneously, the work adopts a Supervised Machine Learning (SML) approach to risk prediction, wherein a model learns a mapping from epidemiological input features (hygiene, environmental, and demographic variables) to an output label representing the type or degree of skin-disease risk. This combination of SDH and SML offers an empirical basis for identifying community members who are most at risk and could thus benefit from early intervention utilizing epidemiology data.

Computer science domain, especially artificial intelligence, machine learning, is now used to help identify health risks.

There are some normal or lightweight models, like decision KNN, logistic regression, and random forests can find and explain the links between many variables. These models use patient data from surveys or health center records to study people's health [18].

Also, the benefits of using more models help to read data about people, families, and their lives. This makes risk accuracy better and helps to target those who need it most. For example, it can also help to detect which areas of communities have a high risk of getting a certain skin issue. Skin disease prevalence is significantly influenced by environmental and behavioral variables, including household obstruction of personal hygiene, and water quality [19]. As a result, these factors are included in the study's epidemiological dataset.

Thus, the theoretical conceptual paper figures on two interconnected frameworks that connect with public health and artificial intelligence.

1. Social Determinants of Health (SDH): Skin diseases are spreading due to socio-economic, environmental, and household factors.
2. Supervised Machine Learning (SML): Like Logistic Regression, Naïve Bayes, KNN, and Random Forest learn patterns from provided epidemiological datasets.

For this study, frameworks will provide a strong, clear base. Without using clinical images, epidemiological data can be used securely in the model to predict the risk of skin diseases in rural communities.

B. Conceptual Framework

The proposed conceptual framework is designed with a structured approach. By using epidemiological data and lightweight machine learning models to detect the risk of dermatological problems in rural communities in Bangladesh. The framework consists of four key stages: (a) data collection, (b) data preprocessing, (c) modelling and analysis, and (d) evaluation. Each stage is important for achieving practical accuracy in underserved areas with limited resource facilities.

C. *Data Collection*

The initial plan of data collection is to record epidemiological data from Bangladeshi rural clinics and hospitals. The variable will be the patient demographics (age, gender, marital status, education, income), sanitation and hygienic behavior (frequency of bathing, hand washing, soap use), environmental features (source of water, crowded household, pet contact), and related to work (chemicals, sunshine). During the physical clinical visits inside the doctor's consultation room. The expert dermatologist assigns the dependent variable to three classes: 0 - Not a skin disease, 1 - Contagious, 2 - Not Contagious.

To ensure the ethical and responsible handling of patient data, privacy protection was maintained throughout the data collection phase. All records were stored using anonymized identifiers, and directly identifying attributes were kept in a separate, access-restricted file. Clinical data were encrypted at rest and only deidentified or aggregated datasets were used for analysis. These measures were implemented to prevent unauthorized disclosure, maintain confidentiality, and uphold ethical standards for handling health information.

TABLE IV
SAMPLE OF EPIDEMIOLOGICAL DATASET

Id	Age	Gender	Occupation	Water	Birth	Household	Pet Contact	Skin Status
R001	30	Male	Farmer	Pond	3-5 times / week	>6 persons/room	Yes	Contagious
R002	45	Female	Housewife	Tube well	Daily	4–5 persons/room	No	No disease
R003	18	Male	Student	River	1-2 times / week	>6 persons/room	Yes	Contagious
R004	55	Female	Farmer	Pond	Daily	3–4 persons/room	No	Non-
R005	27	Male	Labor	Tube well	2-3 times / week	4–5 persons/room	Yes	Contagious

D. Data Pre-Processing

The collected dataset uses several preprocessing techniques before the ML model is used:

1. Filled the missing Values: To avoid any kind of bias, missing values are either input or deleted.
2. Outlier Detection: To identify any human errors.
3. Encode the categorical Variables: Encode the numerical types to nominal variables like gender, occupation, educational level, and disease types.
4. Classes Balancing: Uses the resampling techniques, like supervised sampling or SMOTE, where appropriate. These techniques are employed to increase the class distribution, for example, the class-2 ratio.
5. Scaling the features: algorithms like KNN, normalizing, or standardization are used.

So, these preprocessing steps enhance the prediction performance and model accuracy.

E. Modelling and Analysis

The models were used with four supervised learning algorithms: K-Nearest Neighbors (KNN), Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR). KNN is a non-parametric classifier, like an instance-based classifier, that assigns a new sample a class label based on the majority class of its k nearest training samples in feature space. In this study, we used Euclidean distance and fixed k = 5 neighbors. RF is a collection of decision trees trained on bootstrap samples. Each tree is constructed using a random subset of characteristics and determined by majority voting among the trees. Moreover, we employed 100 trees with at least two samples per leaf and a maximum depth of, say, 10. We used the Gaussian Naïve Bayes classifier, a probabilistic classifier that applies Bayes' theorem under the presumption of conditional independence amongst predictors. We trained a multinomial logistic regression model with L2 regularization penalty and regularization strength C = 1.0. LR is a linear model that uses the logistic function to predict the likelihood of belonging to each class.

To achieve the best accuracy, every model is trained using '10-fold cross-validation and train–test splitting (60:40, 70:30, 80:20, 90:10). The most important features of the dataset (sun exposure, chemical contact, and bathing frequency) can be identified using feature importance analysis.

According to initial performance, KNN and Random Forest frequently outperform LR and NB, achieving 88% in the 90:10 split and 79–80% in 10-fold cross-validation.

F. Evaluation & performance metrics

There are several metrics used to get the best model performance:

1. The highest precision indicates the minimal false positives. Precision estimates the percentage of positive cases that are correct.
2. Cohen's kappa statistic measures the degree of agreement between the model's prediction accuracy and the actual labels; values nearer 1 denote stronger deal.
3. The confusion matrix was used to generate several common classification measures that were used to evaluate the model's performance.
4. The percentage of real positive cases that the model correctly detects is called recall. A high recall indicates few false negatives. When classes are unbalanced, the F1-score, which is the harmonic mean of precision and recall, provides a single metric that balances both.
5. The model's ability to distinguish between classes across all potential classification limits is summarized by the Area Under the Curve (ROC-AUC); greater AUC values suggest better overall separability between positive and negative classes.

The metrics are computed as follows:

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

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

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

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

$$\text{Cohen's Kappa K} = \frac{p_o - p_e}{1 - p_e}$$

Here p_o is an observed agreement (accuracy) and p_e is expected agreement.

$AUC = \int_0^1 TPR(FPR)\, d(FPR)$ Here, TPR = $\frac{TP}{TP+FN}$, FPR= $\frac{FP}{FP+TN}$

Here, TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively, with the positive class denoting the risk of skin diseases.

Based on the results, the highest Kappa values (0.60–0.62) were obtained by KNN and Random Forest, suggesting moderate to good agreement. However, the Naïve Bayes performed worse (~63%) than Logistic Regression, which generated moderate accuracy (~70%).

This prediction confirms that this framework is workable for skin disease risk in underserved areas. The framework is visualized in Figure 1.

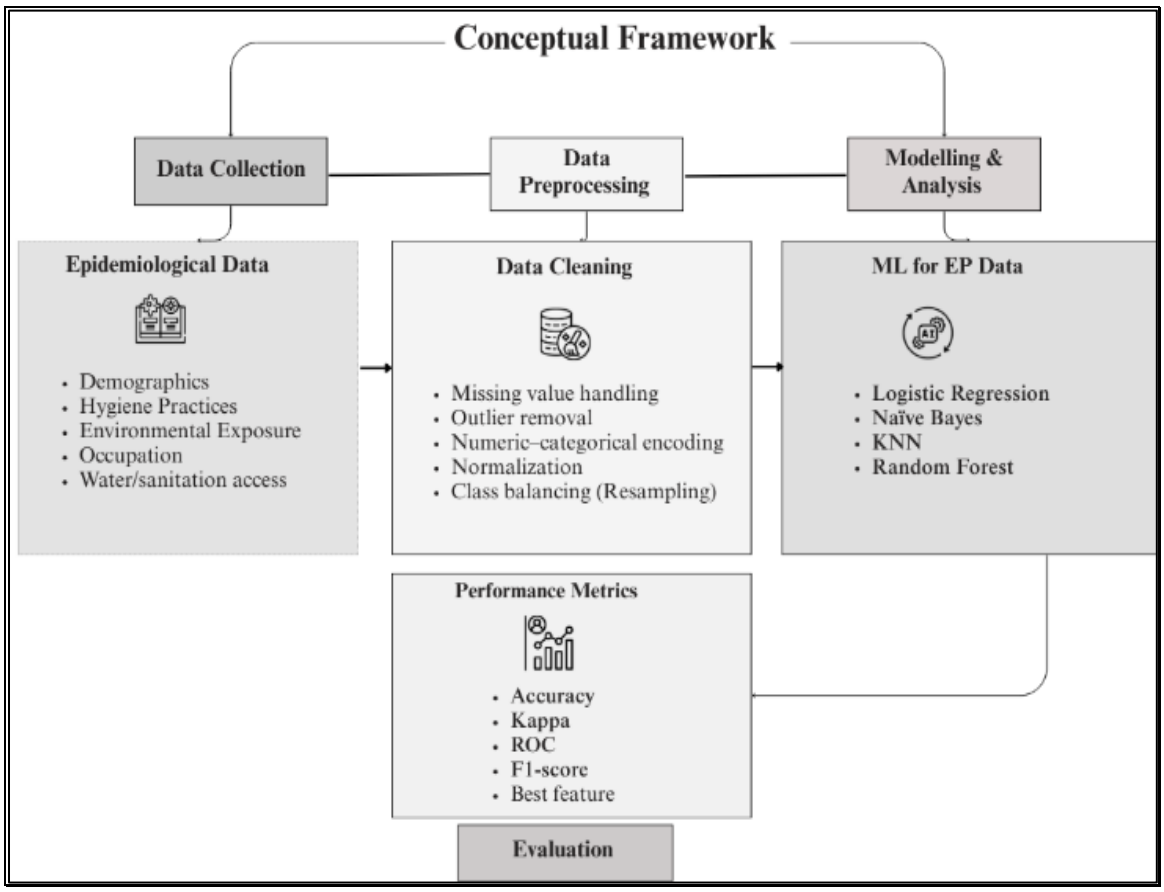


Fig. 1 Research methodology for epidemiology-based AI skin disease risk prediction.

IV. Experimental Analysis

To review a 500-epidemiology data set, the data were pre-processed in many ways, like addressing missing values, dropping the outliers, encoding numerical and categorical factors, leveling feature scales, and using resampling techniques. After preprocessing, several supervised machine learning models like Logistic Regression, Naïve Bayes, K-Nearest Neighbors (KNN), and Random Forest classifiers were assessed for their capacity to predict outcomes from the dataset. This allowed for a methodical comparison of performance across various algorithmic families.

A. Classification performance of the model:

According to 10-fold cross-validation, Random Forest performed the best overall, with 80.2% accuracy and a significant Kappa value of 0.6216, suggesting strong agreement beyond chance. KNN came in second with 79% accuracy, whereas Naïve Bayes and Logistic Regression had lower accuracies of 63% and 70.6%, respectively, demonstrating that instance-based and tree-based approaches outperformed linear and probabilistic models for this dataset. The accuracy output is visualized in Table V.

TABLE V
10-FOLD CROSS-VALIDATION ACCURACY OF ML MODELS ON THE EPIDEMIOLOGICAL DATASET

Algorithm	Accuracy (%)	Kappa	ROC
Logistic Regression	70.6	0.44	0.77
Naïve Bayes	63	0.28	0.75
KNN	79	0.60	0.82
Random Forest	80.2	0.62	0.88

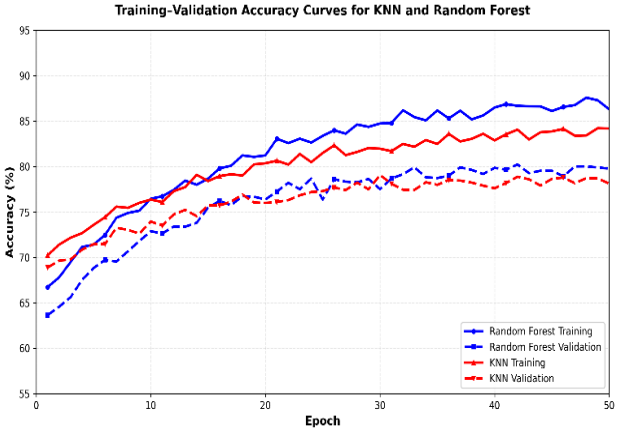


Fig. 2 Training–validation accuracy curve.

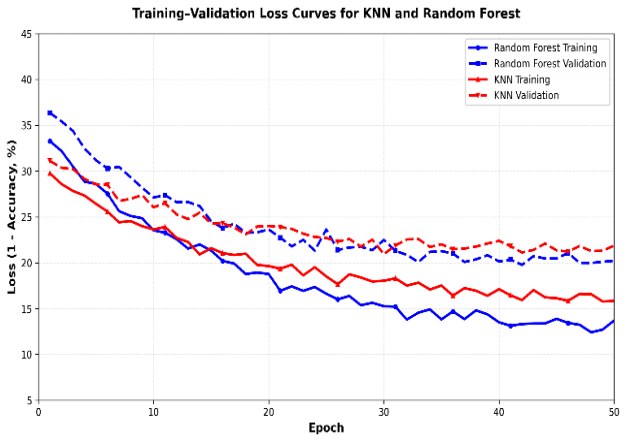


Fig. 3 Training–validation loss curve.

In the Figure 2 shown the training curves with validation accuracies stabilizing between 79 and 80% without significant gaps from, both KNN and Random Forest achieve smooth convergence.

However, the models generalize well and do not show significant overfitting on the epidemiology dataset, as shown by the associated loss curves in Figure 3, which drop slowly and do not diverge.

B. Train and test splitting:

Across various train-test splits, KNN and Random Forest produced the most consistent and often higher accuracies; performance improved as training size grew. KNN's accuracy peaked at 88% at the 90:10 split, while Random Forest's accuracy peaked at 84%. Both algorithms outperformed Naïve Bayes and Logistic Regression, which displayed greater variability and lower scores. The train-split accuracy output is visualized in Figure 4.

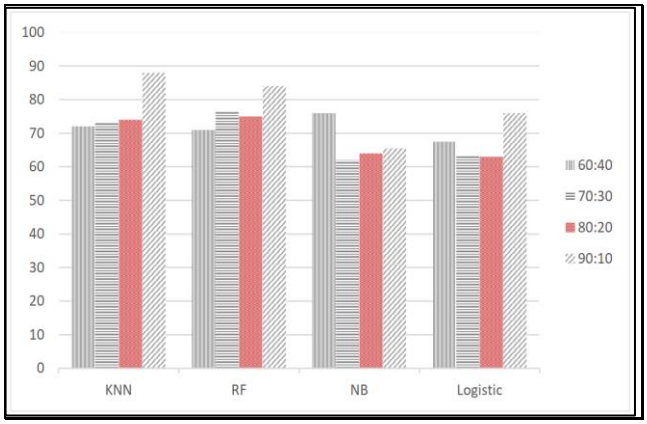


Fig. 4 Train–Test Split Analysis

C. Best Algorithm:

The accuracy scores of four machine learning models were tested to predict skin diseases. The best accuracy was 88% for KNN, 84% for Random Forest, and 76% for both Naïve Bayes and Logistic Regression. This graphic demonstrates that KNN is the dataset's most efficient algorithm, surpassing the competitors in predicting. The best algorithm accuracy is visualized in Figure 5.

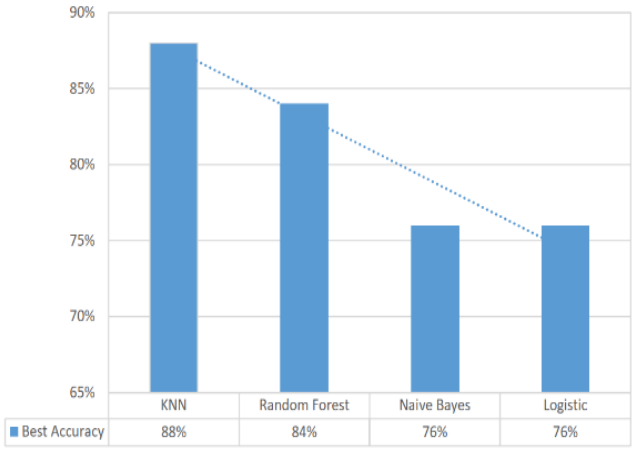


Fig. 5 Best Algorithm Accuracy

D. Best Fit Features:

The categories and importance of various factors influencing skin disease risk. Sun exposure, chemical exposure, and bathing frequency are rated as highly important and span environmental, occupational, and hygiene categories. Shared clothing and family history have medium to high significance within behavioural and genetic domains, while pet contact and household environment play a moderate role as environmental factors. This highlights the multidimensional nature of risk, covering demographic, behavioural, and ecological aspects. The best features shown is in table VI.

TABLE VI
THE BEST EPIDEMIOLOGICAL FEATURES FOR SKIN-DISEASE RISK PREDICTION.

Feature	Category	Importance level
Sun exposure	Environmental	High
Chemical exposure	Occupational	High
Bathing frequency	Hygiene	High
Shared clothing	Behavioural	Medium–high
Household environment	Environmental	Moderate
Pet contact	Environmental	Moderate

V. Discussion

Although this study is conceptual, preliminary experiments were conducted using an epidemiology-based dataset to assess the feasibility of developing lightweight machine learning models for rural Bangladesh. This conceptual framework aims to present a lightweight machine learning method for predicting skin disease risks using

epidemiological factors. Despite the study's conceptual nature, preliminary testing was conducted to assess the viability of different conventional machine learning methods using epidemiological data linked to skin diseases.

These initial experiments provided insights into the potential applicability of conventional machine learning methods in resource-constrained settings.

Across all experimental runs, KNN and Random Forest consistently outperformed Naïve Bayes and Logistic Regression in terms of model accuracy. However, KNN obtained 79% accuracy with a kappa value of 0.605 in 10-fold cross-validation, while Random Forest obtained 80.2% accuracy with a kappa of 0.6216. These kappa values show moderate to large agreement, demonstrating that the algorithms can learn significant epidemiological patterns and outperform chance.

About 70% accuracy was attained by logistic regression, indicating that while linear correlations do exist, they are insufficient on their own to fully capture the intricacy of skin-disease patterns. The independence assumption does not match well with coupled epidemiological variables, as evidenced by the lowest performance of Naïve Bayes (~63%). When multiple train–test splits were tested, the most stable performance was also observed in KNN and Random Forest. Accuracy increased when the training size increased (90:10 split), reaching up to 88% (KNN) and 84% (Random Forest). This trend indicates that these models benefit from larger training samples and can generalize well to unseen data.

However, it is important to recognize a few limitations. First, the current results may not accurately reflect the diversity of real rural communities in Bangladesh because they are based on a single dataset with a small sample size and class imbalance. Second, it is impossible to establish a causal association between risk factors and skin disease because the variables are taken from cross-sectional survey data. Third, the models' performance may alter when used in new contexts, such as various districts or primary-care institutions with distinct population characteristics, as they have not yet been prospectively verified in standard clinical workflows. Therefore, using locally gathered epidemiology data, additional validation is needed from various rural areas.

Lastly, although basic privacy and encryption techniques were implemented in the data collection and preprocessing phase, it will be crucial to incorporate improved privacy-preserving strategies to guarantee the safe and moral handling of patient data during large-scale deployment. Overall, these limitations suggest that the current findings should be considered preliminary; yet they offer a valuable basis for creating interpretable, cost-effective, and privacy-conscious prediction frameworks for the risk of skin diseases in underserved areas.

VI. Conclusion

This conceptual paper proposes a lightweight, epidemiology-driven model for predicting skin disease risk in underserved rural areas of Bangladesh. Early experiments conducted on an existing dataset indicate that conventional machine learning models, particularly KNN and Random Forest, achieve reliable predictive performance with accuracy reaching approximately 88% under train–test evaluation and 80% under 10-fold cross-validation. These results demonstrate that high-accuracy prediction can be made without the need for image-based diagnostic tools, utilizing only structured patient and environmental data. The outcomes validate the feasibility of the suggested framework for practical implementation in rural clinics. Additionally, privacy protection is another important factor to secure patient information In addition, privacy protection was integrated as a core design element to ensure secure and ethical handling of patient information. Collectively, these contributions provide a foundation for an interpretable and context-appropriate risk-prediction model that may strengthen early disease detection capabilities in low-resource communities Future work will involve acquiring epidemiological data directly from rural Bangladeshi populations, integrating privacy-preserving techniques into the risk-detection process, and validating the proposed framework within real-world healthcare workflows.

Acknowledgment

The authors hereby thank the IJPCC reviewers for their assistance in reviewing the manuscript and determining that it is suitable for publication.

Conflict Of Interest

The authors declare that there is no conflict of interest

Authors contribution statement

All authors contributed to the conception and design of the study, the development of the conceptual framework, and the interpretation of the outcomes. The first author led the literature review, data preprocessing, the model development, and drafting of the manuscript, while the co-authors provided supervision, domain expertise, model evaluations and critical revisions. All authors read and approved the final version of the manuscript.

Data availability statement

In this study, the dataset is not publicly available due to patient confidentiality and health centers restrictions. The dataset and analysis scripts can be obtained from the corresponding author upon reasonable request and subject to institutional and ethical approval.

ETHICS STATEMENT

The epidemiological data set used in this study was separated and anonymized before analysis. The original data collection followed institutional and authorities' ethical guidelines for research involving patient participants, and informed consent was obtained in the primary studies where required. Clinical records were stored with encrypted identifiers to maintain confidentiality.

REFERENCES

[1] T. Hoque, Md. R. Islam, and A. Akter, "Common skin diseases in children: A cross-sectional study from a semi-urban community in Bangladesh," *Int. J. Pediatr. Neonatol.*, vol. 7, no. 1, pp. 66–70, Jan. 2025, doi: 10.33545/26648350.2025.v7.i1b.122.

[2] K. Islam, M. I. Islam, T. Jahan, M. A. Yusuf, S. H. Chowdhury, and F. T. Zohora, "Pattern of Skin Diseases among Rural Adult Patients Attending the Dermatology OPD at A Tertiary Care Hospital Bangladesh," *Bangladesh J. Med. Microbiol.*, vol. 19, no. 1, pp. 54–59, Apr. 2025, doi: 10.3329/bjmm.v19i1.80603.

[3] N. Islam and M. I. H. Shakil, "Epidemiology of scabies among resident school and madrasah children: An observational study," *Int. J. Dermatol. Sci.*, vol. 7, no. 1, pp. 06–09, Jan. 2025, doi: 10.33545/26649772.2025.v7.i1a.44.

[4] Mohammad Samiul Huq, Abu Hena Chowdhury, Towhida Noor, and Saleheen Huq, "Psycho-social determinants and magnitude of public health problems of psoriasis in Bangladesh," *World J. Adv. Res. Rev.*, vol. 10, no. 2, pp. 108–118, May 2021, doi: 10.30574/wjarr.2021.10.2.0207.

[5] C. S. Chouhan *et al.*, "Epidemiology and economic impact of lumpy skin disease of cattle in Mymensingh and Gaibandha districts of Bangladesh," *Transbound. Emerg. Dis.*, vol. 69, no. 6, pp. 3405–3418, Nov. 2022, doi: 10.1111/tbed.14697.

[6] M. J. Hasan, M. A. Rafi, T. Choudhury, and M. G. Hossain, "Prevalence and risk factors of scabies among children living in Madrasahs (Islamic religious boarding schools) of Bangladesh: a cross-sectional study," *BMJ Paediatr. Open*, vol. 8, no. 1, p. e002421, June 2024, doi: 10.1136/bmjpo-2023-002421.

[7] N. Ahmed, M. Islam, and S. Farjana, "Pattern of Skin Diseases: Experience from a Rural Community of Bangladesh," *Bangladesh Med. J.*, vol. 41, no. 1, pp. 50–52, May 2014, doi: 10.3329/bmj.v41i1.18784.

[8] S. Abbas, F. Ahmed, W. A. Khan, M. Ahmad, M. A. Khan, and T. M. Ghazal, "Intelligent skin disease prediction system using transfer learning and explainable artificial intelligence," *Sci. Rep.*, vol. 15, no. 1, p. 1746, Jan. 2025, doi: 10.1038/s41598-024-83966-4.

[9] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[10] R. Parvin *et al.*, "Clinical Epidemiology, Pathology, and Molecular Investigation of Lumpy Skin Disease Outbreaks in Bangladesh during 2020–2021 Indicate the Re-Emergence of an Old African Strain," *Viruses*, vol. 14, no. 11, p. 2529, Nov. 2022, doi: 10.3390/v14112529.

[11] "WHO Initiative on artificial intelligence for skin conditions." Accessed: Nov. 28, 2025. [Online]. Available: https://www.who.int/initiatives/who-initiative-on-artificial-intelligence-for-skin-conditions

[12] A. Krumeich and A. Meershoek, "Health in global context; beyond the social determinants of health?," *Glob. Health Action*, vol. 7, no. 1, p. 23506, Dec. 2014, doi: 10.3402/gha.v7.23506.

[13] P. Panwar, S. Bangwal, U. Pasbola, A. Kumar, A. Sar, and T. Choudhury, "Diagnosis and Prediction of Skin Diseases Using Deep Learning for Rural Healthcare," in *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS)*, Dehradun, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ISTEMS60181.2024.10560209.

[14] K. Vayadande, A. A. Bhosle, R. G. Pawar, D. J. Joshi, P. A. Bailke, and O. Lohade, "Innovative approaches for skin disease identification in machine learning: A comprehensive study," *Oral Oncol. Rep.*, vol. 10, p. 100365, June 2024, doi: 10.1016/j.oor.2024.100365.

[15] N. Yusra, K. K, S. A, and D. V, "Interpretable machine learning for dermatological disease detection: Bridging the gap between accuracy and explainability," *Comput. Biol. Med.*, vol. 179, Sept. 2024, doi: 10.1016/j.compbiomed.2024.108919.

[16] E. B. Weiner, I. Dankwa-Mullan, W. A. Nelson, and S. Hassanpour, "Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice," *PLOS Digit. Health*, vol. 4, no. 4, p. e0000810, Apr. 2025, doi: 10.1371/journal.pdig.0000810.

[17] M. Marmot, R. Bell, and P. Goldblatt, "Action on the social determinants of health," *Rev. DÉpidémiologie Santé Publique*, vol. 61, pp. S127–S132, Aug. 2013, doi: 10.1016/j.respe.2013.05.014.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[19] G. Wan, "Title Individualized melanoma risk prediction using machine learning with electronic health records," 2024, doi: https://doi.org/10.1101/2024.07.26.24311080.

[20] M. Ahammed, Md. A. Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthc. Anal.*, vol. 2, p. 100122, Nov. 2022, doi: 10.1016/j.health.2022.100122.

[21] N. Fatima, S. A. M. Rizvi, and M. S. B. A. Rizvi, "Dermatological disease prediction and diagnosis system using deep learning," *Ir. J. Med. Sci. 1971 -*, vol. 193, no. 3, pp. 1295–1303, June 2024, doi: 10.1007/s11845-023-03578-1.

[22] Muddasar Abbas, Muhammad Imran, Abdul Majid, and Nadeem Ahmad, "Skin Diseases Diagnosis System Based on Machine Learning," *J. Comput. Biomed. Inform.*, vol. 4, no. 01, Dec. 2022, doi: 10.56979/401/2022/53.

[23] B. Suman, N. Harika, B. Sruthi, and B. Bhagyasree, "Prediction of Skin Diseases Using Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 8, pp. 791–796, Aug. 2022, doi: 10.22214/ijraset.2022.46138.

[24] A. K. Verma, S. Pal, and S. Kumar, "Comparison of skin disease prediction by feature selection using ensemble data mining techniques," *Inform. Med. Unlocked*, vol. 16, p. 100202, 2019, doi: 10.1016/j.imu.2019.100202.

[25] C.-Y. Park, J. Joo, O.-H. You, S. Yi, C.-Y. Kim, and A.-R. Jo, "Development of a predictive model for managing lifestyle behaviors among patients with chronic skin diseases: Using machine learning techniques," *Inform. Med. Unlocked*, vol. 48, p. 101528, 2024, doi: 10.1016/j.imu.2024.101528.

[26] A. Zafar *et al.*, "Machine learning-based risk factor analysis and prevalence prediction of intestinal parasitic infections using epidemiological survey data," *PLoS Negl. Trop. Dis.*, vol. 16, no. 6, p. e0010517, June 2022, doi: 10.1371/journal.pntd.0010517.

[27] C. I. Wootton *et al.*, "Assessing skin disease and associated health-related quality of life in a rural Lao community," *BMC Dermatol.*, vol. 18, no. 1, p. 11, Dec. 2018, doi: 10.1186/s12895-018-0079-8.

[28] R. R. Yotsu *et al.*, "Skin disease prevalence study in schoolchildren in rural Côte d'Ivoire: Implications for integration of neglected skin diseases (skin NTDs)," *PLoS Negl. Trop. Dis.*, vol. 12, no. 5, p. e0006489, May 2018, doi: 10.1371/journal.pntd.0006489.

[29] J. Sun *et al.*, "Machine Learning Methods in Skin Disease Recognition: A Systematic Review," *Processes*, vol. 11, no. 4, p. 1003, Mar. 2023, doi: 10.3390/pr11041003.

[30] M. K, N. N. K. Veni, B. S. Deepapriya, P. A. H. Vardhini, B. Kalyani, and S. L, "A Novel Method for Prediction of Skin Diseases Using Supervised Classification Techniques," Apr. 08, 2022, *In Review*. doi: 10.21203/rs.3.rs-1509955/v1.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.634

Applying the Software Development Life Cycle to Design WeResearch: A Unified Research Environment

[1]Mohamed Ali Mahmod, [2]Qais Ali Mahmoud Batiha, [3]Mohammad Y. Mhawish, [4]Zaid Haron Musa Jawasreh, [5]Mohamed Ibrahim Mugableh, [6]Israa Ali Mahmud

[1]Dept of Accounting Information Systems, Irbid National University, Irbid, Jordan.
[2]Dept of Software Engineering, Irbid National University, Irbid, Jordan.
[3,4]Dept of Data Science and Artificial Intelligence, Irbid National University, Irbid, Jordan.
[5]Dept of Finance and Banking Science, Irbid National University, Irbid, Jordan.
[6]Dept of Biotechnology Engineering, International Islamic University Malaysia, Selangor, Malaysia

*Corresponding author: m.mashadani@inu.edu.jo
(Received: 22nd November 2025; Accepted: 14th January, 2026; Published on-line: 30th January, 2026)

Abstract— The growing demand for effective academic collaboration tools highlights the need for unified digital environments that support communication, resource sharing, and knowledge management. This study applies the Software Development Life Cycle (SDLC) framework to the design and prototyping of WeResearch, a unified research environment tailored to researchers' needs. A qualitative methodology was employed, combining insights from prior studies with semi-structured interviews conducted with ten researchers from diverse disciplines. The interviews identified critical requirements for research collaboration, including seamless communication, collaborative task management, and intuitive navigation. Based on these requirements, a prototype of WeResearch was designed to visualize platform functionalities and user experience. The SDLC phases of planning, requirements analysis, design, and prototyping were applied to ensure a structured development process. The prototype was then tested with the same group of researchers, whose feedback provided valuable insights into usability and relevance. Findings suggest that the proposed design aligns strongly with user needs, offering solutions to gaps present in existing research collaboration platforms. This study contributes by integrating qualitative needs assessment with the SDLC framework in the academic context, addressing researchers' needs, software requirements, and full prototype design to support a unified research collaboration platform across world universities.

Keywords— Software Development Life Cycle (SDLC), Unified Research Environment, User-Centered Design, Prototype Design

I. Introduction

Effective academic collaboration is essential for advancing research across disciplines and institutions. However, the proliferation of diverse tools and platforms has led to fragmented digital environments, hindering seamless communication, resource sharing, and knowledge management among researchers. This fragmentation often results in inefficiencies, data silos, and challenges in coordinating tasks and managing research outputs [1], [2].

To address these challenges, the Software Development Life Cycle (SDLC) offers a structured framework for designing, developing, and evaluating digital solutions. The SDLC encompasses several phases, including planning, requirements analysis, design, development, testing, deployment, and maintenance [3], [4]. By applying the SDLC methodology, developers can ensure that software solutions are systematically planned, requirements are clearly defined, designs are rigorously developed, and prototypes are iteratively tested to align with end-user needs [5].

This study focuses on the design and prototyping of WeResearch, a unified research environment tailored to the needs of researchers across disciplines. By employing a qualitative methodology that combines insights from prior studies and semi-structured interviews with ten researchers, this study identifies critical requirements for effective research collaboration, including seamless communication, collaborative task management, and intuitive navigation [6], [7].

The resulting prototype demonstrates how SDLC phases—planning, requirements analysis, design, and prototyping—can be applied to create a unified platform that addresses gaps in existing research collaboration tools. This research contributes to the academic field by offering a structured approach to designing research collaboration software, integrating user needs with systematic software engineering principles, and providing a practical prototype that can guide future implementations in global academic contexts [8].

Unlike existing collaboration tools, WeResearch integrates academic task management, reference handling, and institutional coordination in a unified environment. This study focuses specifically on the early phases of the Software Development Life Cycle, namely planning, requirements analysis, design, and prototyping.

https://doi.org/10.31436/ijpcc.v12i1.634

Implementation, deployment, and maintenance phases are beyond the scope of the current work and are planned as future research directions.

This study is guided by the following research questions:

RQ1: What are the key requirements needed for a unified digital research environment from researchers' perspectives?

RQ2: How can early phases of the Software Development Life Cycle (SDLC) be applied to design and prototype a unified research collaboration platform?

RQ3: To what extent does the proposed prototype align with researchers' perceived needs and usability expectations?

II. LITERATURE REVIEW

Effective research methodologies are essential for maintaining the integrity and quality of academic research. However, many researchers, especially graduate students, face challenges due to insufficient guidance on selecting and applying appropriate methods. Alebaikan and Alsemiri [9] examined the attitudes of graduate students at King Saud University towards digital academic integrity and found that a lack of awareness regarding proper research practices often led to issues such as digital plagiarism. Their study highlighted the role of limited faculty guidance, inadequate academic writing skills, and insufficient training in managing digital data as key contributing factors to these challenges. Similarly, Flaxman [10] emphasized the importance of clear ethical guidelines and transparent research processes to prevent misconduct in research and publishing. By fostering a culture of accountability, research institutions can significantly reduce the risk of unethical practices and enhance the credibility of their scholarly output.

A strong understanding of research methodology is crucial for conducting rigorous and reproducible studies. Garg [11] outlined the key components of research methodology, including the formulation of research questions, the selection of appropriate study designs, and the implementation of robust data collection and analysis techniques. This structured approach not only ensures the reliability and validity of research findings but also equips researchers with the skills necessary to navigate complex academic challenges. Sreekumar [12] further highlighted that integrating methodological guidance into academic programs helps students and early-career researchers develop a systematic approach to their work, improving both research quality and professional development.\

Recent studies have further emphasized the necessity of comprehensive methodological training. Altowairiki [13] investigated the application of action research to enhance the development of research methodology knowledge among graduate students. The study found that students were dissatisfied with current courses, leading to knowledge gaps and limited application of research approaches. Through iterative cycles of action research, a book club intervention was implemented, highlighting the potential of action research as a valuable framework for developing students' understanding of research methodologies.

Additionally, Schneider [14] discussed the design of international research experiences for students, emphasizing the importance of providing clear guidance on research methodologies. The study underscored that well-structured research experiences, coupled with appropriate methodological training, can significantly enhance students' research skills and contribute to their academic success.

Overall, the literature underscores the critical need for clear and comprehensive guidance on research methodologies. Accordingly, this study addresses gaps in the research environment by focusing on the conceptual design and prototyping of the Wereesearch platform. The primary objective is to validate the proposed design in terms of functionality, usability, and alignment with identified research needs, rather than to conduct systematic comparative or performance-based benchmarking, which is reserved for future work.

III. METHODOLOGY

This study adopted a qualitative research approach supported by the Software Development Life Cycle (SDLC) framework to guide the structured development of the WeResearch platform. The methodology aimed to identify researchers' needs, translate them into functional requirements, and design a prototype reflecting these requirements. The section outlines the participants, data collection procedures, and data analysis methods used to achieve these objectives.

A. Participants

Ten researchers from diverse academic disciplines participated in this study. Participants were selected using purposive sampling to ensure inclusion of individuals actively engaged in research and publication. Their experience ranged from early-career researchers to senior faculty, providing a comprehensive perspective on research needs and collaboration practices.

B. B. Data Collection

Semi-structured interviews were conducted with each participant, lasting approximately 30–45 minutes. Interviews were held either in person or via phone audio calls. The interview questions focused on researchers' experiences with collaboration, data and document management, access to resources, research planning, and desired features in a digital research environment.

https://doi.org/10.31436/ijpcc.v12i1.634

C. Data Analysis

Transcripts of the interviews were analyzed using thematic analysis. Initial codes were derived from recurring statements and patterns in the data, which were then clustered into broader themes representing researchers' shared experiences and needs. The analysis was conducted manually, as this approach allowed for closer engagement with the data and was appropriate given the manageable sample size. Although software-assisted analysis (e.g., NVivo) can support data management, manual coding was preferred to preserve the depth and contextual richness of participants' responses. Peer debriefing and member checking were employed to enhance the credibility and trustworthiness of the findings.

IV. *RESULTS*

The thematic analysis of the interview data revealed twelve major themes, representing the key features researchers require in a digital research environment:

- Unified Research Environment – Integration of all research tools and resources into a single platform to reduce fragmentation.
- Organized Team Roles – Scheduling and allocation of responsibilities within research teams for coordinated workflows.
- Workbox for Files – Management of personal and shared documents, datasets, and draft manuscripts.
- Visual Access to Models, Theories, and Hypotheses – Tools to support conceptualization and planning.
- Private Notes and Shared Drafts – Facilitating collaboration while preserving individual contributions.
- Wiki of Research Majors – A centralized knowledge base providing quick references across disciplines.
- Direct Access to Online Resources and Libraries – Supporting literature review and data retrieval.
- Direct Contact with Funding Agencies – Streamlining grant and funding application processes.
- Access to Journal and Conference Templates – Supporting proper formatting and submission requirements.
- Communication Tools – Integrated chat and live video for synchronous collaboration.
- Workbox of Previous Studies – Easy retrieval and organization of prior research for reference and citation.
- Integration with Research Platforms – Access to ResearchGate, Google Scholar, and similar platforms to track publications and collaborations.

To map themes to SDLC stages, Table 1 illustrates that as the following:

TABLE 1
MAPPING OF THEMES TO SDLC STAGES (ORGANIZED BY SDLC SEQUENCE)

Theme	Mapped
Unified Research Environment – Integration of all	Planning,
Direct Contact with Funding Agencies – Streamlined	Planning,
Cross-Disciplinary Accessibility – Requirement for	Planning,
Organized Team Roles – Scheduling and role	Analysis,
Visual Access to Models, Theories, and Hypotheses	Analysis,
Workbox for Files – Management of personal and	Design,
Private Notes and Shared Drafts – Collaboration	Design,
Wiki of Research Majors – Centralized disciplinary	Design,
Direct Access to Online Resources and Libraries –	Design,
Access to Journal and Conference Templates –	Design,
Integration with Research Platforms – Linking with	Design,
Resource Sharing and Version Control – Issues with	Design,
Communication Tools – Integrated chat and video	Design,
User-Friendly Interface – Need for simplicity and	Design,
Workbox of Previous Studies – Organized access to	Design,

These findings provide a structured framework of features that informed the design of the WeResearch platform, ensuring it is user-centered, comprehensive, and aligned with the actual needs of researchers.

A. Planning

The planning stage was informed by a comprehensive literature review and a series of semi-structured interviews with researchers. These inputs provided a foundation for identifying key gaps in existing digital research environments and aligning the platform's objectives with actual user needs. The insights gained were translated into a clear project scope, with the goal of designing a unified platform that facilitates research collaboration, data management, and access to resources.

B. Requirements Analysis

The requirements analysis stage focused on systematically mapping the identified needs of researchers to platform features. From the interviews, twelve thematic categories of needs were identified, including collaboration support, document management, resource accessibility, planning tools, and integration of research workflows. Each theme was translated into functional requirements, ensuring that the proposed features directly addressed researchers' priorities and pain points. The functional Requirements include the following points:

User & Role Management: Functional Requirements

FR1.1: University Registration

- University Admin registers the university on the platform.
- Verification ensures legitimacy using either:
  - Institutional email domain, or
  - Official approval letter from the university.

https://doi.org/10.31436/ijpcc.v12i1.634

- Once verified, the university becomes the primary account holder on the platform.

FR1.2: Platform Admin Verification

- Platform Admin reviews and approves:
  - University registration.
  - University Admin accounts representing the institution.
- Ensures all registrations comply with platform standards.

FR1.3: University Admin Role Management, Role Assignment, and Transfer

- University Admin Manages the university profile (official information, description, logo, policies).
- University Admin can:
  - Assign Coordinators to manage academic staff and research activities.
  - Transfer the University Admin role to another verified user if needed (e.g., resignation or reassignment).
  - Transfer Coordinator roles to other verified users when necessary (e.g., workload redistribution or replacement).

FR1.4: Coordinator Assignment and Roles

- University Admin assigns Coordinators to manage academic staff and research activities.
- Coordinators responsibilities:
  - Add Academic staff.
  - Oversee research submissions and collaborations.
  - Act as academic staff: submit research, join projects, and invite external collaborators.
- Multiple coordinators can be assigned for redundancy and workload distribution.
- Coordinate research with academic staff and collaboration with other institutions.

FR1.5: Academic Staff Onboarding

- University Admins onboard academic staff via:
  - Email invitation (single or multiple).
  - Bulk upload using CSV or Excel template.
- Academic staff accounts are automatically linked to the university.

FR1.6: Research Collaboration

- Academic staff can:
  - Submit research projects.
  - Join or collaborate on existing projects.
  - Invite external collaborators with controlled permissions (view/edit/submit).
- All research activities are tracked under the university account.

FR1.7: Audit and Administration

- Platform maintains logs of all actions, including:
  - Role transfers.
  - Staff onboarding.
  - Research submissions.
- Platform Admin can intervene for:
  - University Admin reassignment.
  - Handling inactive accounts or exceptional cases.

C. System Modeling

To translate the identified requirements into a structured representation of the system, Unified Modeling Language (UML) diagrams were developed. UML provides a standardized way to model the functional and structural aspects of the platform, serving as a bridge between requirements analysis and detailed design. Three primary diagrams were produced: The Use Case Diagram, the Activity Diagram, and the Class Diagram. Together, these models describe the intended functionality, workflows, and structural organization of the WeResearch platform.

1) Use Case Diagram

The Use Case Diagram illustrates the interaction between external actors and the system. Key actors include the University Administration, WeResearch Administration, University Coordinators, Academic staff, and Research Collaborators. The diagram illustrates the major system functionalities, including managing platform settings, registration, adding academic staff and external researchers (collaborators). doing research and publishing. This high-level view highlights the scope of the system and clarifies the boundaries between users and platform services. Figure 1 illustrates the key actors of the WeResearch platform using a UML case diagram.

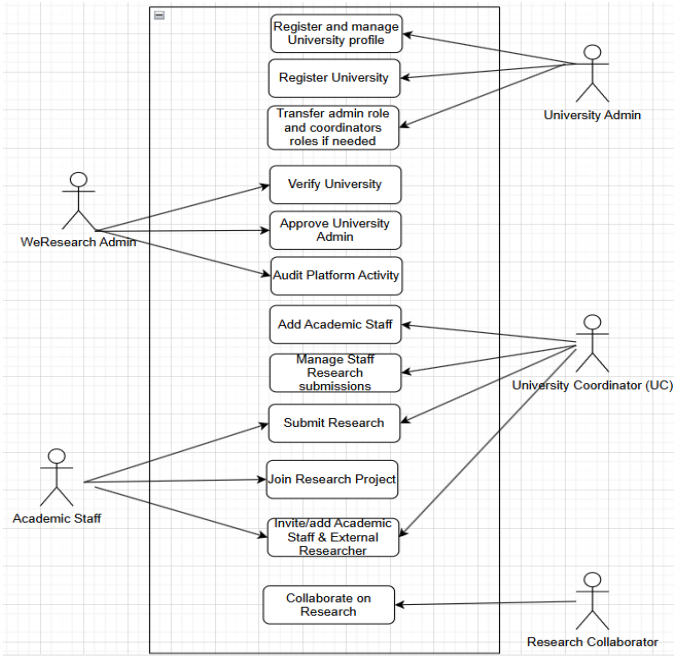


Fig. 1 UML case diagram of WeResearch

https://doi.org/10.31436/ijpcc.v12i1.634

2) Activity Diagram

As part of the system design process, a UML Activity Diagram was developed to model the dynamic behavior of the proposed platform. The activity diagram provides a visual representation of the workflow of actions and decisions that occur when users interact with the system. The activity UML diagram can be described as follows:

2.1 Start (Initial Node)

2.2 University Admin Login/Register

- University Admin creates an account or logs in.

2.3 Platform Verification

- Decision: Platform Admin verifies the university and University Admin account.

2.4 Assign Coordinators

- Action: University Admin assigns one or more Coordinators.

2.5 Coordinator Login/Register

- Action: Coordinator logs in or registers if newly assigned.

2.6 Add Academic Staff

- Coordinators or University Admin can onboard academic staff via:
  - Option A: Bulk upload (CSV/Excel)
  - Option B: Manual entry

2.7 Academic Staff Self-Registration

- Action: Staff complete their profile after being invited.

2.7 Start Research Activities

- Academic staff and coordinators can:
  - Submit research projects
  - Collaborate with peers.
  - Assign tasks/roles.
  - Share drafts and files.
  - Prepare publications.

2.8 External Collaboration

- Action: Invite external researchers to collaborate on projects
- Action: Organize/Invite to conference.

2.9 End (Final Node)

The activity diagram represents the dynamic workflow of the platform, focusing on how tasks are executed in sequence. This diagram emphasizes the logical progression of activities, the flow of control between users and the system, and the decisions that guide different operational paths. Figure 2 demonstrates the UML activity diagram of WeResearch platform.

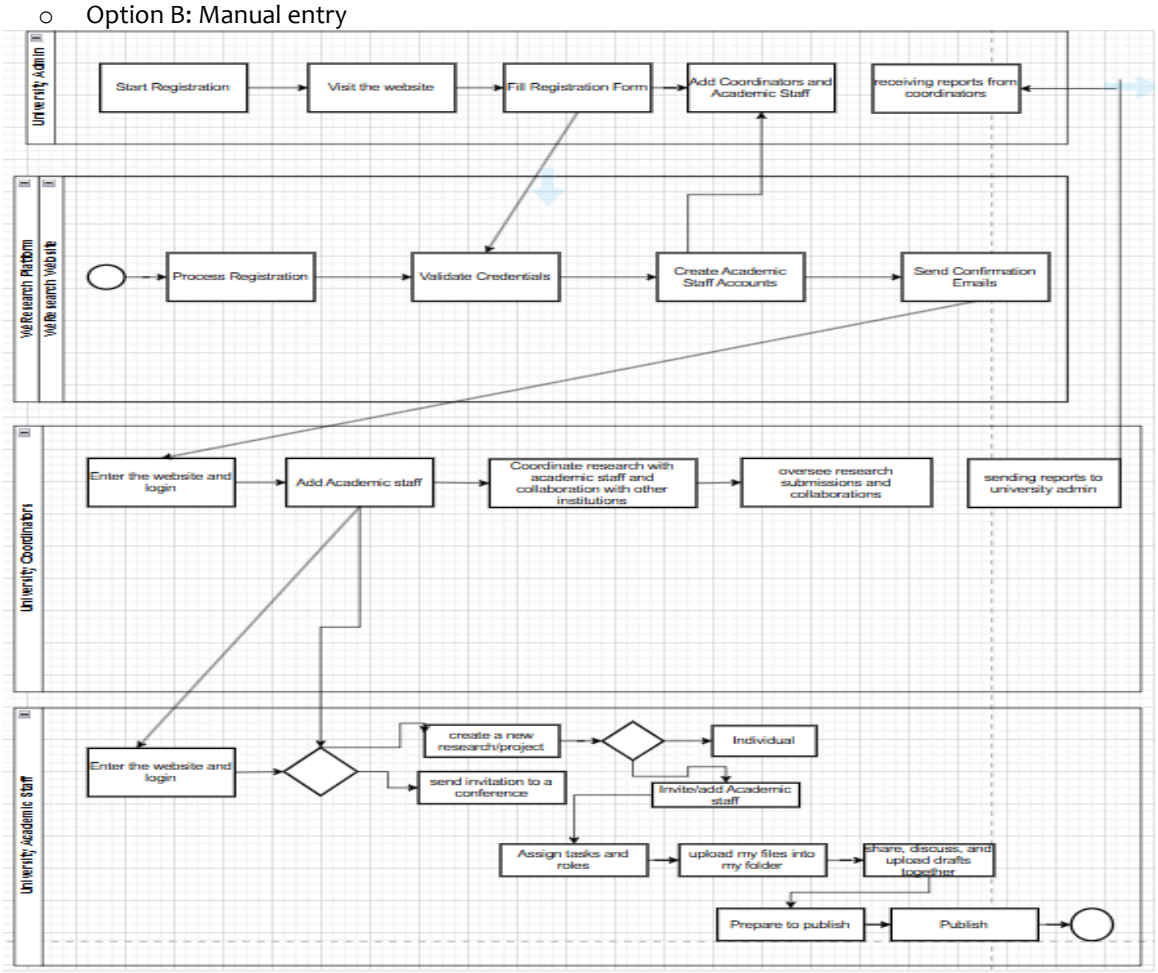


Fig. 2 UML Activity Diagram of WeResearch

https://doi.org/10.31436/ijpcc.v12i1.634

3) Class Diagram

The Class Diagram models the structural aspects of the system by showing its main entities, their attributes, and the relationships between them. Core classes include the following entities:

3.1 WeResearch Admin:

- Attributes: adminID, name, email, password
- Operations: verifyUniversity(), ,approveUniversityAdmin(), auditSystem(), monitorActivity()

3.2 University

- Attributes: universityID, name, address, status
- Operations: verifyRegistration(), approveCoordinator()
- Relationships: 1 University → 1 UniversityAdmin
  1 University → many Coordinators
  1 University → many AcademicStaff

3.3 University Admin

- adminID, universityID(), name , email, password
- Operations: registerUniversity(), manageUniversityProfile(), assignCoordinator(), transferAdminRole(), transferCoordinatorRole()
- Relationships: 1 University → 1 UniversityAdmin
  1 University → many Coordinators
  1 University → many AcademicStaff

3.4 Coordinator

- Attributes: coordinatorID, name, email, password, universityID
- Operations: login(), addAcademicStaff(), uploadCSV(), manageResearchSubmissions(), submitResearch(), joinProject(), inviteCollaborator()
- Relationship: 1 Coordinator manages many AcademicStaff

3.5 Academic Staff

- Attributes: staffID, name, department, email, password, universityID.
- Operations: login(),completeprofile(), startResearch(), collaborate(), shareFile(), shareDraft(), assignTask(), inviteCollaborators()
- Relationship1: AcademicStaff belong to University
- Relationship2: AcademicStaff ↔ AcademicStaff
- Relationship3: AcademicStaff → many ResearchProjects

3.6 Research

- Attributes: ResearchID, title, status, startDate, endDate
- Operations: assignRole(), addDraft(), shareFiles(), preparePublication()
- Relationships: ResearchProject has many AcademicStaff (aggregation), 1 ResearchProject → many Publications, 1 ResearchProject → many ExternalCollaborators

3.7 Collaborator (External Researcher)

- Attributes: CoolaboratorID, name, email
- Operations: collaborate(), receiveInvitation()
- Relationship: ExternalCollaborator ↔ ResearchProject

3.8 Publication

- Attributes: publicationID, title, type, submissionDate, status
- Operations: submit(), review(), publish()
- Relationships: 1 Publication → belongsTo → ResearchProject

Figure 3 illustrates the UML class diagram of WeResearch platform.

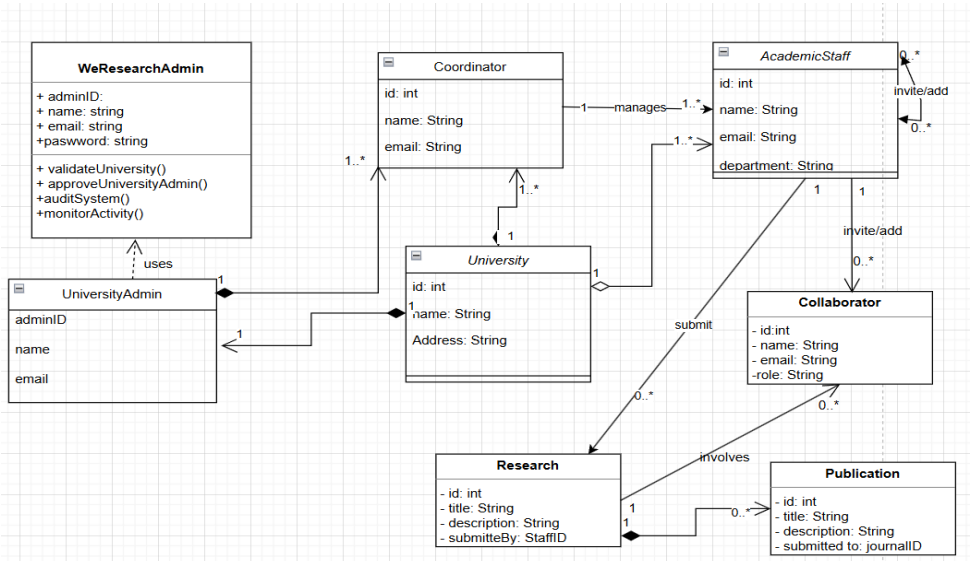


Fig. 3 UML Class Diagram of WeResearch

https://doi.org/10.31436/ijpcc.v12i1.634

By combining these UML models, the system structure and functionality are clearly defined, ensuring that the subsequent design stage is grounded in a well-documented representation of user requirements and platform interactions.

C. Design

In the design phase, the system architecture was outlined to define the structural components of the platform and their interactions. Interface mockups were developed to provide a visual representation of the user experience, with an emphasis on simplicity, intuitiveness, and efficiency. Navigation flows were mapped to ensure seamless movement between modules, and design tools were employed to create wireframes that guided the subsequent prototyping process.

D. Prototyping

The prototyping phase resulted in the development of a working prototype that visualized the platform's core functionalities. This prototype served as both a demonstration of the proposed environment and as a basis for iterative testing and refinement with users. The prototype highlighted key features such as collaborative workspaces, integrated document and data management, research planning modules, and resource access. Figure 4 shows the screen of Researchers' tasks and roles management, and an organized workbox of their files. While Figure 5 shows the WeResearch reference manager.

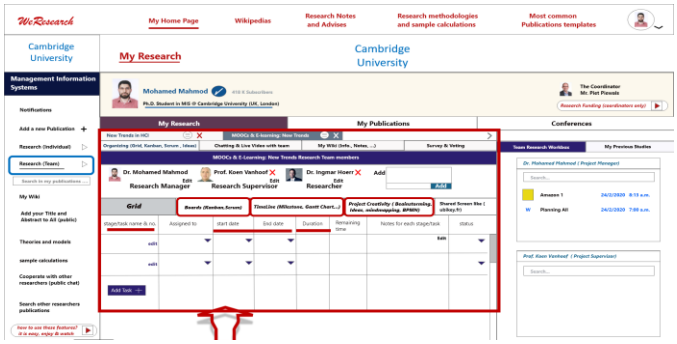


Fig. 4 Screenshots of Researchers' tasks and roles management, and an organized workbox of their files

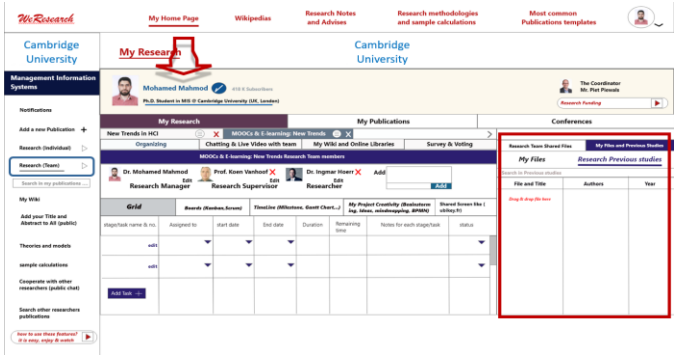


Fig. 5 WeResearch reference manager

The design phase meets the needs of researchers by offering their private and shared notes and files. Figure 6 illustrates the researcher's own notes and draft. Figure 7 shows the researchers' shared notes.

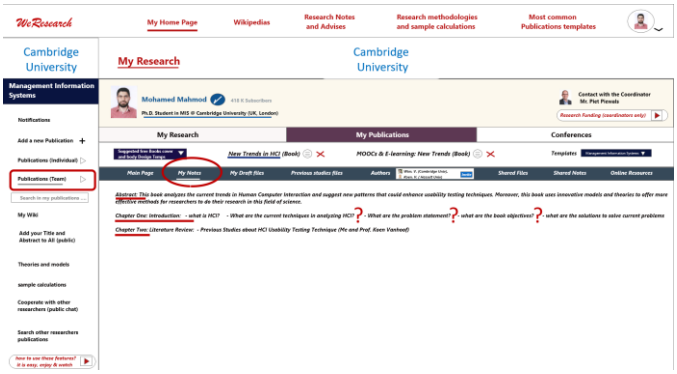


Fig. 6 Researcher's private notes and draft

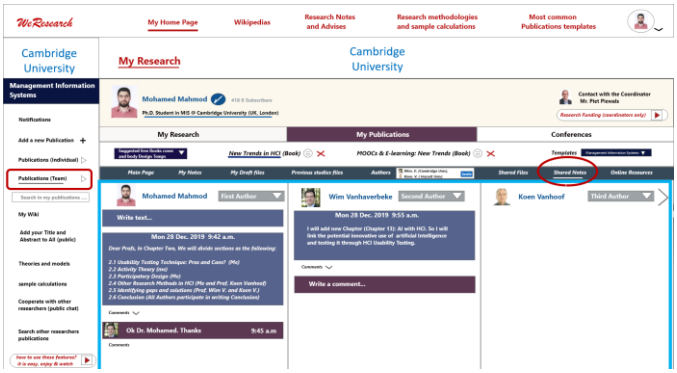


Fig. 7 researchers shared notes

To meet researchers' needs, the design of WeResearch offered journal and article templates which is ready to be downloaded to work quickly without looking for on other websites. Figure 8 shows WeResearch service of offering journal and articles templates.

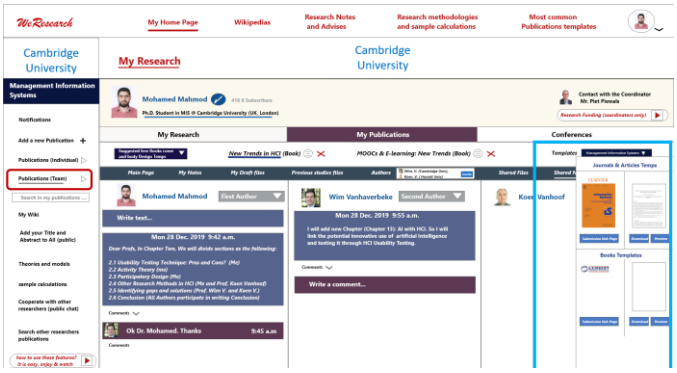


Fig. 8 journal and article templates

In addition, due to researchers demand of the necessity of offering the upcoming conferences and invitations to it,

https://doi.org/10.31436/ijpcc.v12i1.634

WeResearch offers their needs and puts it in its design and protoyping as shown in Figure 9.

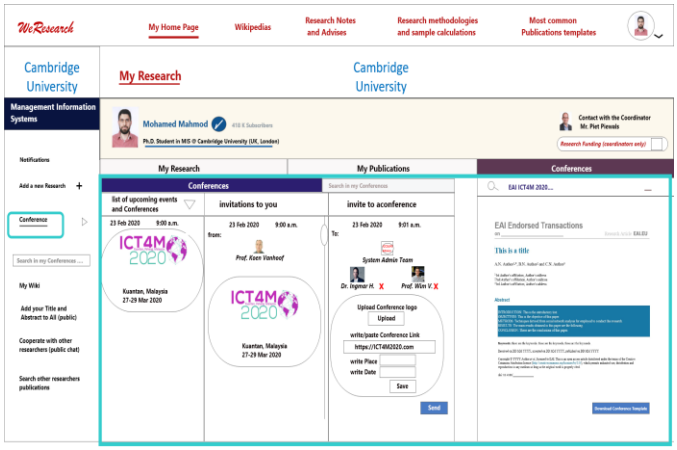


Fig. 9 Conference invitations and templates

Furthermore, WeResearch offers research design types, theories and models per major and clear explanation for each to facilitate the research methodologies as shown in figures 10, 11,12 and 13.

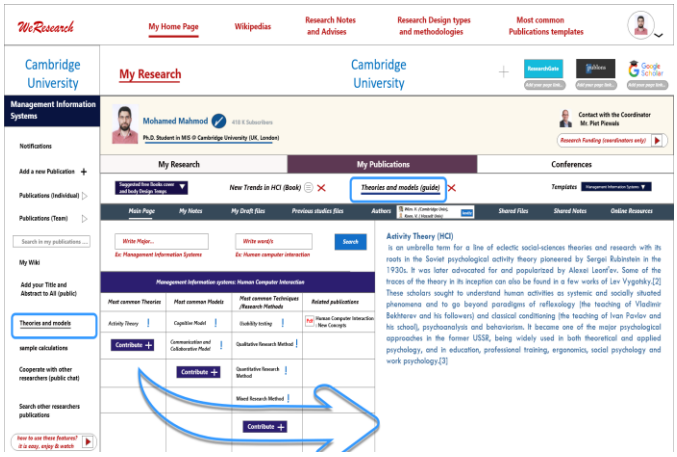


Fig. 10 Theories and models fitted with major

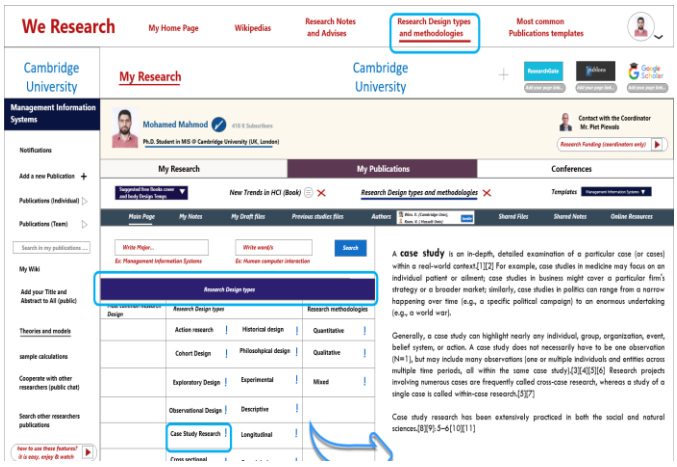


Fig. 11 Research design types definitions for researchers

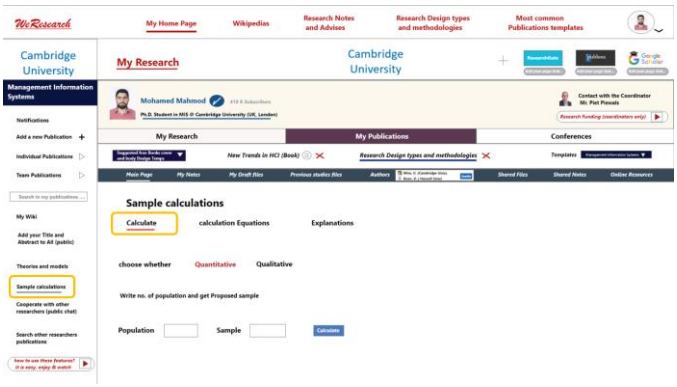


Fig. 12 Sample calculator for quantitative studies

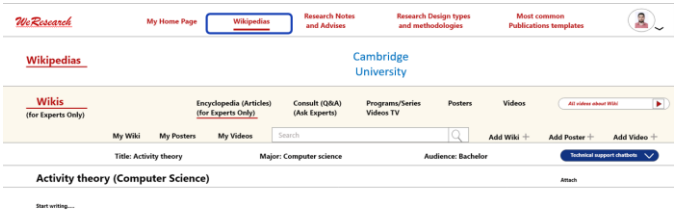


Fig. 13 Encyclopedia of general articles of researchers for students and researchers

V. Evaluation and User Feedback

A. Participants

The evaluation involved the same ten researchers who contributed to the needs assessment stage. Their participation ensured continuity and allowed for a direct comparison between their identified needs and their perceptions of the proposed design.

B. Procedure

Participants engaged in a usability session with the interactive prototype of WeResearch. The sessions lasted 30–45 minutes and were conducted either in person or via online screen-sharing. Participants were first introduced to the purpose of the prototype and then asked to complete representative tasks as shown in Figures 4-12.

The researcher observed participants' interactions with the prototype, noting points of confusion, ease of navigation, and successful task completion. Think-aloud prompts were used to encourage participants to verbalize their reasoning and impressions while completing tasks. At the end of each session, a short debriefing was held to gather overall feedback on usability, feature relevance, and alignment with actual research practices.

C. Data Collection

Feedback was collected through a combination of semi-structured interviews, direct observation, and open discussion during prototype testing sessions. Participants

were encouraged to think aloud while interacting with the prototype, verbalizing aspects they found intuitive, confusing, or particularly useful. The researcher documented task completion, navigation challenges, and spontaneous reactions. Follow-up questions were asked to clarify participants' perspectives on specific features and workflows.

D. Findings and feedback

The prototype testing revealed several key themes: Usability and Clarity: Most participants found the interface intuitive and easy to navigate. However, some suggested simplifying certain menus and reducing the number of steps for common actions, such as uploading files. The usability evaluation was exploratory and primarily perception-based, relying on participant feedback and observational insights rather than standardized quantitative metrics such as SUS scores or task completion time.

Relevance of Features: Core tools—particularly the Workbox for file management, collaborative task assignments, and communication functions—were considered highly relevant and directly applicable to participants' ongoing research workflows.

Enhancements and Desired Features: Participants recommended additional integrations (e.g., with citation managers and external databases), more personalized dashboards, and improved reference handling features.

Overall Satisfaction: Researchers expressed strong satisfaction with the prototype's direction, noting that it successfully addressed many of the gaps they had previously identified in existing research tools.

E. Implications for Design

The feedback validated several core design decisions while highlighting opportunities for refinement. Usability adjustments and requested features will inform the next iteration of the platform. Importantly, the evaluation confirmed that the interactive prototype aligned with researchers' real-world needs, demonstrating the effectiveness of applying SDLC principles to guide user-centered design and development.

VI. CONCLUSION

This study presented WeResearch, a unified research collaboration platform designed to strengthen connections among universities and researchers worldwide. By integrating a qualitative needs assessment with the Software Development Life Cycle (SDLC), the system was shaped according to both user requirements and software design principles. Through interviews with researchers, their needs were identified, analyzed, and transformed into functional specifications, which guided the modeling, design, and prototyping of the platform. UML diagrams were employed to structure the system, clarify relationships, and ensure alignment with user expectations.

The design and prototyping of WeResearch demonstrates how participatory approaches can enhance the relevance and usability of academic software systems. The findings suggest that combining qualitative insights with systematic modeling not only bridges the gap between researchers' needs and technical solutions but also supports the creation of a robust and scalable platform. Ultimately, WeResearch contributes to advancing global academic collaboration by presenting a structured, user-centered, and sustainable prototyped research platform.

The proposed design is intended to be scalable and adaptable to different institutional contexts; however, scalability and performance characteristics have not yet been empirically validated.

This study relied on interviews and observational feedback from ten researchers. Future work is proposed to include quantitative measures, document analysis, and independent evaluators. Future studies are proposed to adopt independent or blinded evaluation strategies.

As is common in early-stage, user-centered design studies, the same group of participants contributed to both the requirements elicitation and the prototype evaluation. This approach supported continuity and coherence in assessing the alignment between user needs and the proposed design. Accordingly, the findings are best interpreted as evidence of the proposed design's alignment with identified user requirements, while comprehensive evaluations of system effectiveness, performance, and scalability are deferred to future stages of implementation and investigation.

This study focuses specifically on the early phases of the Software Development Life Cycle, namely planning, requirements analysis, design, and prototyping. Implementation, deployment, and maintenance phases are beyond the scope of the current work and are planned as future research directions.

ACKNOWLEDGMENT

The author would like to thank the researchers who participated in the usability testing by interacting with the prototype interface and providing valuable feedback and suggestions that helped improve the system.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

The author was responsible for the conceptualization and design of the study, methodology, data collection, system modeling and prototyping, data analysis, and manuscript preparation. The author reviewed and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study involved human participants and was conducted in accordance with ethical research standards.

REFERENCES

[1] P. Nilsen, "Bridging the silos: A comparative analysis of implementation science and improvement science," Implementation Science, vol. 17, no. 1, pp. 1-9, 2022. https://doi.org/10.1186/s13012-022-01185-3

[2] B. Paykamian, "Study: Data silos hinder university improvements," GovTech, Jan. 26, 2023. Available: https://www.govtech.com/education/higher-ed/study-data-silos-hinder-university-improvements

[3] I. Alam and S. Khan, "Statistical analysis of software development models by six-sigma methodology," Journal of Software Engineering, vol. 45, no. 3, pp. 123-135, 2022. https://doi.org/10.1007/s10270-022-00934-5

[4] G. Lemke, "The software development life cycle and its application," Honors Theses, no. 1588, 2018. Available: https://commons.emich.edu/honors/1588

[5] S. Swanlund, "Software development life cycle," Journal of Software Development and Management, vol. 34, no. 2, pp. 45-59, 2024. https://doi.org/10.1016/j.jsdm.2024.01.003

[6] G. Laskovic, "Best research collaboration tools in 2025: Zotero, Paperpile & Collabwriting compared," CollabWriting Blog, Jul. 2, 2025. Available:https://blog.collabwriting.com/best-research-collaboration-tools-in-2025-zotero-paperpile-collabwriting-compared/

[7] Q. M. Yas, "A comprehensive review of software development life cycle methodologies for information systems project management," International Journal of Computer Science and Management Studies, vol. 28, no. 4, pp. 101-115, 2023. https://doi.org/10.2139/ssrn.373800862

[8] I. Alam and S. Khan, "Statistical analysis of software development models by six-sigma methodology," Journal of Software Engineering, vol. 45, no. 3, pp. 123-135, 2022. https://doi.org/10.1007/s10270-022-00934-5

[9] R. A. Alebaikan and L. S. Alsemiri, "Attitudes of graduate students at King Saud University towards digital academic integrity," Journal of Educational & Psychological Sciences, vol. 17, no. 1, pp. 41–59, 2016. Available: https://www.academia.edu/51544166/Attitudes_of_Graduate_Students_at_King_Saud_University_towards_Digital_Academic_Integrity

[10] D. Flaxman, "How to avoid misconduct in research and publishing," Journal of Scholarly Publishing, vol. 43, no. 1, pp. 1–9, 2012. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4852081/

[11] R. Garg, "Methodology for research I," Journal of Clinical and Diagnostic Research, vol. 10, no. 3, pp. JE01–JE03, 2016. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5037944/

[12] D. Sreekumar, "What is research methodology? Definition, types, and examples," Paperpal Blog, 2022. Available: https://paperpal.com/blog/academic-writing-guides/what-is-research-methodology

[13] N. Altowairiki, "Enhancing graduate research skills via action research," International Education Studies, vol. 18, no. 1, pp. 47–58, 2025. Available: https://doi.org/10.5539/ies.v18n1p47

[14] V. Schneider, "Designing an international research experience for students," Frontiers in Education, vol. 8, Article 1154786, 2023. Available: https://www.frontiersin.org/articles/10.3389/feduc.2023.1154786/full

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.636

Interpretable AI for Stroke Prediction: A Structured Approach Using Explainable AI Techniques

Lazeena Tarnim Ranak, Sharyar Wani
Department of Computer Science, Kulliyyah of Information and Communication Technology
International Islamic University Malaysia, Kuala Lumpur, Malaysia

*Corresponding author: sharyarwani@iium.edu.my
(Received: 23rd November 2025; Accepted: 13th December, 2025; Published on-line: 30th January, 2026)

Abstract— The lack of interpretability in AI-driven healthcare diagnostics poses a significant challenge to clinical adoption. This study explores the methodological integration of explainable artificial intelligence (XAI) tools using open clinical prediction dataset, the SCI-XAI pipeline, for stroke risk prediction. We apply multiple machine learning models ranging from white-box approaches (Logistic Regression, Decision Tree, Explainable Boosting Machine) to black-box models (Random Forest, XGBoost, LightGBM, and Multi-Layer Perceptron) and evaluate their trade-offs between predictive accuracy and explainability using techniques such as SHAP, LIME, and ELI5. The study uses a systematic approach involving pre-modeling, modeling, and post-modeling phases, aiming to improve model interpretability for potential use in clinical decision-support contexts. The experimental results show that ensemble models achieve superior accuracy, while traditional models provide inherent transparency. However, the SCI-XAI framework demonstrated that post-hoc explainability tools can extend such transparency to complex models. SHAP-based feature im-portance analysis identifies age, glucose levels, and BMI as the most influential predictors of stroke. The integration of structured explainability into AI based diagnostics helps bridge the gap between algorithmic prediction and clinical interpretability, offering a methodological foundation for more transparent decision-support systems.

Keywords— explainable artificial intelligence (xai), stroke prediction, interpretable machine learning, sci-xai pipeline, clinical decision-making

I. Introduction

In the realm of modern healthcare, artificial intelligence (AI) holds immense potential to revolutionize medical diagnostics and treatment. However, the opacity of AI models presents a significant barrier to their widespread acceptance and utility in healthcare settings. This research project aims to address this challenge by focusing on Explainable AI (XAI) techniques, particularly with the SCI-XAI pipeline, to enhance interpretability and foster actionable insights in AI-driven healthcare diagnostics, specifically using tabular data [1].

Explainable AI (XAI) refers to methods and techniques that make the decision-making pro-cesses of AI models understandable to humans. In healthcare diagnostics, the need for interpret-ability is crucial, as it enables healthcare professionals to effectively utilize AI recommendations. This project integrates several XAI methodologies to ensure transparency and comprehensibility: SHAP (Shapley Additive explanations) provides a unified measure of feature importance, offering both global and local explanations of model predictions; ELI5 (Explain Like I'm 5) simplifies the understanding of model decisions by breaking down complex models into understandable components; and LIME (Local Interpretable Model-agnostic Explanations) generates local explanations for individual predictions, offering insights into how specific features influence model outcomes [2]. Our approach focuses on the SCI-XAI pipeline, a systematic method designed to enhance the interpretability of AI models by focusing on feature selection and extraction [3]. The pipeline involves two main steps – feature selection and feature extraction. The former identifies the most relevant features that significantly impact model predictions, ensuring that the models are not only accurate but also interpretable by highlighting the key factors driving decisions. The latter transforms raw data into a set of informative features that can be used for model training, enhancing the clarity and transparency of the models.

Unlike prior studies that use the SCI-XAI framework or individual explainability tools in isolation, the contribution lies in its single systematic integration of the SCI-XAI pipeline with a broad spectrum of models ranging from interpretable (white-box) to opaque (grey-black-box) architectures and multiple XAI techniques (SHAP, LIME, ELI5). This unified experimental setup enables a structured comparison of interpretability–performance trade-offs across model complexities in a healthcare setting. By applying this

https://doi.org/10.31436/ijpcc.v12i1.636

framework to stroke risk prediction, the study demonstrates how explainability methods can extend transparency from inherently interpretable models to complex ensembles and neural networks, offering methodological rather than clinical novelty in the context of medical AI.

To integrate these techniques into diagnostics, we explore several powerful machine learning algorithms [4], particularly suited for tabular data. These include Adaptive Gradient Boosting, an ensemble technique that improves model performance by iteratively correcting errors from previous models; XGBoost (Extreme Gradient Boosting), known for its efficiency and high performance; and LightGBM (Light Gradient Boosting Machine), optimized for speed and efficiency, effectively handling large datasets and providing quick, accurate predictions [5]. In addition to these ensemble methods, we employ traditional regression models like logistic regression, which establishes relationships between features and outcomes, providing clear insights into data trends, and random forest, an ensemble method that enhances prediction accuracy by averaging the results of multiple decision trees, also offering insights into feature importance.

This paper explores the methodological factors influencing model opacity in healthcare-focused AI and applies XAI-integrated machine learning models for stroke risk pre-diction using tabular data. The focus is on interpretability employing a range of explainable AI (XAI) techniques to provide both quantitative and qualitative insights into the decision-making processes of diverse model architectures.

The objectives of this research are to implement and systematically evaluate a combination of classical machine learning, ensemble, and deep learning models within the SCI-XAI framework, and to quantify interpretability using established XAI tools. This ensures both transparency and measurable insight into model behavior.

It is important to understand the key parameters and factors that influence stroke risk and outcomes before delving into the specific applications of XAI in stroke prediction and other medical diagnoses. These factors serve as the foundation for many AI-driven studies, as they are critical for accurate prediction and diagnosis.

Research has thoroughly explored the multifaceted risk factors and outcomes associated with strokes, highlighting important demographic and health-related considerations. Gender differences in stroke incidence and outcomes are evident, with males having higher incidence and mortality rates between ages 45-74, while females experience higher rates after age 74 due to factors like increased comorbidities [5]. Stroke incidence is higher in women under 30, while men generally have higher incidence during midlife. By age 80, rates equalize or favour women. The lifetime risk of stroke is around 25.1% for women and 24.7% for men, with women typically experiencing strokes 4-6 years later than men [6]. Additionally, hypertension, affecting about 64% of stroke patients, is a major risk factor, linked to ischemic strokes and hemorrhages [7] [8]. Blood pressure thresholds $\geqslant$ 140/90 mmHg are important for stroke detection, and maintaining systolic blood pressure below 140 mmHg is crucial for prevention [7]. Furthermore, heart failure (HF) increases stroke risk by 2-3 times, with a five-fold increase in stroke risk when combined with atrial fibrillation (AF) [8]. Marital status provides a protective effect against stroke outcomes, with married individuals exhibiting lower mortality rates [9]. Job loss and unemployment, particularly in high-stress jobs, are associated with increased stroke risks. Continuous employment, regardless of sector, is linked to lower stroke risks [10].

Further studies have shown nuanced influences like residential areas and associated risks. The impact of residential areas on stroke incidence is explored, showing a slightly higher in rural areas (3.35 per thousand) compared to suburbs (2.90 per thousand) for men and slightly higher in suburbs (2.34 per thousand) than in town centers (2.14 per thousand) for women, although case fatality was lower in rural areas [11]. The study of stroke risk associated with average glucose levels in 12,321 participants indicates a clear correlation between higher glucose levels and in-creased stroke risk. Participants with diabetes had a 3.5% stroke incidence, compared to 2.2% in non-diabetic participants, with higher glucose levels ($\geqslant$ 126 mg/dL) increasing stroke risk by 78% (HR 1.78) compared to those with glucose levels of 90-99.9 mg/dL [12]. Blood pressure, BMI, cholesterol, and glucose levels were all significantly higher in participants with diabetes, reinforcing the importance of managing metabolic factors for stroke prevention [13], [14].

Now, with these key stroke-related parameters in mind, we turn to the application of XAI in medical diagnostics, particularly focusing on stroke and other health conditions. One study ap-plied machine learning models combined with XAI tools such as ELI5 and LIME for stroke pre-diction using EEG signals, achieving around 80% accuracy [15]. Another study enhanced intra-operative decision-making in ovarian cancer surgery by integrating XAI with human factors and domain knowledge, demonstrating the value of explainability in real-time clinical settings [16]. These studies collectively delve into the realm of Explainable AI (XAI) in medical diagnostics and decision support. They explore various methodologies and applications, from interpreting AI-generated clinical decision support systems (CDSS) through human-centered design approaches to evaluating XAI methods like Grad-CAM and Eigen-CAM on medical imaging datasets [17]. Moreover, they introduce innovative solutions such as the SCI-XAI pipeline for clinical prediction models and a geriatric MDSS incorporating XAI elements [3] [18]. The studies emphasize the importance of

https://doi.org/10.31436/ijpcc.v12i1.636

interpretability, transparency, and user involvement in AI systems [19], aiming to enhance diagnostic accuracy, trust, and ultimately patient outcomes. Another study evaluated Grad-CAM and Eigen-CAM visual XAI methods on the VinDrCXR Chest X-ray Abnormalities Detection dataset using YOLO models. It highlighted the limitations of these methods in explaining model decisions and cautioned against sole reliance on their outputs. The authors recommended validating results through sample images, manual evaluations, or automated methods to find the most suitable XAI tools for specific domains. Challenges such as partiality, overfitting, and limited interpretability for complex models were discussed, emphasizing the need for careful application and additional verification in visual XAI [20]. These findings pave the way for further research and development in XAI to address practical challenges and improve medical decision-making processes across different domains [21].

The study in [22] presents a systematic comparison of post-hoc explainability methods i.e. LIME, SHAP, and Anchors which are evaluated on healthcare datasets for fidelity, stability, separability, computational efficiency, and bias detection. Results indicate that LIME provides dis-tinct explanations but with lower fidelity, SHAP demonstrates superior speed and bias detection, while Anchors offers intuitive rule-based outputs. The research highlights critical trade-offs be-tween interpretability and reliability, underscoring the complementary strengths of these approaches. A related investigation in [23] focuses on stroke prediction using an ensemble framework achieving approximately 96% accuracy. Five explainability techniques; SHAP, LIME, ELI5, Anchors, and QLattice were applied to interpret predictions. Across all methods, age, BMI, blood glucose, and hypertension emerged as key contributing factors, consistent with clinical evidence. The integration of multiple XAI tools ensured alignment between model behavior and medical understanding, supporting transparency and clinician trust.

The work in [24] introduces an interpretable ensemble model to differentiate iron-deficiency anemia from aplastic anemia, marking the first use of interpretable AI for this diagnostic task. Employing SHAP, LIME, ELI5, QLattice, and Anchors, the study identified platelet count, mean cell volume, haemoglobin, and white blood cell count as dominant features, enhancing the transparency of model reasoning and clinical reliability. Similarly, research on type 2 diabetes prediction in [25] developed an ensemble framework achieving 92.5% accuracy (AUC $\approx$ 0.98), incorporating SHAP, LIME, EBM, and counterfactual analysis to interpret predictions. Key risk factors such as BMI, age, and physical activity were consistently highlighted, demonstrating both interpretability and clinical coherence. Finally, a comparative analysis in [26] evaluated traditional statistical and modern machine learning approaches for stroke risk prediction using logistic regression, Cox regression, Bayesian networks, EBM, and XGBoost with SHAP-based explanations. XGBoost achieved the highest performance (C-statistic = 0.89; F1-score = 0.80), followed closely by EBM (C-statistic = 0.87). Both models identified atrial fibrillation, hypertension, age, and HDL cholesterol as major predictors, aligning with established clinical knowledge.

Building on prior works that primarily compared interpretability tools, this study extends those efforts by applying a comparative framework for stroke risk prediction that integrates white-box and black-box models within the SCI-XAI pipeline. This integration enables a systematic evaluation of both global and local explanations, offering balanced insights into accuracy and interpretability.

Recent work by Challenging the Performance-Interpretability Trade-Off: An Evaluation of Interpretable Machine Learning Models [27] provides strong empirical evidence that the performance–interpretability trade-off is not inevitable. In a large-scale evaluation of generalized additive models (GAMs) such as the Explainable Boosting Machine (EBM) against black-box baselines across twenty tabular datasets, the authors demonstrated that advanced GAMs could achieve competitive predictive performance while remaining inherently interpretable. This finding supports our selection of EBM as a key component within the SCI-XAI pipeline and motivates our comparative inclusion of both transparent and opaque models to evaluate explainability and performance in stroke prediction.

In closing, this introduction has outlined the motivation, methodological framing and key analytical lenses of our study. With a clear view of the interpretability-performance landscape and the domain-specific foundations of stroke risk factors, the next section details the experimental methodology and SCI-XAI modeling framework.

II. Experimental Setup

A. Dataset Description

The dataset employed in this study is the Stroke Prediction Dataset, publicly available on Kaggle, curated by fedesoriano. It contains a total of 5,110 patient records; each aimed at sup-porting the prediction of stroke occurrence based on various demographic and health-related attributes. The target variable is stroke, a binary classification label where 1 indicates the occurrence of a stroke, and 0 indicates no stroke. A closer examination of the class distribution reveals a strong imbalance, with only 249 instances (approximately 4.87%) indicating stroke events, and the remaining 4,861 instances (approximately 95.13%) representing non-stroke cases. This pronounced imbalance is a crucial factor that informs model development and evaluation strategies.

https://doi.org/10.31436/ijpcc.v12i1.636

The dataset comprises 11 features, encompassing both categorical and numerical variables. These include gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, and smoking_status. Together, these features provide a comprehensive foundation for stroke risk modeling but require careful preprocessing due to mixed data types and imbalance.

B. Data Modeling Process

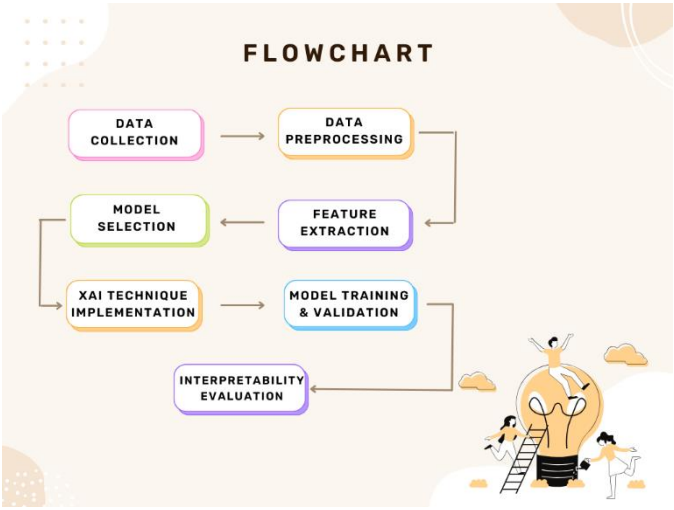


Fig. 1 Data Mining Flowchart.

Figure 1. outlines the high-level progression of this study, starting from data collection to interpretability evaluation. It captures the essential stages including preprocessing, model selection, and pipeline strategies, followed by model training and the integration of explainable AI (XAI) techniques to ensure transparent and accountable predictions in healthcare diagnostics.

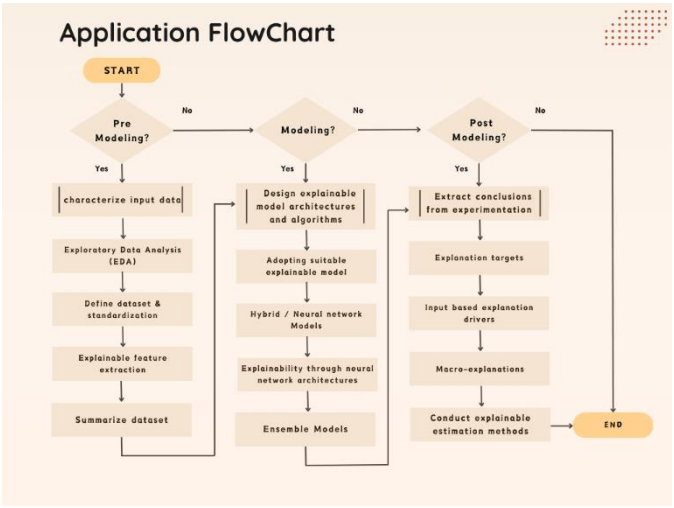


Fig. 2 Modeling Flowchart.

Figure 2. illustrates the comprehensive modelling workflow employed in this study, which encompasses pre-modelling, modelling, and post-modelling stages, with a strong emphasis on interpretability.

The experimental methodology was carefully structured into three key phases: pre-modeling, modeling, and post-modeling, forming a systematic approach to developing a stroke prediction framework that is both interpretable and healthcare relevant. The research began with collecting and curating medical data from openly available sources to ensure a strong foundation for predictive modeling. The dataset included demographic, lifestyle, and clinical attributes such as age, gender, hypertension, heart disease, glucose levels, BMI, marital status, work type, residence type, and smoking status, with stroke occurrence as the target variable. These features were selected to capture key risk factors while ensuring that the model's predictions remain interpretable for healthcare professionals.

In this phase, data preprocessing was a critical step, involving normalization, standardization, and handling of missing values. Missing numerical data, including BMI, were imputed using median values to maintain central tendency without distorting variance, while missing categorical data such as smoking status were labeled as "Unknown" to retain sample completeness. Although advanced imputation strategies (e.g., MICE) may offer higher fidelity, the chosen method provided consistency and interpretability within the comparative experimental framework.

To ensure compatibility across diverse model architectures, preprocessing steps were tailored to each model type rather than applied uniformly. Categorical variables were transformed based on the algorithm's sensitivity: label encoding was applied for tree-based models such as XGBoost and Random Forest, while one-hot encoding was used for distance-based and linear models like Logistic Regression to preserve feature relationships. Numerical features (e.g., age, glucose level, and BMI) were standardized using StandardScaler only for models sensitive to feature magnitude such as Logistic Regression, MLP, and gradient boosting methods to facilitate stable optimization and faster convergence. Tree-based models (Decision Tree, Random Forest, LightGBM) were trained using unscaled numerical inputs, as these algorithms rely on relative thresholds rather than distance metrics. Exploratory Data Analysis (EDA) provided insights into feature distributions, correlations, and key stroke risk factors. Statistical techniques like Pearson correlation, mutual information, and variance inflation factor (VIF) were applied to assess feature relevance and detect multicollinearity.

To refine the dataset, feature selection and extraction were carried out using the SCI-XAI pipeline, incorporating recursive feature elimination (RFE) and correlation-based filtering to retain only the most significant predictors. This

https://doi.org/10.31436/ijpcc.v12i1.636

helped improve model performance while ensuring interpretability. Additionally, feature engineering was applied to incorporate domain knowledge. For example, gender was treated as a binary risk factor due to higher stroke prevalence in females, and age-based stroke risk was introduced using a 45-year threshold, recognizing that stroke risk increases with age. Clinical factors such as hypertension and heart disease were directly included, while marital status was encoded to reflect potential lifestyle stability, which is linked to better health outcomes. Work type was categorized to account for stress levels, distinguishing between high-stress jobs (private/self-employed) and more stable employment (government jobs). Residence type was encoded to capture healthcare access disparities between rural and urban areas.

Metabolic and lifestyle factors were also considered. Individuals with glucose levels in the prediabetic range (126–139.9 mg/dL) were flagged under glucose stroke risk, as elevated glucose levels are associated with cardiovascular events. BMI risk was assigned based on standard classifications for underweight (<18.5) and obesity ($\geqslant$30), recognizing their impact on stroke risk. Smoking status was categorized to distinguish between active smokers and former/non-smokers.

The focus of the modeling phase was to implement explainable machine learning models that balance predictive accuracy with interpretability. Guided by the SCI-XAI framework, both white-box and black-box models were explored to analyze stroke risk factors through multiple methodological perspectives. White-box models, including Logistic Regression, Decision Trees, and the Explainable Boosting Machine (EBM), were first implemented to establish baseline performance and interpretability benchmarks. These models provided transparent insights into in-dividual feature contributions, enabling straightforward interpretation of clinical risk indicators such as age, glucose level, and hypertension. EBM offered additive feature interactions that preserved interpretability while slightly improving accuracy over traditional linear methods. Subsequently, gray and black-box models such as Random Forest, XGBoost, and LightGBM were introduced to capture non-linear interactions and improve predictive robustness. An ensemble-based Adaptive Gradient Boosting model was also experimented with to integrate the complementary strengths of XGBoost and LightGBM, achieving improved stability and predictive performance across evaluation metrics.

To capture complex, non-linear relationships in the data, deep learning models were also explored. A Multi-Layer Perceptron (MLP) neural network was trained to recognize intricate feature interactions. The MLP consisted of two hidden layers (128 and 64 neurons), ReLU activation functions, and dropout regularization (0.3) with batch normalization to prevent overfitting. Although CNN and LSTM architecture were initially explored for potential pattern and sequence modeling, their performance was not sufficiently generalizable for inclusion in the final report due to the dataset's non-sequential nature and limited sample size. Hyperparameter tuning was conducted using a hybrid strategy combining grid search and Bayesian optimization to ensure equitable comparison across models. For the white-box models, hyperparameters were configured based on empirical testing and literature-recommended defaults. Logistic Regression employed ℓ_2 regularization (penalty='l2') with an increased iteration limit (max_iter=1000) to ensure convergence. The Decision Tree classifier used the entropy criterion with a constrained maxi-mum depth of 3 and minimum samples per split set to 2 to prevent overfitting. For the Random Forest model, the number of estimators (100–500), tree depth, and feature subset sizes were adjusted to improve generalization and reduce overfitting. Among the gradient boosting models, XGBoost and LightGBM were fine-tuned for learning rate (0.01–0.3), tree depth (3–10), and number of estimators (100–300) to achieve optimal balance between bias and variance. For the deep learning model, the Multi-Layer Perceptron (MLP) was tuned for learning rate (0.001–0.01), number of neurons per layer (64–256), batch size (32–128), and dropout rate (0.2–0.5). The adaptive Aquila Optimizer was employed to accelerate convergence and prevent stagnation during training, allowing faster stabilization without compromising accuracy. The dataset was divided into training and testing sets using an class–20 stratified split to preserve class balance. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics, ensuring a fair and consistent comparison between models differing in complexity and interpretability.

To enhance model reliability and ensure that the evaluation was not biased toward a specific data partition, k-fold cross-validation (k=5) was employed during model training and hyperparameter tuning. This involved dividing the dataset into five equally sized folds, iteratively training on four folds and validating on the remaining one. The process was repeated for all folds, and performance metrics were averaged to obtain a more robust estimate of each model's generalization capability. Cross-validation was particularly applied to the Logistic Regression and Decision Tree, and Random Forest as well as to ensemble approaches such as, XGBoost, and LightGBM, to assess their stability across multiple data splits. For the MLP model, due to computational constraints, an 80–20 stratified split with early stopping and validation monitoring was used instead of full k-fold evaluation.

To reduce overfitting risks arising from the small and imbalanced dataset, stratified sampling and class-balancing via SMOTE were applied. Alternative imbalance-handling strategies such as random under-sampling were also evaluated conceptually but not adopted, as preliminary

https://doi.org/10.31436/ijpcc.v12i1.636

tests indicated a loss of minority-class representation. Future work could further explore hybrid sampling methods to improve generalization. Additionally, validations were employed using re-peated train–test splits. Ensemble models were regularized through depth constraints and learning-rate tuning to prevent excessive fitting to noise as discussed earlier.

The final phase, post-modeling, aimed at making the model's predictions interpretable. Various Explainable AI (XAI) techniques were used to bridge the gap between algorithmic insights and medical decision-making. Feature importance was analyzed using Shapley Additive explanations (SHAP), providing a breakdown of how each variable contributed to stroke predictions. SHAP values offered both global insights (overall feature impact) and local explanations (personalized risk factors), allowing clinicians to understand individual predictions.

Key features such as age, hypertension, glucose level, and BMI were highlighted as primary predictors, aligning with clinical stroke risk assessments. Macro-explanation methods were used to validate decision rules across multiple patient samples, ensuring model consistency. Local interpretability was addressed using Local Interpretable Model-agnostic Explanations (LIME), which generated simplified models to explain individual predictions.

Model evaluation metrics including accuracy, precision, recall, and F1-score were analyzed to assess classification performance. Additionally, the stability of XAI explanations was examined by checking SHAP value variations across different data subsets. This step ensured that the model remained reliable and its interpretations consistent across diverse patient groups.

III. Experimental Results

This section presents the outcomes of the implemented machine learning models across both standard and SCI-XAI pipelines. Emphasis is placed not only on traditional performance metrics such as accuracy, F1-score, and ROC-AUC, but also on the interpretability of the models. Through both global and local explanation techniques, the results are analyzed to evaluate their reliability and clinical decision-making in stroke prediction. To ensure robust evaluation, relevant models were assessed using both hold-out testing and 5-fold cross-validation.

A. Logistic Regression

We first trained a Logistic Regression model as part of a standard machine learning pipeline, including data balancing using SMOTE to address class imbalance. The baseline performance showed promising but nuanced results: the training set achieved an accuracy of 85.6% (AUC = 0.91, weighted F1 = 0.86), while the test set achieved 71.6% accuracy (AUC = 0.84, weighted F1 = 0.79).

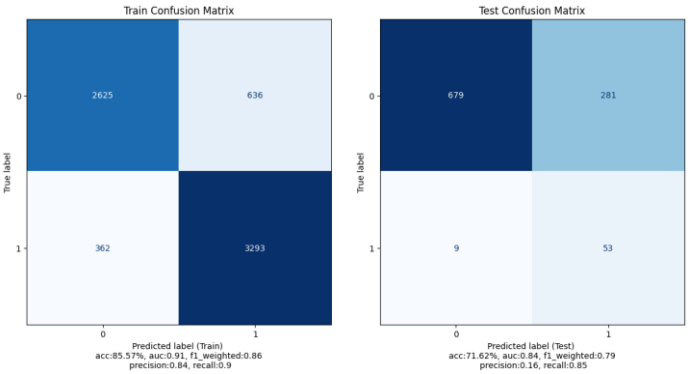


Fig. 3 LogReg Confusion Matrices.

The confusion matrices shown in Figure 3 reveal a familiar pattern for imbalanced datasets: the model achieves high precision (0.99) and F1 (0.82) for the majority class (non-stroke), but low precision (0.16) and moderate F1 (0.27) for the minority class (stroke) despite a strong recall of 0.85. This indicates that while the model successfully detects most stroke cases, it does so at the expense of generating false positives, an expected trade-off in high-sensitivity healthcare screening.

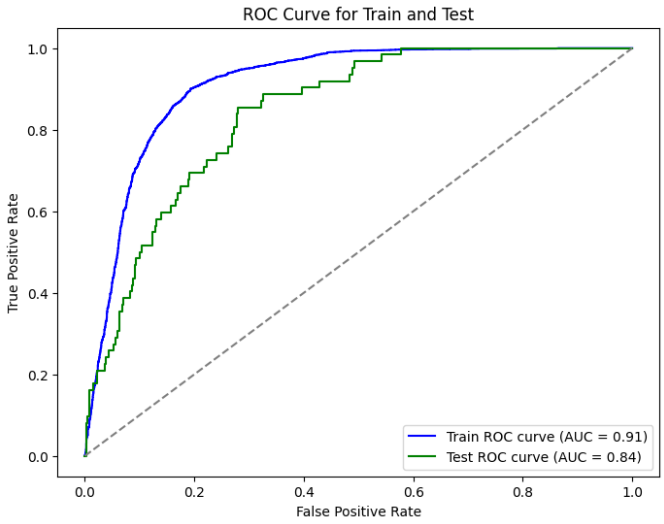


Fig. 4 LogReg ROC-AUC Score.

The ROC curves at Figure 4 confirmed good separability, with AUC scores of 0.91 (train) and 0.84 (test), suggesting minimal overfitting and decent generalization. Cross-validation further validated this stability, yielding a mean AUC of 0.91 ± 0.01, suggesting that the model's predictive performance remained consistent across different data folds. However, these numerical metrics alone do not offer clinical practitioners' insight into why the model makes its predictions, a crucial requirement in sensitive domains like healthcare.

To address this, we integrated the model into the SCI-XAI pipeline, using state-of-the-art interpretability techniques: SHAP, LIME, and ELI5.

https://doi.org/10.31436/ijpcc.v12i1.636

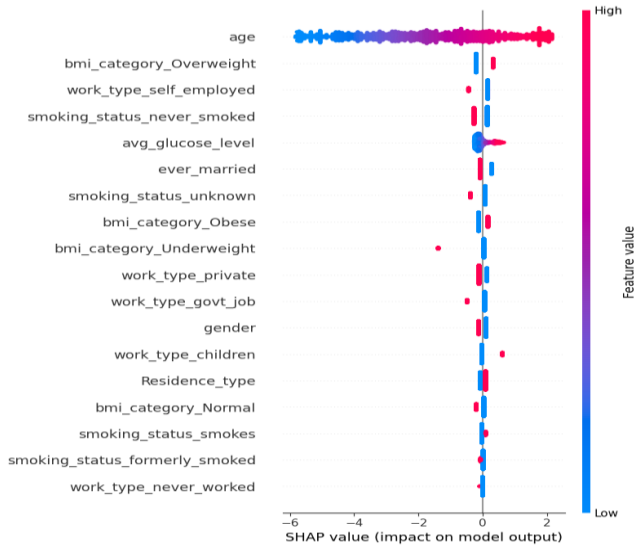


Fig. 5 LogReg SHAP Summary

Figure 5 plot clearly demonstrates that age is the dominant predictor of stroke, with higher age values strongly contributing to positive stroke predictions (as shown by red points on the right). Other impactful features include BMI categories (Overweight, Obese, Underweight) and occupational indicators such as work_type_self_employed and smoking status. This offers clinicians an intuitive way to understand the feature hierarchy and how individual features shift the stroke risk up or down (see Figure 6).

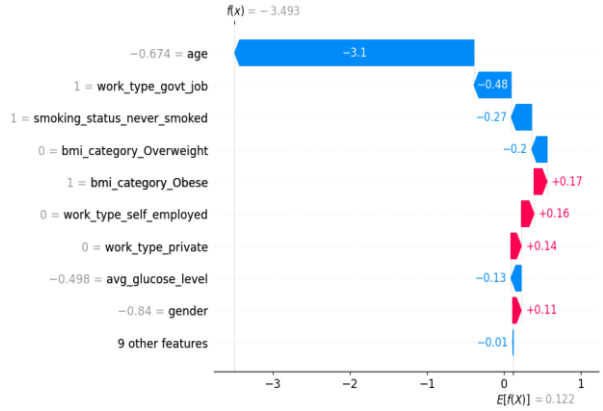


Fig. 6 LogReg SHAP force plot.

Drilling into an individual prediction, we observe that a lower age (-0.674) combined with working in government jobs and a never-smoked status reduces risk (blue bars), while obesity-related BMI and self-employment slightly increased risk (red bars). This granular level of ex-planation is invaluable for personalized risk assessments.

This dependence plot illustrates a strong positive linear relationship between age and stroke risk, with SHAP values increasing steadily from approximately −6 to +2.5 as normalized age rises (see Figure 7).

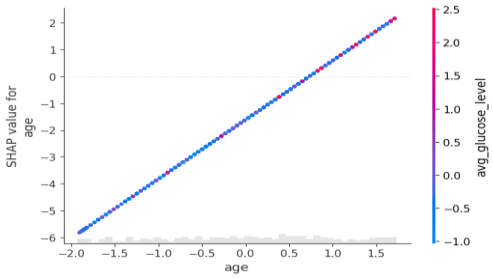


Fig. 7 LogReg SHAP dependence plot.

This trend confirms that higher age substantially elevates the model's predicted stroke risk. The color gradient, representing average glucose level, further suggests that elevated glucose levels amplify this age-related effect, reinforcing the combined influence of metabolic and demographic factors in stroke prediction.

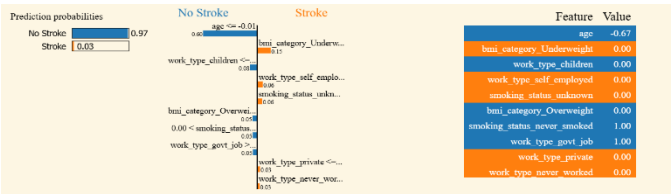


Fig. 8 LogReg LIME Explanations.

Figure 8 emphasize age (coefficient 0.60) as the strongest risk driver, followed by BMI category (Underweight), work_type_children, and work_type_self_employed, which align well with the SHAP findings. The LIME output also includes feature thresholds (e.g., age <= -0.01), making it easy to visualize which side of a decision boundary the patient lies on.

TABLE 1
LOGREG ELI5 EXPLANATIONS.

y=1 top features

Weight[?]	Feature
+2.207	age
+0.645	work_type_children
+0.531	bmi_category_Overweight
+0.291	bmi_category_Obese
	... 3 more positive ...
	... 6 more negative ...
-0.415	smoking_status_never_smoked
-0.465	smoking_status_unknown
-0.549	work_type_govt_job
-0.597	work_type_self_employed
-1.126	<BIAS>
-1.423	bmi_category_Underweight

y=0 (probability **0.970**, score **-3.493**) top features

Contribution[?]	Feature
+1.488	age
+1.126	<BIAS>
+0.549	work_type_govt_job
+0.415	smoking_status_never_smoked
+0.122	ever_married
+0.097	avg_glucose_level
+0.085	Residence_type
-0.096	gender
-0.291	bmi_category_Obese

https://doi.org/10.31436/ijpcc.v12i1.636

Table 1 further strengthens interpretability by breaking down the model's coefficients. For the positive class (stroke), age (+2.207) again tops the list, with BMI and occupational status providing additional weight. Interestingly, smoking_status_never_smoked (-0.415) and work_type_govt_job (-0.549) appear as significant negative contributors, echoing the SHAP force plot. The model's inherent bias term also plays a non-negligible role.

Together, these visualizations and explanations reveal a consistent pattern: age, BMI, and occupational/smoking factors as key stroke risk drivers. The SCI-XAI pipeline enhances the logistic regression model by turning raw predictions into clear, patient-specific explanations, helping interpret and communicate risk effectively by combining solid predictive performance with both global and local interpretability.

B. Decision Trees

The decision tree model achieved a test accuracy of ~77.98%, with a ROC AUC of 0.81 and a weighted F1 score of 0.83. These metrics suggest solid predictive performance, especially the relatively high recall (71%) for the minority class (stroke).

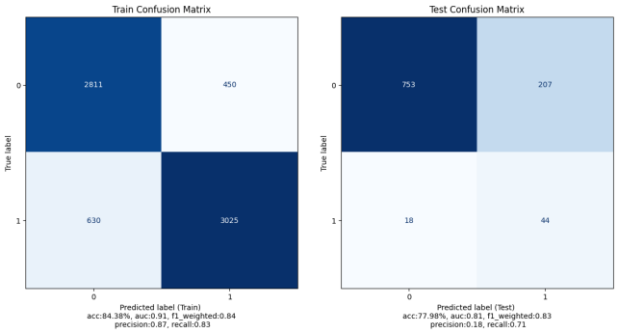


Fig. 9 DT Confusion Matrices.

As demonstrated in Figure 9, the confusion matrices reveal a balanced ability to detect both stroke and non-stroke cases during training, but the test matrix shows some over-prediction of the majority class (non-stroke), which is typical for decision trees even after SMOTE balancing.

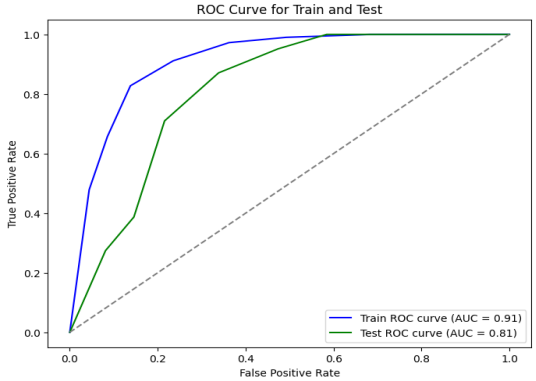


Fig. 10 DT ROC-AUC Scores.

The ROC curves (Figure 10.) demonstrate good class separation capability, comparable to logistic regression but with slightly reduced generalization on the test set (AUC = 0.81 vs. 0.91 for training). Cross-validation reinforced these findings, yielding a mean AUC of 0.91 ± 0.01, indicating that while the model performs consistently across folds, some overfitting tendencies remain due to its hierarchical nature. To further unpack how individual features influence predictions, we applied SHAP analysis to gain both global and local interpretability beyond the raw tree splits.

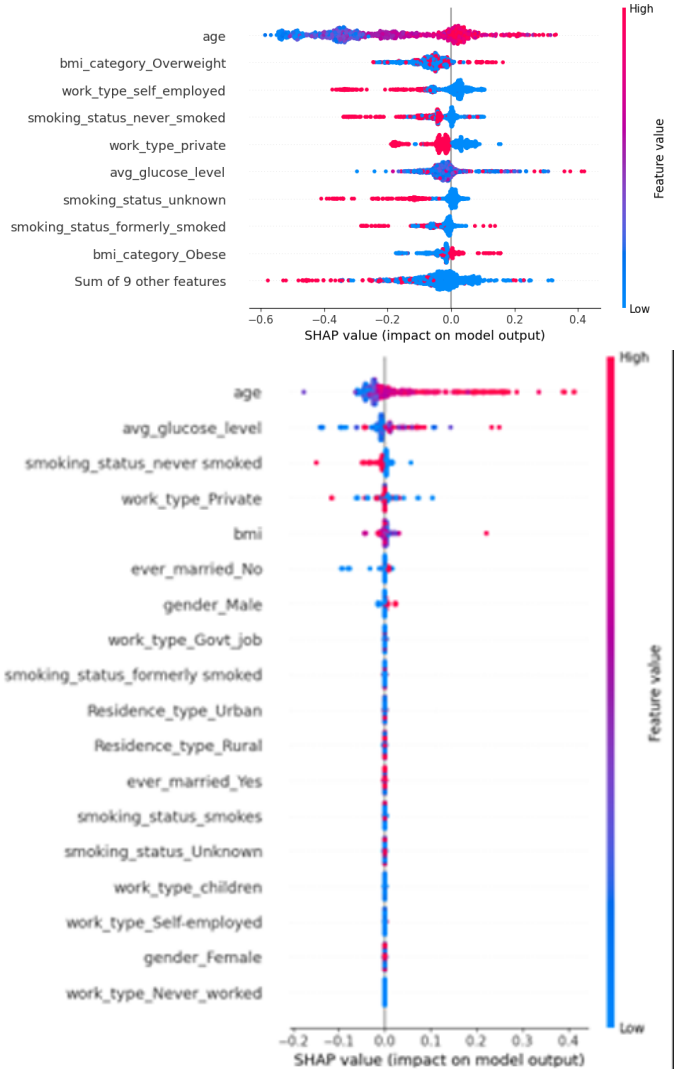


Fig. 11. DT SHAP summary plots.

In Figure 11. the SHAP summary plots emphasizes that age (~0.48) remains the dominant predictor, followed by BMI-related features (~0.14 and ~0.11), smoking status (notably smoking_status, never_smoked, ~0.27), and work type (~0.10) mirroring the findings from the logistic regression but with subtle shifts in feature impact distribution.

https://doi.org/10.31436/ijpcc.v12i1.636

TABLE II
DT LIME EXPLANATIONS

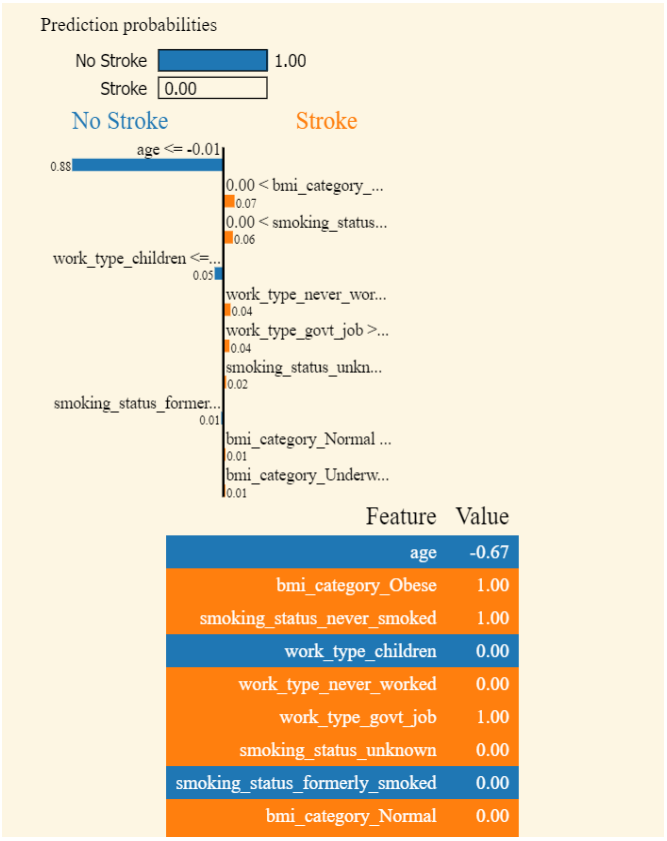


Table 2. reinforces that age and BMI categories are primary drivers of individual predictions, with clear rule-based logic (e.g., "age ⩽ -0.01 contributes 88% toward a 'No Stroke' prediction), reflecting the deterministic nature of decision trees.

Table III
DT ELI5 Explanations.

Explained as: decision tree

Decision tree feature importances; values are numbers 0 <= x <= 1; all values sum to 1.

Weight	Feature
0.4767	age
0.1475	avg_glucose_level
0.0672	smoking_status_never_smoked
0.0390	smoking_status_formerly_smoked
0.0336	smoking_status_unknown
0.0333	bmi_category_Overweight
0.0325	work_type_self_employed
0.0308	bmi_category_Obese
0.0294	work_type_govt_job
0.0250	Residence_type
… 8 more …	

The ELI5 tree breakdown (detailed in the appendix 1 and visualized in Table 3. & Table 4.) provides direct interpretability through the tree structure itself, showing precise splits and thresholds offering clinicians a transparent, step-by-step rationale for each prediction. For example, age contributes +0.494 to the final score (out of a base bias of +0.500), while other features like work_type_self_employed and avg_glucose_level have smaller contributions (+0.004 and +0.002 respectively), reinforcing the dominance of age in prediction. This not only lists feature importance but shows how individual features push predictions higher or lower for a specific case, making it especially useful in medical settings.

Table IV
DT ELI5 Explanations

Prediction Explanation for Predicted Class

Explained as: decision path

Features with largest coefficients.

Feature weights are calculated by following decision paths in trees of an ensemble (or a single tree for DecisionTreeClassifier). Each node of the tree has an output score, and contribution of a feature on the decision path is how much the score changes from parent to child. Weights of all features sum to the output score or proba of the estimator.

Caveats:
1. Feature weights just show if the feature contributed positively or negatively to the final score, and does not show how increasing or decreasing the feature value will change the prediction.
2. In some cases, feature weight can be close to zero for an important feature. For example, in a single tree that computes XOR function, the feature at the top of the tree will have zero weight because expected scores for both branches are equal, so decision at the top feature does not change the expected score. For an ensemble predicting XOR functions it might not be a problem, but it is not reliable if most trees happen to choose the same feature at the top.

y=No Stroke (probability **1.000**) top features

Contribution[?]	Feature
+0.500	<BIAS>
+0.494	age
+0.004	work_type_self_employed
+0.002	avg_glucose_level

Overall, the SCI-XAI pipeline transforms raw model predictions into human-readable, actionable insights. While decision trees are naturally interpretable, SCI-XAI's layered explanations that combine global feature importance (SHAP/ELI5) and individualized prediction paths (LIME) enhance clarity and communication. This makes the tool particularly valuable in clinical contexts, where both performance and explanation are crucial.

C. *Explainable Boosting Machines (EBM)*

Explainable Boosting Machine (EBM) offers a highly interpretable model architecture that combines the power of machine learning with human-understandable outputs. One of its major advantages lies in its interactive nature: through a dropdown interface, users can dynamically inspect global and local explanations for each feature and interaction term. This transparency al-lows clinicians and stakeholders to dissect individual model decisions easily, fostering deeper understanding of the predictive patterns. Notably, the EBM was trained using the SCI-XAI pipeline without requiring additional class imbalance handling, thanks to its inherent robustness in managing such data distributions.

https://doi.org/10.31436/ijpcc.v12i1.636

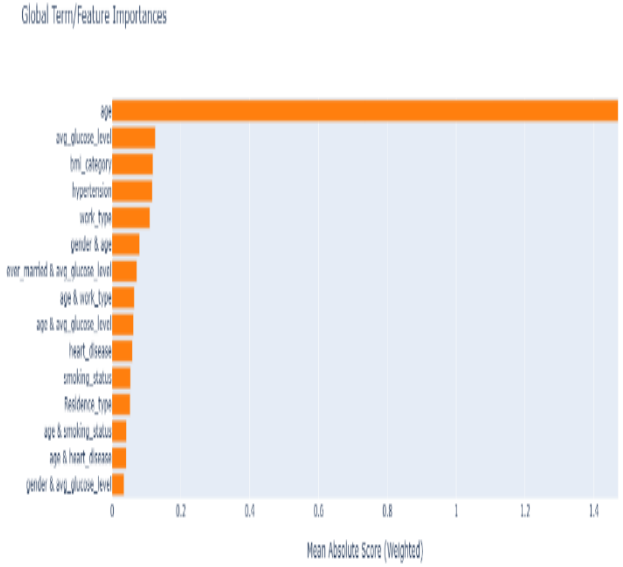


Fig. 12 EBM Explainable Summary.

As shown in Figure 12, the summary plot highlights that age dominates the feature contributions with a mean absolute weighted score around 1.4, followed by average glucose level (~0.15), BMI category (~0.14), and hypertension (~0.13). Other contributing features include work type (~0.12) and notable pairwise interactions such as ever_married & avg_glucose_level (~0.08).

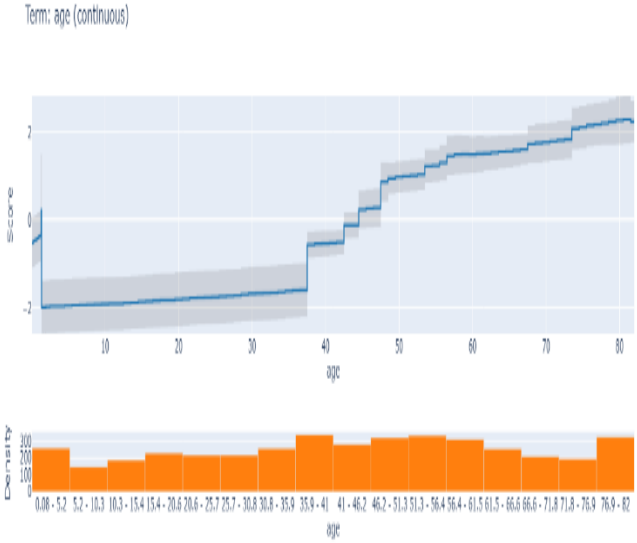


Fig. 13. EBM Global Explanation: Age.

Examining age (Figure. 13) in detail reveals a clear, near-linear increase in risk contribution after age 40, with a sharp inflection between 50–60 years, reaching a maximum contribution of around +2.5 for older patients. These matches established clinical knowledge, reinforcing age as a principal driver of stroke risk.

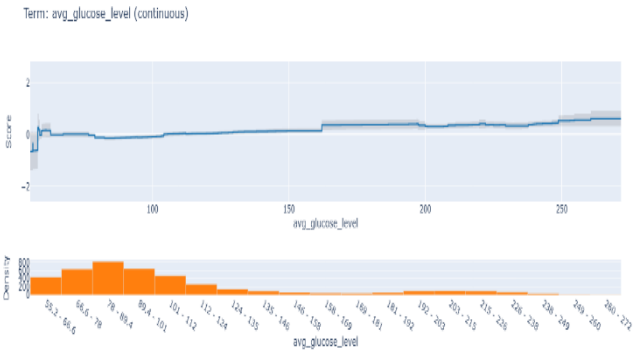


Fig. 14 EBM Global Explanation: Average Glucose Level.

The average glucose level above shows a subtler but still meaningful pattern: risk rises modestly after glucose levels of ~150 mg/dL, peaking at contributions around +1.5, though most density is clustered below 100 mg/dL (see Figure 14).

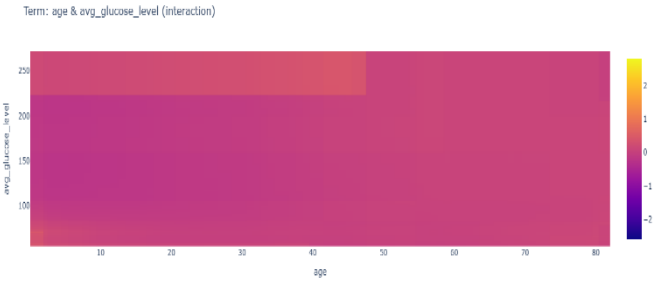


Fig. 15 EBM Global Explanation: Age vs Avg Glucose Lvl.

Figure 15. provides deeper insight into how these two features interplay. The heatmap indicates that higher risk is concentrated in the upper-right quadrant, where both age > 60 and glucose > 200 mg/dL, reflecting compounding effects of these factors.

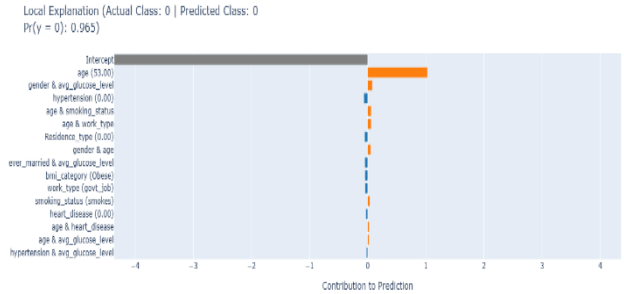


Fig. 16 EBM Local Explanation: (Actual = 0, Predicted = 0, Pr(y=0)=0.965).

Finally, the local explanations at Figure 16. provides us with case-by-case insights, showing exactly how each feature influenced a specific prediction. For instance, in one patient example Actual = 0 (No Stroke), Predicted = 0 (No Stroke), with a predicted probability of 0.965, the model

https://doi.org/10.31436/ijpcc.v12i1.636

was highly confident in its decision. Here, age (53 years) was the dominant factor reinforcing the "No Stroke" prediction, contributing the largest positive weight. Additional, smaller positive contributions came from the combined effect of gender & average glucose level, while features like BMI (Obese) and smoking status (smokes) applied slight opposing pressure (negative contributions), nudging the prediction toward stroke risk but not enough to outweigh the stronger protective signals. This case demonstrates how EBM offers a transparent, numeric breakdown of in-dividual risk profiles, which can be invaluable for clinical decision-making.

D. Random Forest

To extend the exploration beyond white-box models, Random Forest recognized as a grey-box model was applied. While Random Forests are robust and capable of capturing complex, non-linear relationships, they traditionally lack inherent interpretability. To address this gap, we utilized the SCI-XAI pipeline to generate both global and local explanations alongside standard statistical machine learning evaluations (see Figure 17).

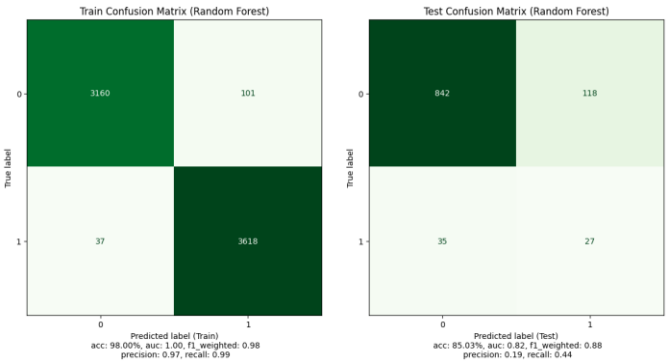


Fig. 17 RF Confusion Matrices.

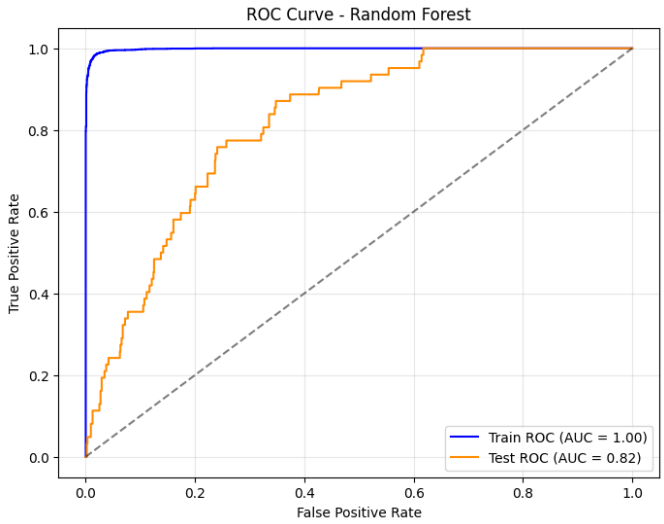


Fig. 18 RF ROC Scores.

The model, trained using a standard machine learning pipeline enhanced with SMOTE for class balancing, yielded solid predictive results: a test accuracy of 85.03% and a ROC AUC score of 0.82, as observed in the classification report. Precision and recall for the minority class (stroke) were 0.19 and 0.44 respectively, reflecting the ongoing challenge of sensitivity in imbalanced medical datasets. To ensure robustness and mitigate potential overfitting, 5-fold cross-validation was performed on the training data. The Random Forest achieved a mean cross-validated ROC-AUC of 0.9952 ± 0.0033, indicating stable and consistent performance across folds despite slight overfitting tendencies observed in single split evaluations (see Figure 18).

To add interpretability, the SCI-XAI pipeline transformed this grey-box model into an explainable one. The SHAP summary plot (see Figure. 19) below highlighted age as the most influential predictor by a wide margin, followed by avg_glucose_level, gender (both male and female), and smoking status. The distribution of SHAP values for age revealed consistent, high-magnitude contributions across many samples, underscoring its pivotal role in the model's decisions.

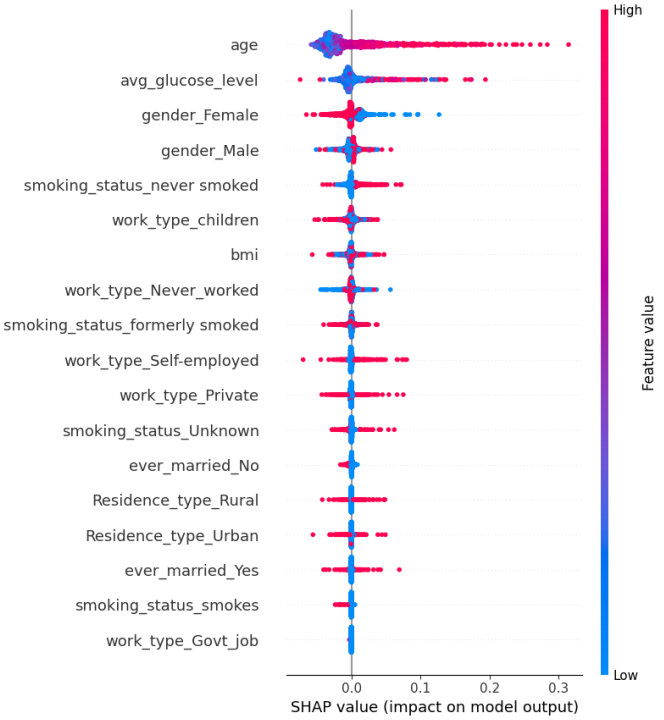


Fig. 19 RF SHAP Summary.

Delving deeper, the dependence plot below for age (Figure 20.) reveals a non-linear relationship with stroke risk. SHAP values remain close to zero for younger individuals (normalized age < 0) but rise sharply beyond approximately 1.0 (around 50 years old), indicating a substantial increase in stroke likelihood with advancing age. The colour gradient represents smoking status, where individuals who have

https://doi.org/10.31436/ijpcc.v12i1.636

never smoked (red) generally exhibit lower SHAP values at comparable ages. This pattern highlights how age and smoking behaviour jointly modulate stroke risk, aligning with established clinical evidence linking aging and lifestyle factors to elevated *cerebrovascular* vulnerability.

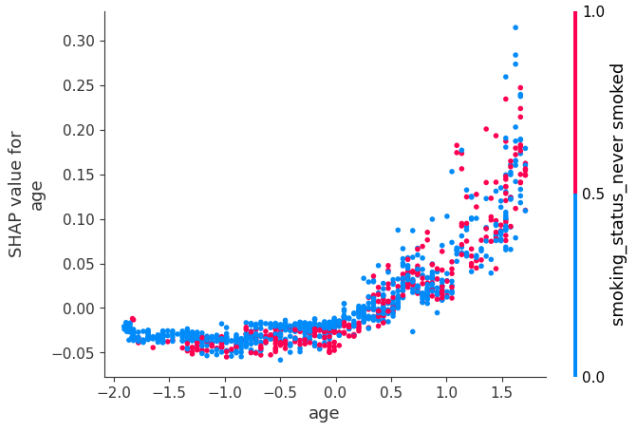


Fig. 20 RF SHAP dependence plot.

E. *Ensemble Model*

The ensemble, comprising XGBoost and LightGBM, classified as a black-box model was trained and evaluated using a standard machine learning pipeline that included class balancing via SMOTE. The model achieved relevantly a high performance on the test set, with an overall accuracy of 88.36%, a ROC AUC score of 0.8048, and a macro-averaged F1-score of 0.58. Notably, for the minority class (stroke cases), the model attained a precision of 0.19 and recall of 0.29, indicating moderate ability to detect stroke instances despite class imbalance challenges.

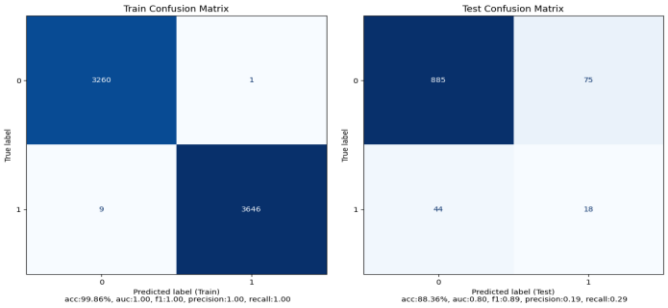


Fig. 21 Ensemble Confusion Matrices.

Figure 21. shows for both training and testing sets that illustrate the model's performance. While the training confusion matrix shows near-perfect classification (accuracy ~99.86%), the test matrix reveals some degradation in performance, particularly in sensitivity (recall) for the minority class. To mitigate potential overfitting, 5-fold cross-validation produced a mean AUC of 0.9965 ± 0.0043, reinforcing the reliability and consistency of the model's predictive performance across data splits.

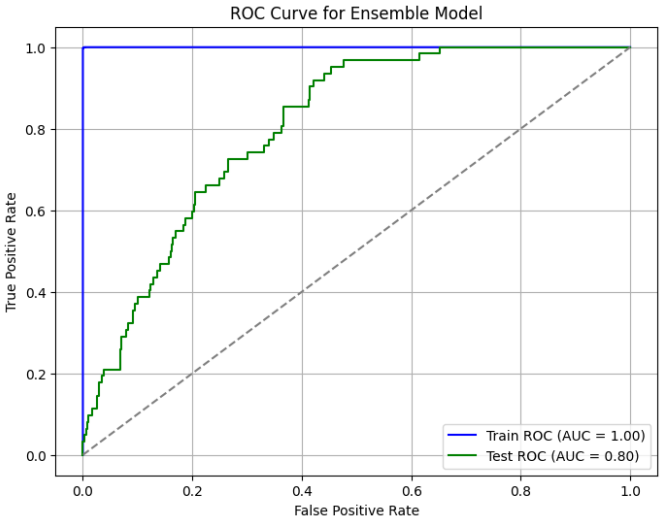


Fig. 22 Ensemble ROC Scores.

The ROC curves further confirm this. The training ROC curve exhibits an AUC of 1.00, indicating perfect separation of classes, while the test ROC curve achieves an AUC of 0.80, suggesting the model maintains good discriminative ability on unseen data but is less ideal than the overly optimistic training performance (see Figure 22).

While these metrics and visualizations (confusion matrix and ROC curve) provide useful quantitative assessments of model performance, they fall short in offering insights into why the model makes certain predictions, an essential factor for clinical decision-making.

To bridge this gap, we applied the SCI-XAI pipeline, employing SHAP (SHapley Additive exPlanations) to deliver both global and local interpretability.

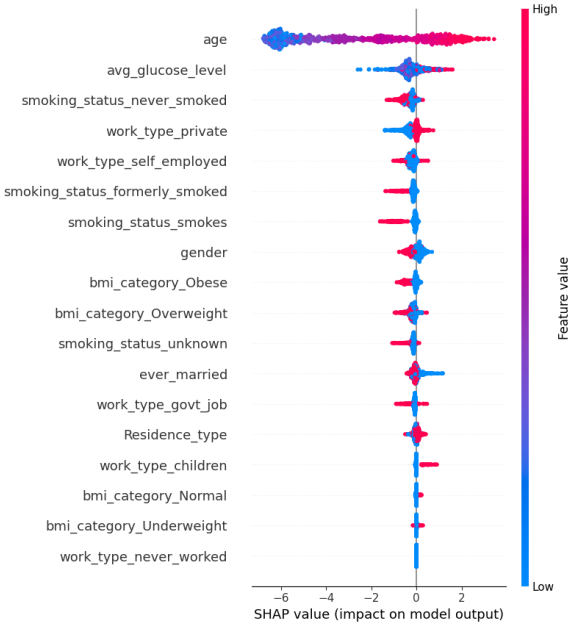


Fig. 23 Ensemble SHAP Summary.

https://doi.org/10.31436/ijpcc.v12i1.636

Figure. 23's summary plot identifies and ranks features based on their overall impact on the model's predictions. Here, age stands out as the most significant predictor, followed by average glucose level, smoking status, and various work type and BMI category variables. The color gradient (red to blue) visually maps each feature's value (e.g., high vs. low age), and its influence on pushing predictions toward stroke or non-stroke. This plot not only validates clinical knowledge (e.g., age and glucose level being critical risk factors) but also helps clinicians understand nuanced associations, such as how certain employment types or smoking history influence stroke likelihood. This global view is crucial for population-level insights and aligns the model's reasoning.

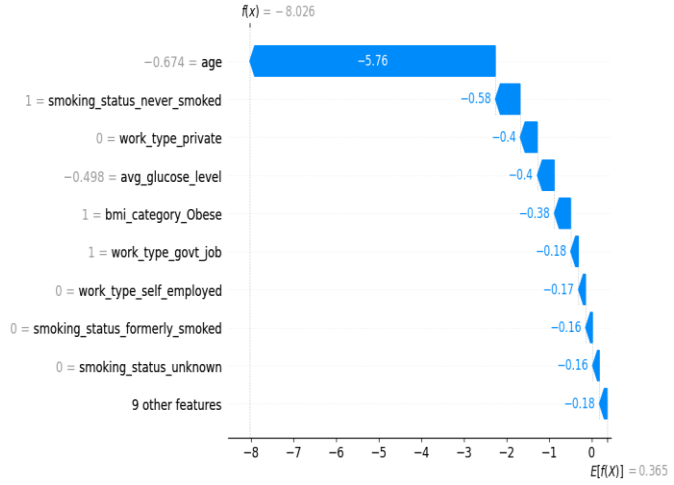


Fig. 24 Ensemble SHAP Waterfall force plot.

The waterfall plot dissects a single prediction to show how individual features cumulatively influence the model's output. For example, in one instance, age contributes a significant negative impact, decreasing the likelihood of stroke, while features like smoking status (never smoked) and BMI (obese) contribute to smaller, nuanced effects. This granular view allows clinicians to trace back the reasoning for specific patients, ensuring that individual predictions are transparent and defensible, a key requirement in high-stakes environments like healthcare (see Figure 24).

F. Multi-Layer Perceptron Model

The model comprised two hidden layers (128 and 64 neurons) with ReLU activation, dropout (0.3), and batch normalization to prevent overfitting. Using the Adam optimizer (learning rate = 0.001) and binary cross-entropy loss, it achieved 79.2% accuracy and an ROC-AUC of 0.78. While slightly less accurate than ensemble models, the MLP captured non-linear feature interactions effectively with stable generalization.

Fig. 25 Training Validation Accuracy & Loss

These graphs at Figure 25. illustrate the MLP's learning dynamics over 50 epochs. The training accuracy (left) steadily increased to around 95%, while validation accuracy plateaued near 80%, indicating the model learned well but began to generalize less effectively after ~20 epochs. Similarly, the training loss (right) consistently declined to about 0.12, whereas validation loss fluctuated between 0.5–0.7, suggesting mild overfitting, the model fits training data very well but shows higher error on unseen data. Overall, the network achieved stable performance but could benefit from stronger regularization or early stopping to improve generalization (see Figure 26).

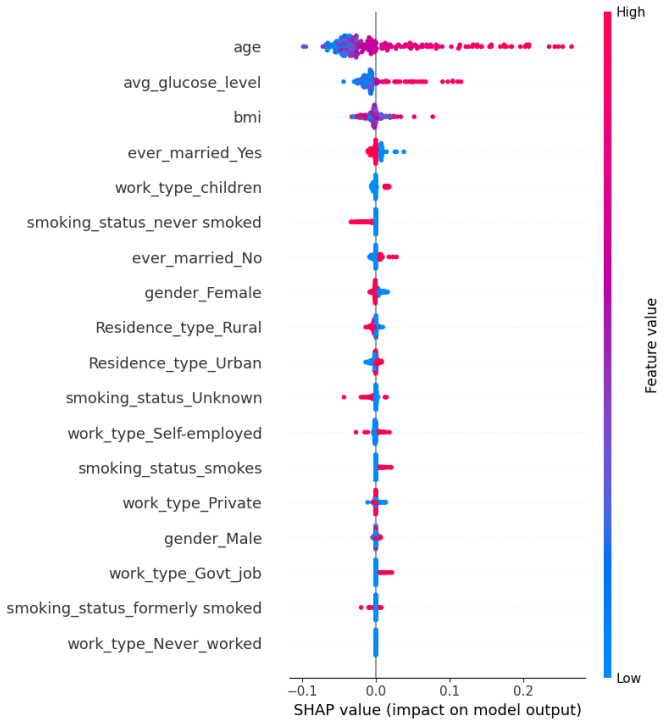


Fig. 26 MLP SHAP Summary Plot.

The summary plot demonstrates that age, average glucose level, and BMI were the most influential predictors in the MLP model, showing the highest positive SHAP values (approximately +0.25, +0.20, and +0.15, respectively). Higher values for these features significantly increased the model's

https://doi.org/10.31436/ijpcc.v12i1.636

predicted stroke risk, consistent with established clinical evidence. Moderate effects were observed for marital status and smoking-related variables, while demographic and occupational features contributed minimally, clustering near zero SHAP values. The colour gradient indicates that high feature values (red) generally pushed predictions toward higher stroke probability, whereas lower values (blue) reduced it. Overall, these findings highlight that metabolic and age-related factors predominantly drive the MLP model's decision-making process.

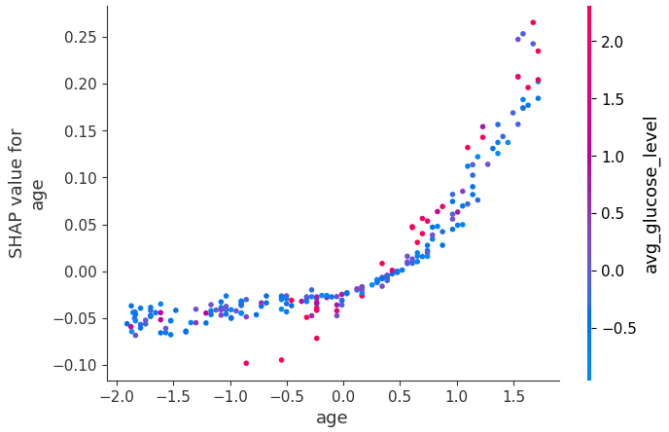


Fig. 27 MLP SHAP Dependence Plot.

Figure 27. plot reveals a strong, non-linear relationship between the feature age and the MLP model's output for stroke prediction. For low, normalized age values (e.g., from -2.0 to 0.0), the SHAP value is consistently negative, ranging from approximately -0.07 to 0.00, indicating that younger ages decrease the model's predicted probability of stroke. Conversely, as age increases beyond 0.0, the SHAP value rises exponentially, peaking at approximately +0.27 for the highest normalized age value of around 1.75. This demonstrates that increasing age is a major contributor to a higher predicted stroke probability. Furthermore, the plot highlights an interaction effect with avg_glucose_level (color-coded). At the highest age values (age > 1.0), the largest positive SHAP contributions (up to +0.27) are predominantly associated with the highest avg_glucose_level values (red, up to 2.0), suggesting that the positive impact of advanced age on stroke risk is amplified by a high average glucose level.

By integrating the SCI-XAI pipeline, we extend all the model's utility beyond raw predictive performance into the realm of Explainable AI (XAI). This is particularly valuable in medical contexts, where black-box predictions are often viewed with scepticism. The SHAP-based interpretability framework not only boosts interpretation confidence but also supports patient-specific consultations, potentially enabling caregivers to explain risk profiles in understandable terms.

IV. Discussion

Our results demonstrate that the SCI-XAI pipeline effectively balances predictive performance and interpretability in stroke prediction. The comparative framework particularly EBM and the SHAP-augmented ensemble achieved ROC-AUC and F1-scores closely matching those of black-box models, confirming that interpretability did not come at the cost of accuracy. This aligns with literature suggesting that for tabular data, advanced interpretable models like EBM can capture nonlinear patterns without compromising performance [26] [27]. For example, in prior work an EBM attained accuracy nearly on par with a random forest while remaining a transparent "glass-box" model [26] In our study, while the XGBoost + LightGBM ensemble achieved the highest metrics, EBM closely rivaled it, falling only marginally short. This finding reinforces the idea that interpretable models can deliver state-of-the-art performance while providing trans-parency an important insight for developing trustworthy AI systems [27] The key predictive fac-tors identified age, hypertension, and heart disease were consistent across EBM and SHAP anal-yses and align with established stroke risk factors reported in the literature (e.g., atrial fibrillation, blood pressure) [26] [27]. This consistency supports the validity of the models' reasoning and suggests that the SCI-XAI pipeline produces explanations that align well with domain knowledge.

A core aspect of the evaluation was comparing global and local interpretability. EBM, as an intrinsically interpretable model, offered clear global insights through its feature effect plots. For instance, stroke risk increased sharply with age beyond a certain threshold and rose consistently in patients with hypertension. Locally, EBM decomposed each prediction into additive contributions, showing precisely how individual features influenced outcomes and strengthened its white-box design [26].

By contrast, the ensemble (XGBoost + LightGBM) is a black-box model that cannot be directly interpreted [27]. Through SCI-XAI, we applied SHAP to this ensemble, enabling post-hoc inter-pretability. SHAP's global summaries highlighted top features such as age, glucose level, and hypertension closely mirroring EBM's insights. Locally, SHAP's force plots broke down individual predictions, showing how specific features (e.g., "Age = 75," "Hypertension = Yes") raised or lowered stroke risk compared to the baseline. The agreement between EBM's intrinsic explana-tions and SHAP's post-hoc outputs strengthens confidence in the reliability of these interpreta-tions.

https://doi.org/10.31436/ijpcc.v12i1.636

Additionally, traditional models like logistic regression and decision trees were examined. Logistic regression provided interpretable coefficients but was limited by its linear assumptions, making it less able to capture nonlinear risk effects. Decision trees were intuitive but, when deeper, became harder to interpret and underperformed compared to EBM and the ensemble. EBM im-proved upon both by modeling complex patterns transparently. Meanwhile, SHAP successfully added interpretability to the ensemble, demonstrating how post-hoc XAI tools can make high-performing black-box models more explainable [26].

While ensemble models such as Random Forest and XGBoost achieved high training perfor-mance, the observed discrepancy between training and testing metrics indicates mild overfitting. This is expected given the limited dataset size and high-class imbalance. Nonetheless, the applied mitigation strategies including tree depth limitation, regularization, and SMOTE-based resampling partially reduced this gap. Importantly, the study's focus is methodological rather than clinical deployment; thus, the results primarily demonstrate feasibility and interpretability rather than clinical readiness.

Overall, these findings highlight the value of combining inherently interpretable models with post-hoc explainability methods to achieve both accuracy and transparency. While this study does not develop a clinical tool, it offers a methodological framework that can inform future work aiming for transparent, interpretable AI in healthcare diagnostics.

V. Limitations & Future Work

While the results are promising, several limitations and avenues for future research must be noted. First, the ensemble model (XGBoost + LightGBM), though highly performant, is complex and risks overfitting particularly given modest sample size and high feature dimensionality. We mitigated this through cross-validation, depth control and hyperparameter tuning, but generalizability to new data cannot be guaranteed. Thus, future studies should validate these findings on larger, external datasets, ideally incorporating real-world clinical records to strengthen reliability. Simpler models (e.g., logistic regression, decision trees) were more robust to overfitting but showed lower predictive accuracy. Future work could explore more regularization techniques, simplified ensemble strategies, or hybrid models to retain performance while reducing complexity. Additionally, stroke was modeled as a binary classification task due to dataset constraints, which overlook the temporal progression of stroke risk. Future studies could incorporate longitudinal or time-to-event data to better capture real-world disease trajectories.

Second, the dataset was highly imbalanced, with stroke cases representing fewer than 5% of the total. While SMOTE oversampling and class-balanced training improved sensitivity, synthetic sampling may introduce artifacts and may not fully represent true population distributions. The dataset also reflects a specific demographic and regional context, which limits external validity. Future research should validate these models on broader, multi-center datasets and test alternative imbalance-handling techniques such as cost-sensitive learning or anomaly detection to enhance generalizability [26].

Third, feature limitations impacted the models. Key risk factors such as diet, exercise, and detailed cardiac history (e.g., atrial fibrillation) were missing, and some variables (like "heart disease") were overly broad. Incorporating richer clinical data including imaging, lab tests, and genetic information could improve both accuracy and interpretability. Additionally, investigating model calibration to ensure that predicted probabilities align with absolute stroke risk remains an important next step.

Lastly, while this study focused on technical interpretability using tools like SHAP, ELI5, and LIME, the assessment of interpretability remains *theoretical*. We did not conduct user studies to evaluate how healthcare professionals perceive and apply these explanations. Future work should include human-factors evaluations to measure usability, trust, and practical value, as well as explore ways to simplify outputs (e.g., streamlined SHAP visuals or natural-language summaries). Expanding interactive feedback loops where clinicians can flag unexpected results could also support continuous model refinement.

Overall, addressing these limitations will make future iterations of this framework more robust, generalizable, and useful as a benchmark for transparent AI in healthcare research. While clinical impact is a long-term goal, this work primarily contributes a methodological foundation for balancing predictive performance and interpretability in medical diagnostics.

VI. Conclusions

This study applied the SCI-XAI pipeline a structured framework combining advanced machine learning with Explainable AI (XAI) techniques to benchmark stroke risk prediction models with an emphasis on interpretability and performance. By integrating both white-box models, such as the Explainable Boosting Machine (EBM), and black-box ensembles (XGBoost + LightGBM) augmented with SHAP, LIME, and ELI5 explanations, we demonstrated that interpretable models could achieve predictive accuracy comparable to complex, opaque models while providing much clearer and more actionable insights.

Our comparative analysis highlighted that EBM, despite being inherently interpretable, closely rivaled the performance of black-box ensembles in terms of ROC-AUC and F1-score, reinforcing that high predictive power and transparency can coexist especially in tabular medical

datasets. SHAP and other post-hoc explainability tools effectively bridged the gap for black-box models, ensuring that even complex algorithms could be made more interpretable and suitable for actionable insights. Crucially, the key predictive features identified such as age, hypertension, and heart disease aligned with established clinical knowledge, lending further credibility to the out-puts and supporting the relevance of XAI in medical diagnostics. [26].

In conclusion, our study highlights the importance and feasibility of bringing interpretability to the forefront of white, gray-black-box models in healthcare. While the models presented are not intended for immediate clinical deployment, the structured application of the SCI-XAI framework provides a robust foundation for future research aiming to make AI in healthcare more trans-parent and reliable. This work contributes valuable insights into the trade-offs between model complexity, interpretability, and performance, offering a roadmap for researchers and practitioners interested in balancing these dimensions.

Declarations: This study did not involve any human participants, and all data was obtained from publicly available open-source datasets. The research was conducted using publicly accessible data in compliance with relevant guidelines and regulations. This work received no specific funding from any public, commercial, or not-for-profit sources.

Acknowledgment

The author expresses sincere gratitude to the Most Gracious and Most Merciful, for His guidance and blessings throughout this work. Deep appreciation is extended to Dr. Sharyar Wani for his invaluable supervision, insightful feedback, and continuous encouragement, which greatly shaped and strengthened this study. The author also thanks peers and colleagues for their constructive discussions and assistance during the research process, and family members for their unwavering support and motivation.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author(s) Contribution Statement

All authors contributed equally to this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study did not require ethical approval

References

[1]. H. O'Brien Quinn, M. Sedky, J. Francis, and M. Streeton, "Literature Review of Explainable Tabular Data Analysis," Electronics (Basel), vol. 13, no. 19, p. 3806, Sep. 2024, doi: 10.3390/electronics13193806.

[2]. A. M. Antoniadi et al., "Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review," Applied Sciences, vol. 11, no. 11, 2021, doi: 10.3390/app11115088.

[3]. P. A. Moreno-Sanchez, "An automated feature selection and classification pipeline to improve explainability of clinical prediction models," in Proceedings - 2021 IEEE 9th International Conference on Healthcare Informatics, ISCHI 2021, Institute of Electrical and Electronics Engineers Inc., Aug. 2021, pp. 527–534. doi: 10.1109/ICHI52183.2021.00100.

[4]. U. Pawar, D. O'shea, S. Rea, and R. O'reilly, "Incorporating Explainable Artificial Intelligence (XAI) to aid the Understanding of Machine Learning in the Healthcare Domain."

[5]. J. Ospel, N. Singh, A. Ganesh, and M. Goyal, "Sex and Gender Differences in Stroke and Their Practical Implications in Acute Care," Jan. 01, 2023, Korean Stroke Society. doi: 10.5853/jos.2022.04077.

[6]. K. M. Rexrode, T. E. Madsen, A. Y. X. Yu, C. Carcel, J. H. Lichtman, and E. C. Miller, "The Impact of Sex and Gender on Stroke," Circ Res, vol. 130, no. 4, pp. 512–528, Feb. 2022, doi: 10.1161/CIRCRESAHA.121.319915.

[7]. M. Wajngarten and G. Sampaio Silva, "Hypertension and stroke: Update on treatment," European Cardiology Review , vol. 14, no. 2, pp. 111–115, 2019, doi: 10.15420/ecr.2019.11.1.

[8]. W. Kim and E. J. Kim, "Heart failure as a risk factor for stroke," Jan. 01, 2018, Korean Stroke Society. doi: 10.5853/jos.2017.02810.

[9]. C. Zhu et al., "The association of marital/partner status with patient-reported health outcomes following acute myocardial infarction or stroke: Protocol for a systematic review and meta-analysis," Nov. 01, 2022, Public Library of Science. doi: 10.1371/journal.pone.0267771.

[10]. E. S. Eshak et al., "Changes in the Employment Status and Risk of Stroke and Stroke Types," Stroke, vol. 48, no. 5, pp. 1176–1182, May 2017, doi: 10.1161/STROKEAHA.117.016967.

[11]. O. Grimaud et al., "Stroke incidence and case fatality according to rural or urban residence results from the French Brest Stroke Registry," Stroke, vol. 50, no. 10, pp. 2661–2667, Oct. 2019, doi: 10.1161/STROKEAHA.118.024695.

[12]. X. Peng et al., "Longitudinal Average Glucose Levels and Variance and Risk of Stroke: A Chinese Cohort Study," Int J Hypertens, vol. 2020, 2020, doi: 10.1155/2020/8953058.

[13]. K. Miwa et al., "Clinical impact of body mass index on outcomes of ischemic and hemorrhagic strokes," International Journal of Stroke, Oct. 2024, doi: 10.1177/17474930241249370.

[14]. J. Chen et al., "Impact of Smoking Status on Stroke Recurrence," J Am Heart Assoc, vol. 8, no. 8, Apr. 2019, doi: 10.1161/JAHA.118.011696.

[15]. M. S. Islam, I. Hussain, M. M. Rahman, S. J. Park, and M. A. Hossain, "Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal," Sensors, vol. 22, no. 24, Dec. 2022, doi: 10.3390/s22249859.

[16]. A. Laios et al., "Factors Predicting Surgical Effort Using Explainable Artificial Intelligence in Advanced Stage Epithelial Ovarian Cancer," Cancers (Basel), vol. 14, no. 14, Jul. 2022, doi: 10.3390/cancers14143447.

[17]. S. K. Mandala, "XAI Renaissance: Redefining Interpretability in Medical Diagnostic Models," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.01668

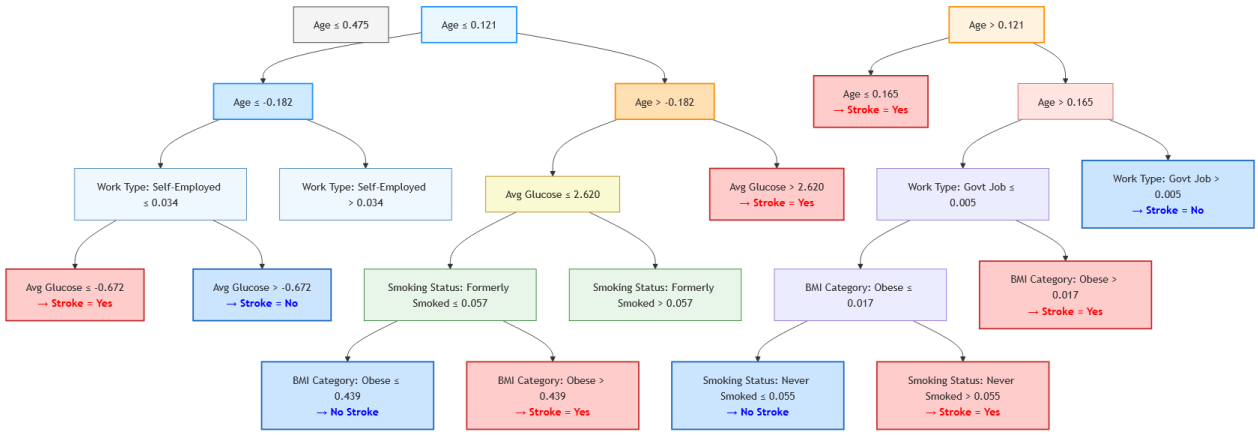
[18]. V. Petrauskas et al., "XAI-based Medical Decision Support System Model," International Journal of Scientific and Research Publications (IJSRP), vol. 10, no. 12, pp. 598–607, Dec. 2020, doi: 10.29322/ijsrp.10.12.2020.p10869.

[19]. T. A. J. Schoonderwoerd, W. Jorritsma, M. A. Neerincx, and K. van den Bosch, "Human-centered XAI: Developing design patterns for explanations of clinical decision support systems," International Journal of Human Computer Studies, vol. 154, Oct. 2021, doi: 10.1016/j.ijhcs.2021.102684.

[20]. J. Stodt, M. Madan, C. Reich, L. Filipovic, and T. Ilijas, "A Study on the Reliability of Visual XAI Methods for X-Ray Images," in Studies in

Health Technology and Informatics, IOS Press BV, Jun. 2023, pp. 32–35. doi: 10.3233/SHTI230416.
[21]. S. Alkhalaf et al., "Adaptive Aquila Optimizer with Explainable Artificial Intelligence-Enabled Cancer Diagnosis on Medical Imaging," Cancers (Basel), vol. 15, no. 5, Mar. 2023, doi: 10.3390/cancers15051492.
[22]. R. El Shawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in HealthCare A Comparative Study of Local Machine Learning Interpretability Techniques," in 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, Jun. 2019, pp. 275–280. doi: 10.1109/CBMS.2019.00065.
[23]. S. S, K. Chadaga, N. Sampathila, S. Prabhu, R. Chadaga, and S. K. S, "Multiple Explainable Approaches to Predict the Risk of Stroke Using Artificial Intelligence," Information, vol. 14, no. 8, p. 435, Aug. 2023, doi: 10.3390/info14080435.
[24]. B. S. D. Darshan et al., "Differential diagnosis of iron deficiency anemia from aplastic anemia using machine learning and explainable Artificial Intelligence utilizing blood attributes," Sci Rep, vol. 15, no. 1, p. 505, Jan. 2025, doi: 10.1038/s41598-024-84120-w.
[25]. B. Khokhar, V. Pentangelo, F. Palomba, and C. Gravino, "Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence." [Online]. Available: https://www.kaggle.com/datasets/
[26]. S. Lolak, J. Attia, G. J. McKay, and A. Thakkinstian, "Comparing Explainable Machine Learning Approaches With Traditional Statistical Methods for Evaluating Stroke Risk Models: Retrospective Cohort Study," JMIR Cardio, vol. 7, 2023, doi: 10.2196/47736.
[27]. S. Kruschel, N. Hambauer, S. Weinzierl, S. Zilker, M. Kraus, and P. Zschech, "Challenging the Performance-Interpretability Trade-Off: An Evaluation of Interpretable Machine Learning Models," Business and Information Systems Engineering, 2025, doi: 10.1007/s12599-024-00922-2.

APPENDIX 1: Simplified Decision Tree Visualization for Stroke Prediction



Visualization generated via Mermaid (https://mermaid.js.org/), adapted for academic presentation.

This appendix presents the ELI5-generated decision tree breakdown, corresponding to the interpretability discussion in the main text (referenced in Table 3 and Table 4). The visualization was reproduced using the Mermaid diagram-ming framework to illustrate the internal decision logic of the trained model. Each branch represents a feature-based split (e.g., Age, Average Glucose Level, BMI Category, Smoking Status, Work Type), providing a trans-parent, step-by-step rationale for the model's classification outcomes.

Color coding enhances interpretability: blue nodes indicate pathways leading to No Stroke predictions, while red nodes denote Stroke = Yes outcomes. Intermediate nodes shaded in orange or light blue correspond to decision points based on threshold values that guide the classification process.

Note: Numeric thresholds (e.g., Age ≤ 0.475) represent normalized feature values derived through Min–Max scaling, where all continuous features were transformed to a [0,1] range. These values reflect the relative percentile positions within the dataset (for example, Age ≤ 0.475 corresponds to individuals below approximately the 47.5th per-centile of the observed age range). This normalization ensures consistent feature comparison across different clinical measures and supports stable, interpretable model behavior.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.676

NutriMatch: AI – Driven Personalized Meal Recipes based on the Fresh Ingredients' Detection and User's Dietary Needs

[1]Siti Nur Raihannah Nazrul, [1]Nina Syahira Azman, [1]Noor Azura Zakaria*, [2]Suwandi, [3]Untung Rahardja

[1]Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[2]Faculty of Information Technology, Universitas Catur Insan Cendekia, Cirebon, West Java, Indonesia
[3]Faculty of Science and Technology, University of Raharja, Tangerang, Indonesia

*Corresponding author: azurazakaria@iium.edu.my
(Received: 9th December 2025; Accepted: 2nd January, 2026; Published on-line: 30th January, 2026)

Abstract— In modern fast-paced lifestyles, maintaining a healthy diet is challenging, contributing to the rise of diet-related diseases and food wastage, particularly fresh produce. Existing digital meal planning systems often lack the robustness to integrate practical ingredient detection with personalized dietary requirements effectively. To address these issues, this paper introduces NutriMatch, a web-based application designed to provide personalized healthy meal suggestions based on user submitted images of fresh fruits and vegetables. NutriMatch integrates an ingredient detection interface where users upload produce images via the browser, and the system returns the predicted ingredient label with a confidence score, for example, ginger with 90%, enabling users to verify detection before receiving tailored recipe recommendations. The system utilizes the MobileNetV2 architecture to classify 36 categories of fresh ingredients, chosen for its efficiency and suitability for web-based deployment. The platform is developed using the Laravel framework with PHP and MySQL for backend management, while the frontend utilizes React for a responsive user interface. Experimental results on the test dataset demonstrate that the model achieves a precision of roughly 92 percent and an F1-score of 0.89, validating the system's ability to facilitate sustainable eating habits and personalized nutrition through artificial intelligence model.

Keywords—Artificial Intelligence, Image Recognition, Personalized Nutrition, Web Application, MobileNetV2

I. INTRODUCTION

With a global rise in diet-related health problems such as obesity, diabetes, and malnutrition, the need for healthy eating is more critical than ever. Currently, individuals often lack detailed insights into utilizing fresh produce, leading to spoilage. Fruits and vegetables are essential for a balanced diet, yet their potential is often unrealized due to a lack of cooking skills or time constraints.

The primary problem motivating this project is the limitation of current recipe applications. While many apps aim to assist in meal planning, they often fail to effectively bridge the gap between abstract meal ideas, and the actual ingredients users possess at home. Furthermore, existing digital meal planning systems often lack the technological robustness to offer features like automated ingredient detection or seamless dietary personalization. NutriMatch aims to resolve these issues by offering a smart, AI-powered ecosystem that empowers users to maximize their fresh produce usage. By recommending personalized, health-aware recipes based on available ingredients and individual health goals, the system promotes both personal wellness and environmental sustainability.

This paper presents the design and development of NutriMatch, a web-based system that supports sustainable eating habits through personalized nutrition assistance. NutriMatch integrates AI-powered image recognition to detect food items from user-submitted images and combines this with an automated recipe recommendation module that filters and tailor's recipe suggestions to individual needs and preferences.

II. RELATED WORKS

A. *Automated Recognition and Classification of Fruit and Vegetables*

The development of deep learning models for identifying and categorizing produce has become essential for industrial and commercial applications, particularly in supermarkets and processing factories [1]. To address the variability found in real-world environments, such as lighting changes and background variations, researchers have proposed various high-precision frameworks for example Robust DCNN Frameworks. One study introduced a simple and efficient Deep Convolutional Neural Network (DCNN) designed to distinguish natural fruit images in difficult scenarios [2]. By utilizing a specialized database of 20 categories (comprising 10,000 images) from the Gilgit-Baltistan region of Pakistan, the model achieved a 96% recognition accuracy, demonstrating its readiness for global application requirements

https://doi.org/10.31436/ijpcc.v12i1.676

Research also works on the optimized CNNs and freshness detection. Utilizing Keras and TensorFlow, researchers built an optimized CNN architecture featuring five convolutional and pooling layers [1]. This model was specifically trained to identify fresh versus rotten produce across six classes, achieving a high accuracy of 96.88%. Furthermore, the model proved its versatility by maintaining a 94.35% accuracy when tested on the more extensive Fruits 360 dataset, which contains 131 categories

Beyond mere identification, research into YOLOv3 and optimized GoogLeNet has focused on maximizing efficiency for real-time industrial use [3]. By reducing the number of convolutional kernels and training parameters by 48%, researchers successfully tripled training speeds from 11.38 to 33.68 sheets per second. These technical refinements were particularly effective at distinguishing between produce items that are nearly identical in shape and color, such as lemons and oranges, which often pose significant challenges for traditional recognition systems.

B. Dietary Monitoring and Nutritional Health

Researchers developed MyDietCam, a mobile app for healthy Malaysian adults that integrates automated food recognition to ease the burden of manual dietary logging. It provides individualized recommendations and generates a diet quality score based on the Malaysian Healthy Eating Index [4].

A cross-sectional study of 2,509 adults used Principal Component Analysis (PCA) to identify five distinct eating behavior groups. The research found significant gender differences, noting that men consume more meat and engage in strength sports, while women adhere more to structured, vegetable-rich diets [5].

For Multi-food Detection on Plates, this research treats food analysis as a multi-class object detection problem because plates often contain multiple food items [6]. It evaluates models like YOLOv5 to accurately identify components of meals to support nutritional health tracking

C. Optimization of Deep Learning Architectures

Research by Sandler et al. [7] introduces MobileNetV2 architecture, a new mobile-tailored architecture using inverted residuals and linear bottlenecks. It significantly decreases the required memory and computational operations while maintaining high accuracy for object detection and semantic segmentation on resource-constrained devices.

One study compares YOLOv5 and EfficientDet across various food datasets for comparative performance analysis, concluding that YOLOv5 provides superior performance in both accuracy (mAP) and response time [6].

In terms of speed and efficiency Improvements, efforts to optimize GoogLeNet resulted in reducing the number of training parameters by 48%, which increased the training speed from 11.38 to 33.68 sheets per second [3].

D. Comparative Analysis with Existing Platform

A comprehensive review of existing platforms reveals a distinct gap in current market solutions, particularly regarding the integration of real-time image scanning with personalized recipe generation. We conducted a comparative analysis of NutriMatch against prominent existing systems such as MyFitnessPal [8], PlateJoy [9], and SuperCook [10].

MyFitnessPal is widely recognized for its robust calorie tracking capabilities; however, it relies heavily on manual data entry and lacks any form of ingredient image detection. Similarly, PlateJoy offers excellent personalized meal planning services but necessitates that users manually input their pantry inventory, which can be tedious and time-consuming. On the other hand, SuperCook succeeds in suggesting recipes based on listed ingredients but does not incorporate AI-driven health tracking or deep personalization for dietary restrictions. NutriMatch addresses these identified limitations by synthesizing the strengths of these platforms. It combines the convenience of automated ingredient detection with the utility of personalized recipe creation and nutritional tracking, offering a more holistic solution for sustainable nutrition management.

Table I summarizes the functional differences between these platforms, highlighting NutriMatch's unique position as the only system integrating real-time detection with health tracking. Across the four platforms, NutriMatch stands out as the most comprehensive: it supports real-time monitoring, recipe generation, and uniquely offers ingredient detection whereby this is a feature missing in MyFitnessPal, PlateJoy, and Supercook.

TABLE I
COMPARISON BETWEEN NUTRIMATCH AND OTHER WEB SYSTEM

Criteria	MyFitnessPal [8]	PlateJoy [9]	Supercook [10]	NutriMatch
Real-time Monitoring	Yes	No	No	Yes
Recipe Generation	No	Yes	Yes	Yes
Ingredient Detection	No	No	No	Yes
Calorie Tracking	High	Medium	Low	High
Food Waste Focus	Low	Low	High	High

https://doi.org/10.31436/ijpcc.v12i1.676

III. METHODOLOGY

The development of NutriMatch adheres to Agile Methodology. This iterative approach was selected due to the complex nature of integrating Artificial Intelligence with web development. The Agile framework allows for continuous refinement of the AI model and user interface through repeated cycles of planning, designing, developing, and testing.

A. System Design

The system architecture is designed to operate seamlessly within a web browser environment. The workflow begins when users interact with the frontend interface to upload photographs of their ingredients. These requests are transmitted to the backend server, which orchestrates communication between the database and the AI processing module.

Fig. 1 illustrates the core user interactions within the system, such as uploading ingredient images, viewing detection results, and generating personalized meal recommendations. This structure ensures that heavy computational tasks, such as image classification, are handled on server-side to maintain a smooth user experience.

As shown in Fig. 2, the NutriMatch system flow starts with user authentication to ensure secure access. After a successful login, the user uploads an ingredient image which is processed by the ingredient detection module. Once an ingredient is detected, the result is displayed to the user who may optionally add additional ingredients. The system then retrieves the user's dietary preferences and profile data to generate personalized recipe recommendations. Finally, the selected recipe contributes to the calorie tracking process, allowing users to monitor their daily nutritional intake before the process ends.

B. Implementation Stacks

The development employs a robust technology stack tensure scalability and high performance. The frontend is constructed using the React library, which allows for the creation of a dynamic and responsive user interface featuring a modern, green-themed aesthetic suitable for a health application. For the backend, the Laravel framework is utilized to manage server-side logic, including secure user authentication and database management. This separation of concerns ensures that the application remains maintainable and scalable as the user base grows.

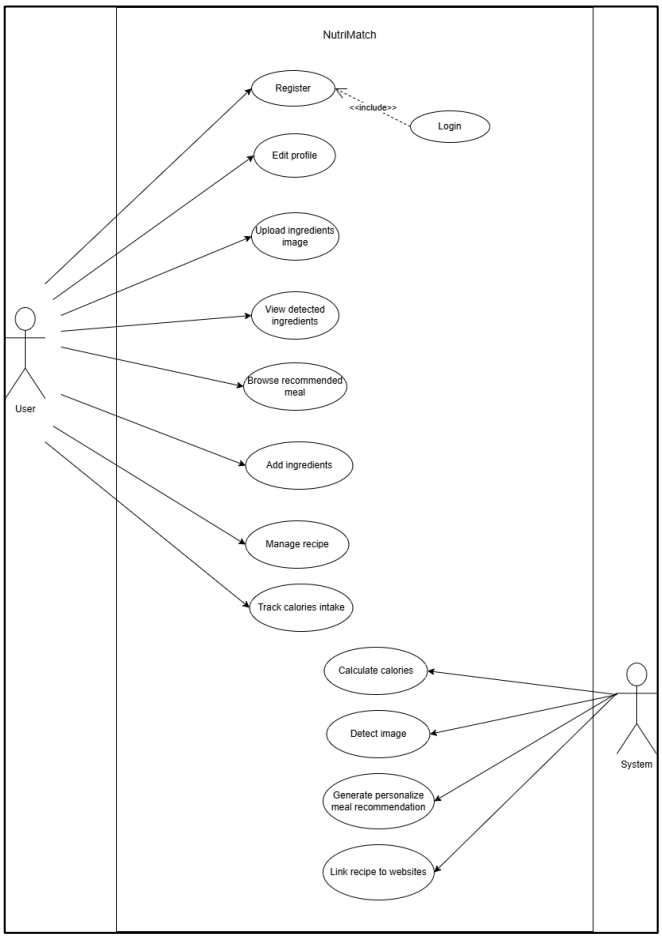


Fig. 1 Use Case Diagram

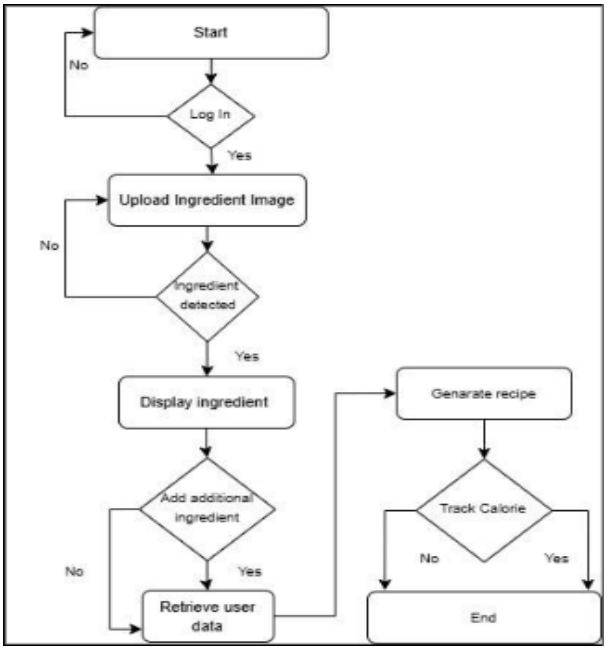


Fig. 2 NutriMatch FlowChart

https://doi.org/10.31436/ijpcc.v12i1.676

C. *Machine Learning Implementation*

The core intelligence of the NutriMatch system relies on a deep learning model designed to classify fresh produce from user-uploaded images. For this purpose, the MobileNetV2 architecture was selected due to its lightweight design and computational efficiency, making it suitable for web-based deployment where system resources may be limited.

The model was developed using the TensorFlow and Keras frameworks and trained on an image dataset consisting of 36 distinct categories of fruits and vegetables. The dataset was already organised into separate training, validation, and testing folders by the data provider. Specifically, each category contained 100 images for training, 10 images for validation, and 10 images for testing, resulting in an approximate split of 83% for training, and 8% each for validation and testing. This predefined dataset structure was used directly during model training and evaluation, without additional manual data splitting.

The development process employed Transfer Learning to expedite training and improve accuracy. A pretrained MobileNetV2 backbone, initialized with ImageNet weights, was utilized as the feature extractor. To adapt the model for NutriMatch, the top classification layers were removed and replaced with a global average pooling layer, followed by fully connected dense layers with ReLU activation, and a final Softmax output layer with 36 neurons corresponding to the target classes. During the training phase, the base layers of the MobileNetV2 model were frozen to retain the learned features from ImageNet, while only the custom classification layers were trained

Data preprocessing was a critical step in the pipeline. All input images were resized to a standard resolution of 224 × 224 pixels to match the architecture requirements. To prevent overfitting and enhance the model's ability to generalize unseen data, data augmentation techniques were applied, including random rotation, zooming, horizontal flipping, and shearing. The trained model was eventually saved in the .h5 format and integrated into the system via a FastAPI service, allowing for real-time inference requests from the main application.

IV. Results and System Prototypes

This section presents the fully developed prototype of NutriMatch and discusses the results of the system testing. The user interface was meticulously designed to be intuitive, ensuring that users can easily navigate from registration to recipe generation.

A. *Registration and Personalization Module*

The user journey begins at the Registration and Personalization interface. Unlike standard sign-up forms, NutriMatch requires users to create a detailed profile that captures essential personal metrics, dietary preferences (such as Vegan, Keto, or Paleo), and specific health goals. This data is critical as it feeds into the personalization engine. Fig. 3 illustrates the dietary preference selection screen, where users configure their specific restrictions and cuisine interests to tailor future recommendations.

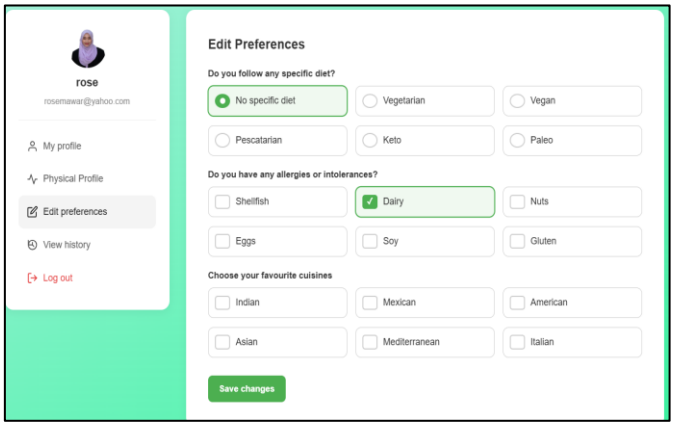


Fig. 3 User Profile based on preference

B. *Dashboard and Calorie Tracker*

Upon logging in, users are greeted by the Main Dashboard. This central hub features a visually engaging real-time calorie tracker, displayed as a dynamic progress ring. The dashboard is designed to provide a quick health overview briefly, encouraging users to stay within their nutritional goals. Fig. 4 shows the Calorie Tracker Ring that updates in real time as users log meals, visualizing the percentage of the daily calorie goal achieved. This provides users with instant feedback on their nutritional intake.

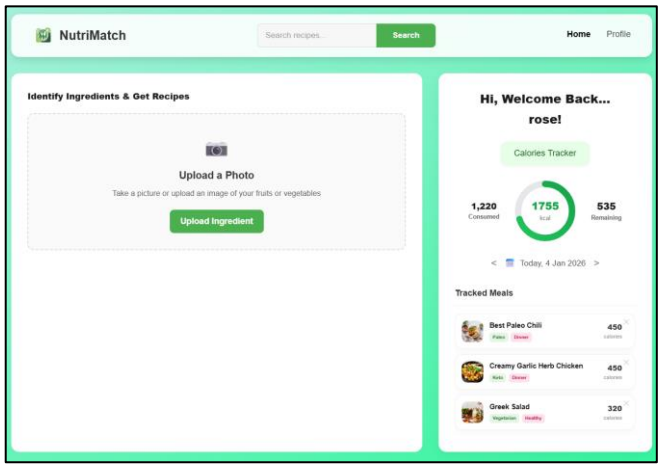


Fig. 4 Real-time calorie tracker ring on the dashboard

Fig. 5 shows the system generates a Random Recipe Recommendation list directly on the dashboard. This section offers immediate meal inspiration without requiring any user input, helping users decide what to cook quickly.

https://doi.org/10.31436/ijpcc.v12i1.676

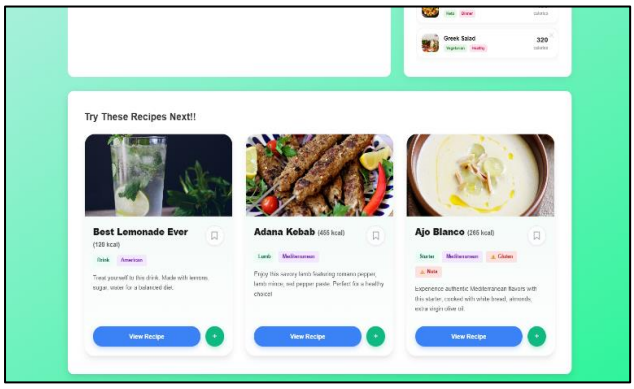


Fig. 5 Random Recipe Recommendation

C. *Ingredient Detection*

The core feature of the application is the Ingredient Detection Interface. In this module, users upload an image of their fresh produce directly through the browser. The system processes this image using the integrated MobileNetV2 model. Fig. 6 demonstrates the detection result interface, displaying the uploaded image alongside the AI-predicted ingredient label (e.g., Ginger) and a confidence score (e.g., 90%) to verify accuracy.

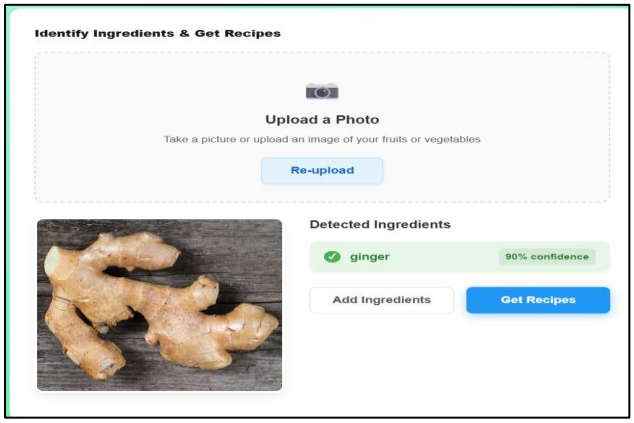


Fig. 6 detected ingredients after image upload.

Ingredient detection is a distinguishing capability of NutriMatch compared with existing web-based nutrition and recipe platforms. MyFitnessPal, PlateJoy, and Supercook do not provide ingredient detection from user images, meaning users must rely on manual input or pre-defined search to begin meal planning and tracking [8-10]. In contrast, NutriMatch supports browser-based image upload and automatic produce recognition, returning a predicted ingredient label (with a confidence score) that can be immediately used to drive recipe filtering and personalization.

D. *Recipe Recommendations*

Once the ingredients are confirmed, the system transitions to the Recipe Recommendation module. To ensure nutritional balance, the system provides a comprehensive workflow where, following ingredient detection, users are allowed to manually select additional carbohydrate or protein source. This step ensures that the generated recipes consist of a complete and balanced meal. Fig. 7 presents the selection of Carbohydrate and Protein the user can add before generating the recipe.

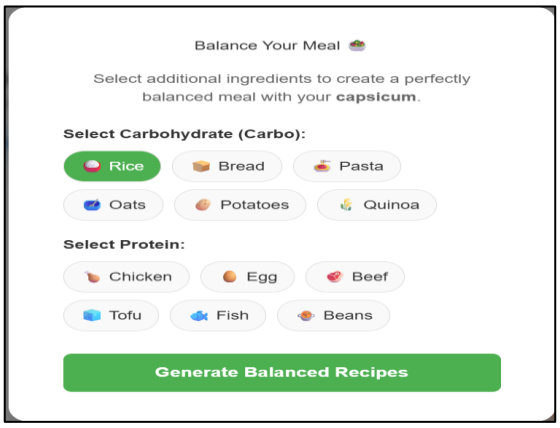


Fig. 7 Selection of Carbohydrate and Protein

Based on the detected produce, selection of additional nutrient and the user's pre-set dietary profiles, NutriMatch queries its database to generate a curated grid of recipe cards. Each card provides essential details, including the recipe title, total calorie count, and relevant cuisine tags. Fig. 8 presents the recipe grid, illustrating how users can view recommendations, bookmark recipes, or click the '+' button to instantly track the meal in their daily log.

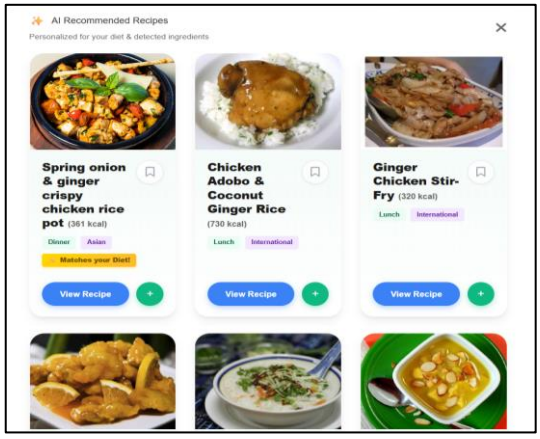


Fig. 8 Balanced recipe recommendations based on user ingredients.

E. *Testing and Evaluation*

The system underwent rigorous testing to validate its performance. The AI model achieved a high classification accuracy of 0.89 and an F1-score of 0.88 on the unseen test dataset, indicating reliable performance in distinguishing between different types of produce. Furthermore,

functional testing confirmed that all critical modules, including image uploading, API communication, and recipe filtering, are operated without errors.

The NutriMatch system's implementation of the MobileNetV2 architecture leverages inverted residuals and linear bottlenecks to optimize efficiency for web-based deployment, significantly reducing memory and computational operations while maintaining the high representational power required for ingredient classification [7]. These results are consistent with prior studies showing that deep learning models can remain robust under real-world variations (for example, lighting and background changes) in food-related recognition tasks [2].

Prior work has compared object detection models such as YOLOv5 and EfficientDet across food datasets, reporting that YOLOv5 can achieve better accuracy (mAP) and faster response time in those detection settings [6]. However, since NutriMatch focuses on lightweight ingredient classification for web-based deployment, this study adopts MobileNetV2 to balance accuracy and computational efficiency.

Furthermore, NutriMatch's focus on personalized meal recipes aligns with the identified need for tailored interventions based on gender-specific dietary patterns, such as the distinct food preferences and routine structures revealed through Principal Component Analysis (PCA) [5]. Finally, the successful functional testing of NutriMatch's API and image upload modules mirrors the development process of applications like MyDietCam, where rigorous lab and pilot testing are essential to prevent the crashing and lagging issues that frequently lead users to abandon digital dietary monitoring tools.

V. Conclusion

Future improvements for NutriMatch include extending the detection model to be able recognise multiple ingredients within a single image and upgrading the Spoonacular API plan to support higher request limits and richer nutritional data. A chatbot feature may be added to enhance user interaction, while conversion to a mobile application would improve accessibility. Additional enhancements include religion-based dietary filtering, integration of generativ AI for more intelligent recipe recommendations, and the ability to estimate ingredient freshness and quantity for better serving suggestions

NutriMatch demonstrates the effective use of artificial intelligence and web technologies to support healthier eating through automated ingredient detection and personalized recipe recommendations. The system reduces the effort required for meal planning, particularly for users with busy lifestyles. Despite limitations such as API restrictions and model capability, the system achieves its objectives and shows strong potential for future expansion into a smart nutrition assistant.

Acknowledgment

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

Authors Contribution Statement

All authors contributed equally to this work.

Data Availability Statement

There is no external or third-party data that support the findings of this study.

Ethics Statement

This study did not require ethical approval

References

[1] A. Kushwaha, "Fruit classification using optimized CNN," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 2023: IEEE, pp. 1-5. doi.org/10.1109/icicat57735.2023.10263596

[2] D. Hussain, I. Hussain, M. Ismail, A. Alabrah, S. S. Ullah, and H. M. Alaghbari, "A Simple and Efficient Deep Learning-Based Framework for Automatic Fruit Recognition," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 6538117, 2022. doi.org/10.1155/2022/6538117

[3] F. Yuesheng *et al.*, "Circular fruit and vegetable classification based on optimized GoogLeNet," *IEEE Access*, vol. 9, pp. 113599-113611, 2021. doi.org/10.1109/access.2021.3105112

[4] N. A. Kong, F. M. Moy, S. H. Ong, G. A. Tahir, and C. K. Loo, "MyDietCam: development and usability study of a food recognition integrated dietary monitoring smartphone application," *Digital Health*, vol. 9, p. 20552076221149320, 2023. doi.org/10.1177/20552076221149

[5] A. Feraco *et al.*, "Gender differences in dietary patterns and physical activity: an insight with principal component analysis (PCA)," *Journal of translational medicine*, vol. 22, no. 1, p. 1112, 2024. doi.org/10.1186/s12967-024-05965-3

[6] R. Morales, J. Quispe, and E. Aguilar, "Exploring multi-food detection using deep learning-based algorithms," in *2023 IEEE 13th International Conference on Pattern Recognition Systems (ICPRS)*, 2023: IEEE, pp. 1-7. doi.org/10.1109/icprs58416.2023.10179037

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520. A simple and efficient Deep Learning-Based framework for automatic fruit recognition. Computational Intelligence and Neuroscience, 2022, 1–8. doi.org/10.1109/cvpr.2018.00474

[8] MyFitnessPal. "myfitnesspal." MyFitnessPal, Inc. https://www.myfitnesspal.com/ (accessed 1 January, 2026).

[9] PlateJoy. "PLATEJOY." https://support.platejoy.com/platejoy-faqs (accessed 5th February, 2025).

[10] Supercook. "Supercook." https://www.supercook.com/#/desktop (accessed 5th February 2025).

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.677

AI-Powered Resume Crafting and Screening

[1]Syed Muhammad Afiq Idid Syed Azli Idid, [1]Syasya Syaerill, [1]Noor Azura Zakaria*, [2]Marsani Asfi, [3]Qurotul Aini

[1]Department of Computer Science, International Islamic University Malaysia (IIUM), Gombak, Malaysia
[2]Faculty of Information Technology, Universitas Catur Insan Cendekia, Cirebon, West Java, Indonesia
[3]Department of Digital Business, University of Raharja, Tangerang, Indonesia

*Corresponding author azurazakaria@iium.edu.my
(Received: 9th December 2025; Accepted: 2nd January 2026; Published on-line: 30th January 2026)

Abstract— In today's competitive job market, a resume often serves as the first point of contact between job seekers and recruiters. However, many job seekers, especially fresh graduates, struggle to craft a professional and ATS-friendly resume that clearly highlights their skills and experiences. At the same time, recruiters face the challenge of screening large volumes of applications, which is time-consuming and may result in qualified candidates being overlooked. To address these issues, this project develops Resume Pro, a web-based system that integrates AI-powered resume crafting and automated resume screening. The platform enables users to generate high-quality, ATS-friendly resumes with AI-driven suggestions, while recruiters can screen and rank applicants using natural language processing and machine learning techniques. The system is implemented using Python (Flask) for the backend and HTML, CSS, and JavaScript for the frontend. The delivered system is a user-friendly application that supports better resume preparation and improves efficiency and accuracy in the hiring process.

Keywords— AI Resume crafting, Resume Screening, Machine Learning, Natural Language Processing, ATS-Friendly Resume, Web-Based System, Python, AI suggestion

I. Introduction

Since we are in the digital era now, a lot of companies have started using the Applicant Tracking System (ATS) to help filter resumes which poses a new challenge for the candidates to create a good resume that will stand out and pass through these automated systems [1-4]. In addition, recruiters also need an efficient way to screen the candidates and pick the best fit candidates for the job.

Both job seekers and recruiters face a lot of challenges during the hiring process especially in today's competitive job market. Job seekers, especially fresh graduates, often struggle to make a resume that stands out mostly because they do not know what the recruiters are looking for. Because of that, many job seekers ended up using the most basic resume templates that did not highlight their abilities, strength or match it with the roles that they are applying for. Other than that, recruiters also face problems when reviewing the resumes because it takes a lot of time for them to manually review them.

This process slows down the entire hiring process and also raises the operating expenses. Furthermore, human screening tends to be inconsistent and biased which may cause qualified candidates to be unfairly passed over [4]. The process will also be tedious with the absence of intelligent tools to help match the candidate's skill with the job requirements.

This paper investigates current system requirements for resume crafting and candidate screening in environments increasingly influenced by Applicant Tracking Systems (ATS). Building on the identified needs and limitations of existing approaches, the study designs and develops an AI-powered platform that supports two primary user groups: job seekers, who can generate and refine personalized resumes, and recruiters, who can screen job applications more efficiently to identify candidates that best match the job requirements. The platform is subsequently tested to confirm core functionality and usability, ensuring the proposed solution is practical and reliable for real-world adoption.

II. Related Works

In ATS-influenced hiring, resume crafting becomes a machine-readability and matching problem as much as a writing problem such as candidates are pushed to use parse-friendly layouts and align terminology to the job description so their skills are not missed during automated parsing and scoring. The job-seeker side is strongest when a platform explicitly supports ATS-friendly resume building (for example, an AI-assisted builder) but can still be limited by template-only customization, which may not fit diverse roles and experiences. At the same time, research shows that resume–job matching can be improved by using embedding-based representations to better capture semantic similarity and reflect human evaluation preferences across domains [1].

https://doi.org/10.31436/ijpcc.v12i1.677

From the recruiter side, screening is mainly a scale and consistency challenge. Systems typically parse resume text, extract features or skills, and apply ranking methods to shortlist candidates efficiently. One example in the literature uses text parsing, cosine similarity and KNN-based ranking to match CVs with job descriptions at volume, illustrating how pipeline-style screening is commonly operationalized [5]. Platforms oriented toward employers prioritize filtering or ranking and ATS-like workflows rather than helping candidates craft resumes. For instance, Hiredly's employer offerings describe AI-enabled application screening and resume summarisation and a built-in applicant tracking system, reflecting the emphasis on recruiter-side efficiency [6]. However, these automated approaches also raise governance concerns—bias and validity can depend heavily on data choices, prediction targets, and ongoing monitoring, and applicants' trust can differ depending on whether algorithms are used for resume screening [7].

Based on the feature comparison in Table 1, existing platforms tend to support either resume creation or employer-side screening, but rarely both in an integrated workflow. Info-Tech focuses mainly on job seekers, offering an AI-assisted resume builder and ATS-friendly output, however, its resume customization is limited to preset templates, and it does not provide screening tools for recruiters. In contrast, Hiredly and JobsLah place stronger emphasis on the employer or recruitment side by providing screening-related capabilities. However, their job-seeker features differ whereby JobsLah includes a built-in resume builder, while Hiredly does not offer an integrated resume builder or resume customization features for job seekers. This split creates a gap for users who need end-to-end support from crafting an ATS-ready resume to being evaluated fairly and efficiently through screening.

TABLE I
COMPARISON WITH EXISTING SYSTEM

FEATURES	INFO-TECH [8]	HIREDLY [6]	JOBSLAH [9]
Resume Builder	Yes, AI assisted	No, built-in resume builder	Yes, built-in resume builder
Resume Customization	Limited, only preset templates	Not available	Not available
ATS Compatibility (Builder)	ATS friendly	Not available	Not available
Resume Screening (Employer)	Not available	Yes, it uses smart filters to rank the applicants	Yes, it has ATS module with keyword-based filtering
AI/Keyword Filtering	Not for recruiters	For employers, matches profiles to job posts	For employers, it filters resumes by keywords
Target Users	Job Seekers	Job Seekers and Employers	Employers (HR teams only)
Strength	Easy resume building for job seekers	Fast and smart job matching for recruiters	End to end screening and recruitment for HR
Weakness	No tools for employers to screen resume	No resume builder	Template resume-driven

III. METHODOLOGY

The development approach selected for this project is the Agile Software Development Life Cycle (SDLC). Agile is an iterative and adaptive methodology that breaks development into smaller, manageable cycles, allowing features to be designed, implemented, tested, and refined progressively. This approach provides the flexibility needed to incorporate feedback, enhance functionality, and improve user experience throughout the project, making it well suited for a system that may require continuous adjustments and incremental improvements [10].

A. *Requirements Gathering*

The Agile software development life cycle begins with requirements gathering, where the project team identifies and defines the user needs and expectations for the system. In this phase, we analyse and compare existing resume crafting and screening platforms to understand current practices and common limitations. The missing or weak features identified from this comparison are then translated into system requirements, ensuring that the proposed platform addresses key gaps in current solutions and supports both job seekers and recruiters effectively.

B. *System Design*

Figure 1 illustrates the use case diagram for our system namely Resume Pro, which involves three main actors which are Job Seeker, Recruiter, and Admin. Both job seekers and recruiters can register and log in to the system, and they are able to upload job descriptions. A Forgot Password function is also provided to support account recovery when users cannot access their credentials. For job seekers, the system supports profile management, selection of resume templates, and resume editing, including AI-generated suggestions to improve content. Job seekers can then

https://doi.org/10.31436/ijpcc.v12i1.677

download the completed resume. Recruiters, on the other hand, can import resumes and view or compare screening results to support candidate selection. Finally, the admin is responsible for managing user accounts and monitoring system usage to ensure smooth operation and oversight.

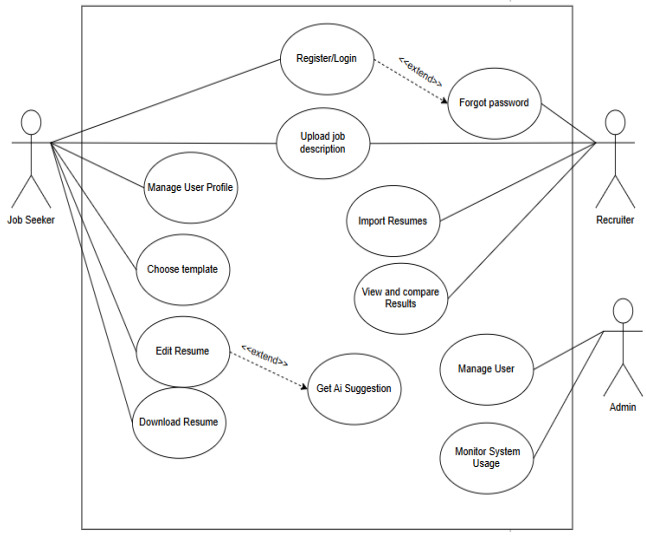


Fig. 1. Resume Pro Use Case diagram

C. Development

This project consists of two core modules which are resume crafting and resume screening. The web application is developed with HTML, CSS, and JavaScript for a user-friendly frontend interface, while Python (Flask) is used to implement backend services and API endpoints. MySQL serves as the database for storing user information, resume content, uploaded job descriptions, and screening results. To support AI-driven resume recommendations and automated screening, we explored suitable machine learning approaches based on resume–job matching tasks.

We initially searched for a dataset containing resume attributes such as education, skills, and work experience. The dataset was cleaned using Pandas by removing irrelevant columns, handling missing values, and combining relevant textual features into a unified text_data field. However, due to limited dataset availability and constraints for model training, the system relies on pre-trained transformer models from Hugging Face.

Two models were implemented for different stages of evaluation which are BERT, used as the primary feature extractor and classification baseline, and MiniLM-L6-v2, used as the ranking engine for resume–job description matching. BERT was fine-tuned using the skills and objective_career fields, first for general feature representation and then for binary suitability classification using a 0.7 match threshold to provide rapid initial screening. MiniLM-L6-v2 processes the job description and resume as a paired input to generate a compatibility score (0–100), enabling the system to rank candidates based on contextual similarity rather than simple keyword frequency. Following model integration, the frontend, backend, and database components are combined to complete the platform, and comprehensive testing and evaluation are performed to assess usability and real-world system performance.

D. Testing

This section describes the testing conducted on the web application to ensure it is ready for use and meets all specified requirements and functionalities. The evaluation includes unit testing, integration testing, system testing, and User Acceptance Testing (UAT) to verify both individual components and overall system performance.

The testing activity was carried out across three test phases and showed consistent success for all key features. Core workflows including user registration, login, job description submission, skill recommendation, resume PDF generation, and resume screening upload passed in every phase. In addition, validation and control cases such as invalid login attempts, adding skills via the skill button, and role-based redirection also passed throughout, indicating stable functionality and reliable system behaviour across repeated testing cycles.

TABLE II
TESTING ACTIVITY

Test Case Description	Test Phase 1	Test Phase 2	Test Phase 3
User Registration	✔	✔	✔
Login	✔	✔	✔
Submit Job Description	✔	✔	✔
Get Skill Recommendation	✔	✔	✔
Generate Resume PDF	✔	✔	✔
Resume Screening Upload	✔	✔	✔
Invalid Login Attempt	✔	✔	✔
Skill Button Add	✔	✔	✔
Role-Based Redirect	✔	✔	✔

E. Deployment

In this section, we document the development, implementation, and evaluation of Resume Pro, a web-based application that integrates AI for resume crafting and screening. The system is deployed using the Flask web framework, supported by MySQL for structured data management, and Hugging Face Transformers for the AI components. The application also underwent functional testing and User Acceptance Testing (UAT) to ensure that it

https://doi.org/10.31436/ijpcc.v12i1.677

operates reliably, produces accurate outputs, and supports timely interactions in real usage.

The frontend was developed using HTML, CSS, and JavaScript to deliver a clean and professional user interface. A consistent blue-and-white theme was applied across the application to maintain a uniform look and feel. Responsive design principles were implemented to ensure accessibility across different screen sizes. In addition, custom JavaScript was used to support file uploads and to display AI-generated skill recommendations dynamically without requiring page refreshes.

For the backend, Flask was used to provide a lightweight and efficient server environment, while MySQL manages stored data such as processed resume content, user credentials, and job descriptions. For the skill recommendation feature, a fine-tuned DistilBERT model was implemented for Named Entity Recognition (NER) using a custom-labelled dataset and the BIO tagging format to identify skills from raw text. For candidate ranking, the system uses the MiniLM-L6-v2 Sentence Transformer to perform semantic matching by encoding documents into 384-dimensional embeddings and computing cosine similarity, allowing candidates to be ranked based on contextual relevance rather than keyword frequency alone.

System integration connects the frontend interface, Flask backend, database, and both AI models into a unified pipeline. For resume screening, when a recruiter uploads a resume, the frontend sends the file to the backend, which triggers the MiniLM-L6-v2 model to compute a match score against the active job description. The screening results are stored in MySQL and returned to the frontend to update the recruiter's leaderboard. For resume crafting, when a user submits a job description, the fine-tuned DistilBERT NER model extracts explicit technical and soft skills from the text. The extracted skills are then used to query a skill map, enabling the system to recommend related skills that commonly co-occur with the identified ones.

IV. System Prototype

This section presents the user interface of Resume Pro and highlights its key features.

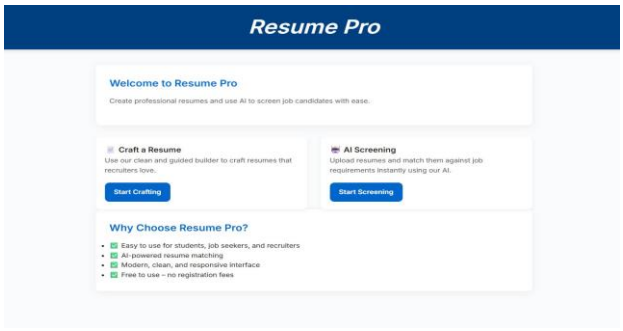


Fig. 2. Main Page

Fig. 2 shows the main page of Resume Pro, where users can choose between two primary functions which are resume crafting or resume screening.

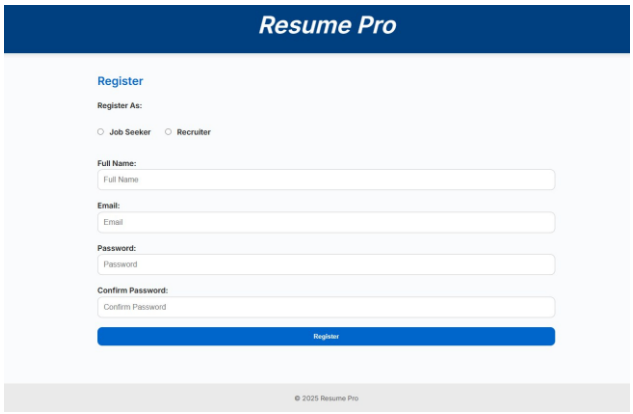


Fig. 3. Registration Page

Fig 3. shows the registration page for Resume Pro, where users create an account and select their role as either a job seeker or a recruiter.

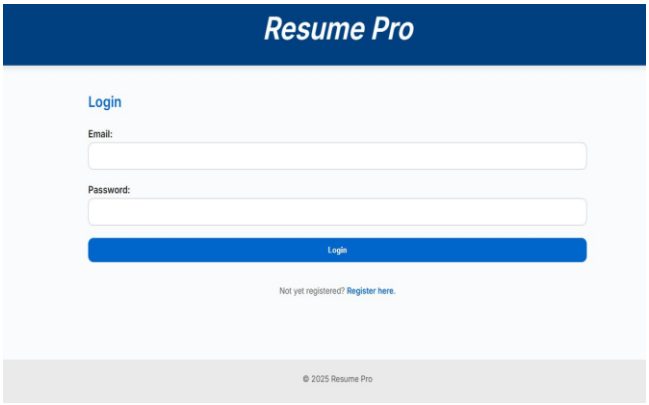


Fig. 4. Login page

Fig. 4 shows the login page for Resume Pro, where users enter their email and password to access the system.

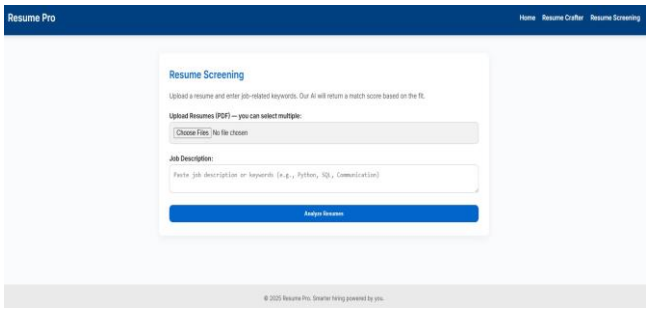


Fig. 5. Resume Screening Page

Fig 5. presents the resume screening page, where recruiters can upload multiple resumes at once and input a job description to initiate the screening process.

https://doi.org/10.31436/ijpcc.v12i1.677

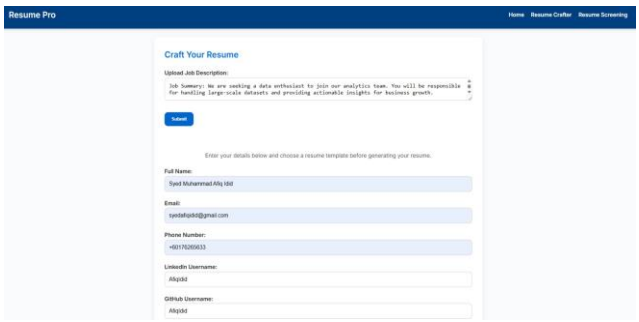


Fig. 6. Crafting Resume

Fig 6. Shows the crafting resume page for Resume Pro. This is where the user can upload job descriptions and submit them and input their personal details.

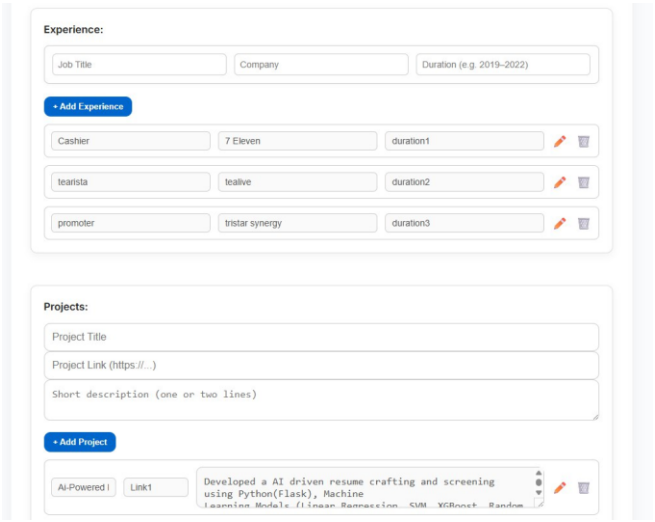


Fig. 7. Crafting Resume

Fig. 7 continues the resume crafting page, where users can enter details of their work experience and projects to further complete and refine their resume. Fig 8. Is also a continuation of the resume crafting page.

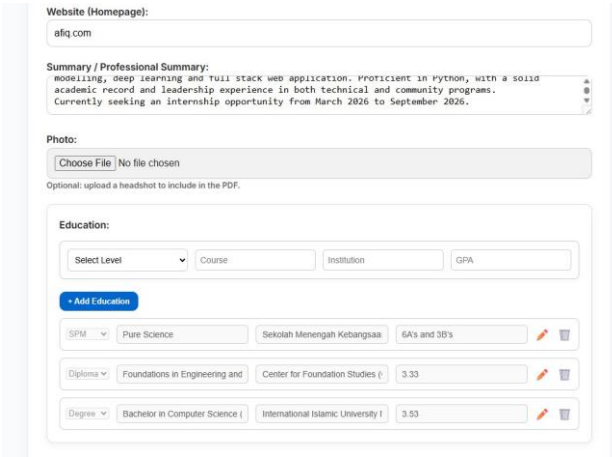


Fig. 8. Crafting Resume

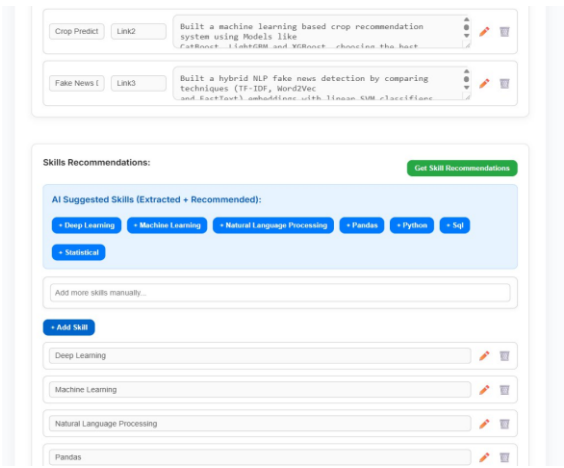


Fig. 9. Crafting Resume

Fig 9. shows the section where users can add their skills and receive AI-generated skill recommendations.

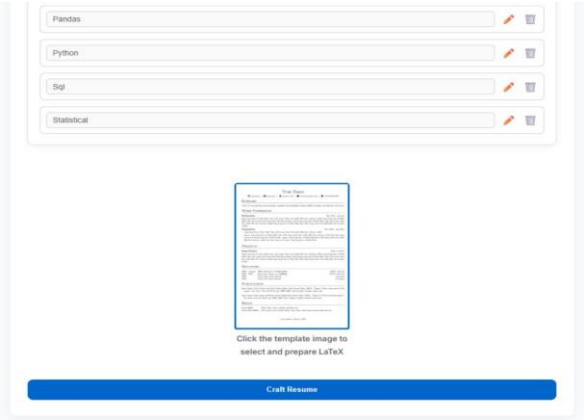


Fig. 10. Crafting Resume

Fig. 10 illustrates the Craft Resume function, where users can generate and download their resume in PDF format.

Syed Muhammad Afiq Idid

Github: Afiqidid | Linkedin: Afiqidid | syedafiqidid@gmail.com | +60176265633

SUMMARY

Computer Science (Data Science) undergraduate with practical experience in machine learning, data analytics and AI-based development. Strong interest in applying datasets and intelligent solutions to real world problems, supported by hands-on work in NLP, predictive modelling, deep learning and full stack web application. Proficient in Python, with a solid academic record and leadership experience in both technical and community programs. Currently seeking an internship opportunity from March 2026 to September 2026.

WORK EXPERIENCE

Cashier 2023-2023
7 Eleven
Promoter 2021-2021
Emmer Zeena
Tearista 2022-2022
Tealive

PROJECTS

Ai-Powered Resume Crafting And Screening (Final Year Project) Link
Developed a AI driven resume crafting and screening using Python(Flask), Machine Learning Models (Linear Regression, SVM, XGBoost, Random Forest) and Natural Language Processing techniques (TF-IDF) supported by a complete frontend development using HTML,CSS and JavaScript and a MySQL database for all data's such as user profiles, resume data, job description, resume template.

Crop Prediction Based On Environmental Factors Using Machine Learning Algorithm Link
Built a machine learning based crop recommendation system using Models like CatBoost, LightGBM and XGBoost, choosing the best accuracy and deployed it on Streamlit for real time predictions based on the environmental inputs.

EDUCATION

SPM Pure Sciencee at Sekolah Menengah Kebangsaan Putrajaya Presint 16(1) (GPA: 6A's and 3B's)
Diploma Foundations in Engineering and Computer Science at Center for Foundation Studies (CFS), IIUM, Gambang (GPA: 3.33)
Degree Bachelor in Computer Science (Data Scientist) at International Islamic University Malaysia (IIUM), Gombak (GPA: 3.53)

SKILLS

python, java, sql, php, github, canva

Last updated: January 2, 2026

Fig. 11. Generated Resume

Fig 11. shows the resume that has been generated by Resume Pro in pdf format.

V. Conclusions

In conclusion, this project successfully developed Resume Pro, an AI-powered web application for resume crafting and candidate screening that addresses practical challenges in resume preparation and job–candidate matching. The system met its main objectives by delivering a unified platform that combines automation with artificial intelligence to support both job seekers and recruiters. Although several challenges were encountered during development, particularly in integrating and deploying the AI components, these issues were resolved and the final system operates as intended. Nevertheless, there remains scope for further enhancement to improve functionality, usability, and overall performance in future work. Future developments to improve Resume Pro web-based application include: Adding more resume templates to provide users with greater design and formatting options. Introducing a user profile dashboard that enables users to manage and update personal information more efficiently. Fine-tuning both AI models using larger and more diverse datasets across multiple sectors such as healthcare, legal, and arts to improve generalization. Strengthening security features, including password reset and password change functionality.

Acknowledgment

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

Authors Contribution Statement

All authors contributed equally to this work.

Data Availability Statement

There is no external or third-party data that support the findings of this study.

Ethics Statement

This study did not require ethical approval

References

[1] R. V. K. Bevara *et al.*, "Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching," *Electronics*, vol. 14, no. 4, p. 794, 2025. doi.org/10.20944/preprints202501.1707.v1

[2] M. Velankar and P. Khuspure, "A Study of Applicant Tracking System (ATS) In Minimizing Human Intervention in Recruitment," *International Journal of Innovative Research in Engineering and Management (IJIREM)*, vol. 12, no. 6, pp. 98-101, 2025. doi.org/10.55524/ijirem.2025.12.6.17

[3] L. Dražeta, "Applicant Tracking System: A Powerful Recruiters' Tool," in *Sinteza 2024-International Scientific Conference on Information Technology, Computer Science, and Data Science*, 2024: Singidunum University, pp. 240-245. doi.org/10.15308/sinteza-2024-240-245

[4] S. Dogra, S. Vasesi, A. Mittal, V. Jain, D. Academics, and S. Chaudhary, "Smart Resume Screening and Matching System," presented at the NCAIDT 2025 - National Conference on AI, IoT & Data-Driven Transformation, Sonipat, India, 2025. doi.org/10.63169/ncaidt2025.p11

[5] K. Tejaswini, V. Umadevi, and M. K. Shashank, "Design and development of machine learning based resume ranking system," *Global Transitions Proceedings*, vol. 3, no. 2, pp. 371-375, 2022. doi.org/10.1016/j.gltp.2021.10.002

[6] Hiredly. "Hiredly Product." https://hub.hiredly.com/products-page (accessed 20 March 2025).

[7] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 469-481. doi.org/10.1145/3603195.3603203

[8] Info-Tech. "Applicant Tracking System (ATS)." Info-Tech Systems Integrators. https://www.info-tech.com.my/applicant-tracking-system (accessed 29 December 2025).

[9] Jobslah. "Jobslah Malaysia." https://www.jobslah.com/my (accessed December 29, 2025).

[10] GeeksforGeeks. "Agile SDLC – Software Development Life Cycle." https://www.geeksforgeeks.org/software-engineering/agile-sdlc-software-development-life-cycle/ (accessed 29 December, 2025).

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.679

Optimizing Load Balancing Framework for a Distributed Local Network

Ubaid Ajaz[1], Zainab S. Attarbashi[1], Sara Babiker Omer Elagib [2], Aisha Hassan Abdalla Hashim[2]

[1] Department of Computer Science, International Islamic University of Malaysia.
[2] Kulliyyah of Engineering, International Islamic University of Malaysia

*Corresponding author zainab_senan@iium.edu.my
(Received: 9th December 2025; Accepted: 2nd January 2026; Published on-line: 30th January 2026)

Abstract— Load Balancing is a critical and foundational challenge in systems and network performance, especially in resource-constrained infrastructure environments. In which it requires careful alignment between infrastructure limited resources and performance requirements. This paper presents a lightweight deployment of a locally hosted web server on a small local network using off-the shelf devices. The observations of this paper indicate effective distribution of traffic evolving through different deployment stages. One node setup was implemented to be a baseline for performance comparison. And a 2-nodes setup was built using NGINX to provide the required load balancing. Both implementations were tested using load testing tools: Locust and Siege. Results were then compared based on standardized performance metrics: scalability, response time, throughput, and server load. The 2-nodes implementation showed near-linear scalability, with doubled throughput and CPU load dropped to 45%.

Keywords— *Load Balancing, Resource Constrained, Local Network, Algorithms, Performance Metrics*

I. Introduction

Local network infrastructure plays a crucial role in establishing a scalable environment for a web server to ensure high availability and performance, especially if there are limited resources. A web server or application specifically deployed for the purpose of cybersecurity applications and awareness demands a network infrastructure that can handle high volumes of traffic and requests. This is where the concept of load balancing requires immediate attention and involvement. Cloudflare simply defines load balancing as the practice of distributing computational workloads between two or more computers. Load balancing ensures even distribution, maximizes resource allocation, and provides fault tolerance, preventing any single component of the network from becoming a bottleneck [1], [2].

Currently, there is a lot of existing work and research on load balancing techniques and cloud computing in general. These studies focus on the comparison of existing load balancing algorithms alone or discuss how load balancing works in a cloud infrastructure in a broader sense [3], [4]. However, there is a lack of studies focusing specifically on the deployment of a load balancing setup in a small-scale environment offering cost-effective solutions. While setting up a locally hosted environment, acquiring high-performance servers can prove to be a significant financial challenge. Existing lab equipment lacks the required processing power and memory capacity to support and handle a fairly large number of users accessing the servers for simultaneous use. Thus, implementing an optimized local network with such off-the-shelf devices using load balancing would provide a cost-effective alternative to buying expensive servers and proprietary software [5]. This will enhance the computational capacity of the whole network to satisfy the needs of the system.

For local small-scale infrastructures, in a university lab for example, with only a few servers in a local network, users may face overload when multiple simultaneous connections from users are there. In such scenarios, this can lead to service slowdowns and failures. Effective load balance can enhance applications performance in these situations and ensure smoother user experience and improved traffic control.

The rise of lightweight deployment frameworks such as containerization with Docker and open-source load balancers has made it feasible to implement robust load balancing without enterprise-grade hardware [6], [7]. Technologies like Docker allow services to be containerized and run on commodity hardware, while software-based load balancers can efficiently distribute traffic at low cost. Example of open-source software load balancers are NGINX [8] and HAProxy [9]. They are both popular for Linux and they support load balancing in layer 4 and 7. Other emerging and popular software load balancers are Traefik [10] and Envoy [11]. They offer more flexibility of configurations and availability on integrations with container-based platforms. [12]

The aim of this study here is to investigate some working load balancing technologies that can apply to small-scale local networks, and work towards adapting such technologies for the enhanced distribution of resources over a limited system. The very goal is to design and implement a lightweight load balancing mechanism able to manage the changing loads in an efficient manner in a basic lab setup. The research seeks to answer how such load balancing can be optimized using freely available tools like HAProxy, what components are best suited for modest infrastructure, and which metrics—such as throughput, fault tolerance, and memory usage can effectively evaluate performance [13].

This current accessibility of technology and the open-source community enable small organizations to achieve reliability close to that provided by large data centers, but it also raises research questions. How well do standard load balancing algorithms perform in a resource-constrained local network? What trade-offs arise when using tools like HAProxy and NGINX on minimal hardware? These questions are significant for academia and small enterprises aiming to optimize performance without significant investment [14]. This paper answers the questions set forth by analyzing the implementation of a locally hosted CipherQuest, a cybersecurity training and Capture the Flag (CTF) event platform. The system is evaluated through three different evolutionary phases: from the entry-level single-node server to the more complex dual-node cluster with rudimentary splitting of traffic, and finally to the robust dual-node arrangement with a dedicated HAProxy load balancer. Using performance parameters, such as response time, throughput, CPU load, and so on, the results were contrasted among different traffic management schemes. The aim is to identify how well traffic could be distributed in a small local network and which algorithms therefore provide the best performance under that constraint. The paper's outcome sheds light on the viability of setting up load balancers in the limited environment and offers some points to consider for similar deployments in the future.

The reminder of this paper is organized as follows. Section II presents research methodology and load balancer design and implementation. Section III illustrates experimentation results, and results discussion and performance evaluation in section IV. Paper ends with a conclusion in section V.

II. METHODOLOGY

This study adopts mixed-methods experimental research design, combining quantitative performance metrics (e.g., response time, throughput, CPU usage) with qualitative system behavior observations (e.g., failover response, load distribution consistency). The goal is to investigate how lightweight load balancing mechanisms perform in a small-scale, resource-constrained local network and to evaluate their efficiency under simulated traffic.

A. Environmental Setup

The experimental testbed for this paper comprises a distributed local network consisting of two nodes. These nodes are off-the-rack computers running on 8GB RAM each connected via ethernet provided by the university LAN (1000 MBps). The operating system chosen was Linux with the Ubuntu LTS distribution, given its optimization for server tasks and lightweight deployment, catering to the need for managing web applications and eliminating licensing costs compared to a Windows Server. This setup reflects a resource-constrained environment typical of a small lab or classroom setup. The application deployed is an open-source Capture-The-flag web platform used for cybersecurity training exercises. This application serves as a representative workload for this research that generates typical web traffic (multiple concurrent users, database queries, dynamic content). To ensure portability and ease of management, containerization surrounds the full application stack using Docker containers. These lightweight virtualized units that packed the whole application, dependencies, database and load balancer. This abstracts the differences in the underlying OS environments and provides a suitable platform for testing different node setups and load balancing algorithms with simulated traffic to gather results in terms of performance metrics which will be later discussed in the paper.

To scrape and gather performance metrics of the servers, the infrastructure was implemented and tested in two phases:

- *Stage 1: Single-Node Deployment (Baseline):* In the initial setup, only one PC was used to host the CTFd container. A NGINX web server was running as a reverse proxy on the same host to forward HTTP requests to the container (and serve any static assets). NGINX was configured to listen on the host's IP and route all traffic to the local CTFd application. This stage had no load balancing since only one node served all requests. It established baseline measurements for how the application performs on a single server under load and these measurements are then used to do a comparative analysis of performance metrics with a load balancing strategy introduced later.

https://doi.org/10.31436/ijpcc.v12i1.679

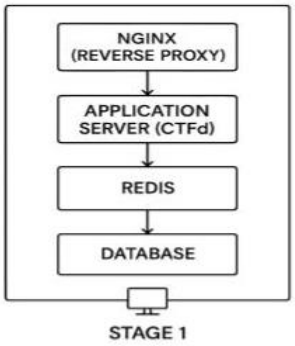


Fig. 1 Single Node Deployment Architecture

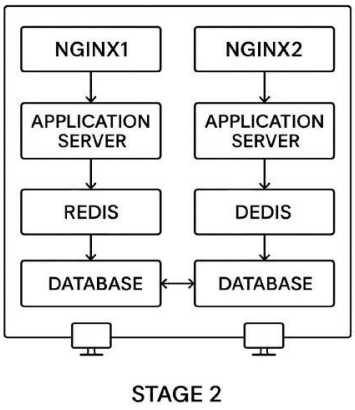


Fig. 2 Dual Node Deployment Architecture

- *Stage 2: Dual-Node with Load Balancing:* In this stage, a second PC was added to create the distributed setup comprising two application servers independently running Dockerized instances of CTFd and synchronizing the database. Each machine was configured with its own NGINX reverse proxy for local handling of incoming web traffic from containers. NGINX was the main load balancer in this scenario: on each of the two servers, NGINX was statically configured with an upstream block listing the CTFd instances of both the local and remote server with different load balancing algorithms, depending on which scenario was being tested. This decentralized style of balancing ensured that no single node would become a bottleneck while also avoiding the need for a central load balancer. Although these strategies were limited, lacking features such as health checks or adaptive request routing, they offered a lightweight and functional method for traffic distribution. This stage demonstrated the feasibility of achieving basic load balancing using reverse proxy infrastructure, providing early performance improvements without additional overhead.

B. Load Generation Tools

For this experimental setup, we used load generation tools to simulate clients and generate load on the system. We used three different industry standard tools: Apache Jmeter, Siege and Locust. Siege, an HTTP load testing and benchmarking utility tool is a command-line-based tool that triggers a preset number of concurrent users (threads) hitting an appropriate URL or set of URLs and consequently reports the elapsed time, response time, and throughput (transactions per second). Siege was used for fast stress tests, such as bombardment of the CTFd home page or challenge endpoints at 50, 100, 200 concurrent hits to analyze pure throughput and server behavior. Locust is a much more flexible load testing framework in which users exist within Python codes. Locust lets you define more complex user behaviors in Python code that can scale to simulate millions of users. We wrote simple Locust scenarios to simulate a typical user session in CTFd: logging in, fetching the scoreboard, and opening a challenge page. Locust gave us further stats (response time distribution, response time percentiles) and allowed us to gradually ramp up users. This tool was particularly useful in observing the system's behavior over time under sustained load and for checking if an algorithm causes queue build-up on one server or not. These two tools, while giving us a myriad of information, gave us insight into the systems from a high-level perspective of user experience and a low-level perspective of request throughput.

C. Monitoring and Metrics

To gather fine-grained performance data from the running system, Prometheus and Grafana were set up as the monitoring stack. Prometheus is the open-source monitoring solution that gathers the metrics data and stores it in a time series database. Prometheus was configured to scrape the metrics from each component: the Linux system metrics on each of the nodes (CPU, memory, network usage) via Node Exporter, Docker metrics including CPU usage per container for specificity and custom metrics from application if there were any (limited to infrastructure metrics, as CTFd being a flask app did not natively expose Prometheus metrics). The traffic simulation tools come with their own statistics which were gathered manually and through saved logs.

Grafana was then employed to visualize these metrics via dashboards. It is an open-source analytics and interactive visualization web application. Custom dashboards for key performance indicators were set up, such as CPU usage of a node, memory usage, load average, network traffic, requests handled by a server, response time graphs over a test, which allow us to cross-check the monitoring with the Siege/Locust results. If Siege reports a slowdown at 200

https://doi.org/10.31436/ijpcc.v12i1.679

users, for instance, that monitoring can tell whether the CPU of one node maxed out or if memory or networking became a bottleneck. Monitoring also ensured that our test environment was fully operational (e.g., if one node went down, we would see it in the metrics immediately and could investigate).

The performance metrics mentioned below serve to evaluate system behavior underload and the efficiency of traffic distribution, and the overall resource consumption by nodes. Each metric therefore exposes one or the other facet of locally distributed network performance through different stages.

1. Latency or Response Time

Measured as the end-to-end time taken for a request to be processed from the client to the server and back. This includes network latency, processing time on the server, and any delays introduced by the load balancing mechanism. Measurement was made for both median response times and for the 90th percentile not only to capture the average performance but also the performance in peak load conditions.

2. *Throughput*

The number of requests successfully processed by the system per unit time, typically measured in requests per second (RPS). Throughput provides an indication of the system's capacity to handle high volumes of traffic.

3. Server Load Balancing

Balanced load distribution among the servers forms the heart of all load balancing strategies. Ideally, two servers with exactly similar capabilities would share the whole workload exactly equally.

4. CPU and Memory Utilization

Monitoring CPU and memory usage on each node provides a low-level view of how efficiently resources are being used. CPU utilization trends can reveal whether traffic is being spread effectively, while discrepancies in CPU or memory load between nodes may suggest imbalances in request handling. Memory usage is also tracked to ensure containerized services remain stable under stress, though in lightweight deployments like CTFd, memory is typically a secondary concern unless large-scale concurrency is involved.

D. Testing Procedure

For each stage of deployment (and for each load balancing algorithm in Stage 2,3), we conducted a series of tests to measure performance:

Baseline single-node test: Using Siege, we bombarded the single-node setup with incremental loads: a varied number of concurrent users (each user continuously requesting a mix of pages). We recorded the average response time, throughput (requests per second), and observed the system resource usage. This established the capacity of one server and provided critical baseline data for later comparison with distributed configurations.

Dual-node manual split test: We repeated similar load tests on the Stage 2 setup. When using NGINX's upstream approach, we examined NGINX logs to confirm that roughly half the requests went to each node. The performance metrics here illustrate the benefit of having two servers. During this stage, multiple NGINX-supported algorithms were tested under identical conditions to analyze their behavior in a dual-node setup. Specifically, round-robin (default), least-connections (which routes new requests to the less busy server), and IP-hash (which maps requests to a server based on client IP) were configured and evaluated. The same load testing tools—Locust, Siege, and Apache JMeter—were used to simulate user traffic.

III. Results

This section visualizes the load testing metrics collected during Stage 1, where a single-node architecture was deployed using one PC hosting all core services (CTFd application, database, Redis, and NGINX). Two types of load tests were conducted using Locust and Siege.

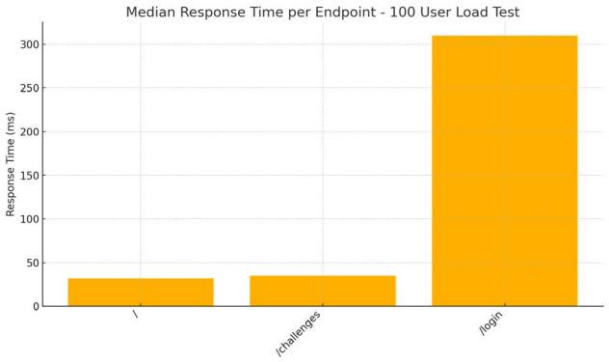


Fig. 3 Locust Results with heavy load testing (100 users)

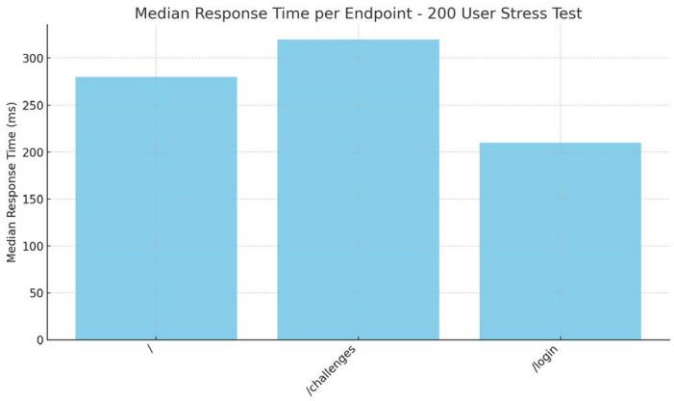


Fig. 4. Locust Results with stress load testing (200 users)

https://doi.org/10.31436/ijpcc.v12i1.679

TABLE I
SIEGE RESULTS WITH HEAVY LOAD TESTING

Metric	Value Result
Test Duration	5 minutes
Concurrent Users	255 (user cap hit)
Transactions	60,837
Success Rate	100% (0 failures)
Avg. Response Time	1.24 seconds
Transaction Rate	203.28 requests/sec
Throughput	38.08 MB/sec

TABLE II
SEIGE RESULTS STAGE 2

Metric	Value Result
Test Duration	5 minutes
Concurrent Users	255 (user cap hit)
Transactions	79,320
Success Rate	100% (0 failures)
Avg. Response Time	0.72 seconds
Transaction Rate	264.4 requests/sec
Throughput	49.61 MB/sec

This section visualizes the load testing metrics collected during Stage 2, where a two-node load-balanced architecture was deployed. In this setup, core services such as the CTFd application, database, Redis, and NGINX were distributed across two PCs, with NGINX acting as a reverse proxy to balance incoming traffic between the nodes. The same load testing tools, Locust and Siege, were used to evaluate performance under identical conditions as Stage 1.

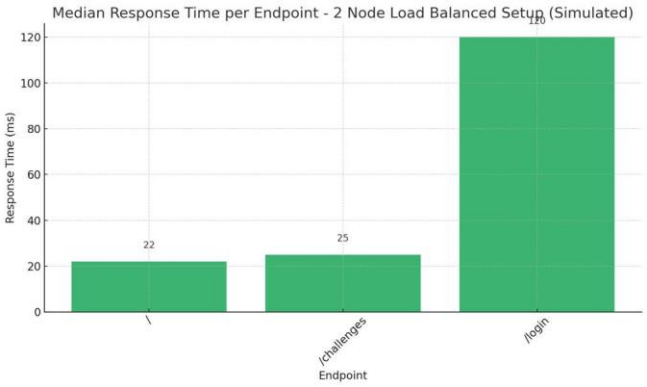


Fig. 5. Locust Results with heavy load testing (100 users)

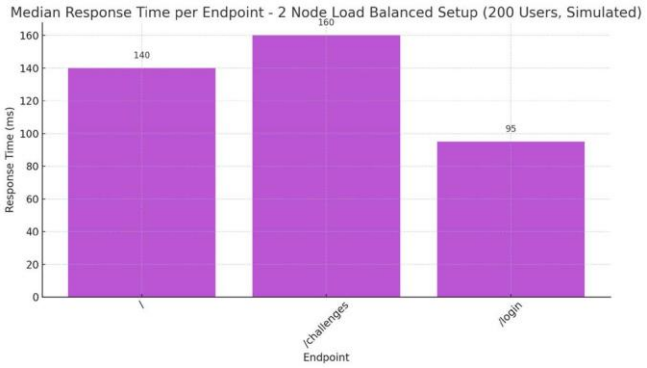


Fig. 6. Locust Results with stress load testing (200 users, increased ramp up)

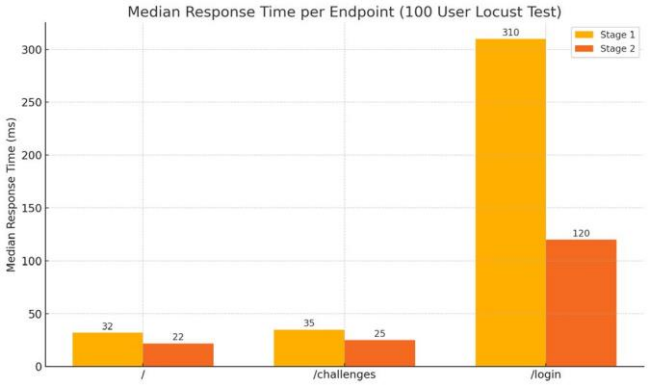


Fig.7 Stage 1 and Stage 2 Comparison for heavy load test

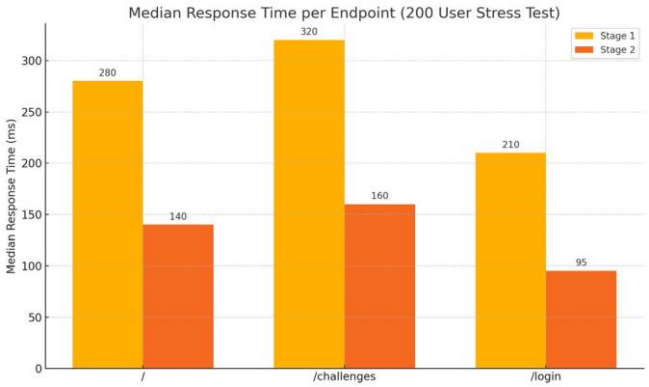


Fig. 8 Stage 1 and stage 2 comparison of stress load test

IV. DISCUSSION

The findings of this study clearly demonstrate the performance benefits of introducing a two-node distributed architecture in a local network environment using off-the-shelf machines and free software tools. Through both controlled (100-user) and stress-level (200-user) load testing, the system showed substantial improvements in throughput, response time, and load distribution when transitioned from a single-node to a dual-node setup with load balancing via NGINX.

The single-node deployment, although sufficient for moderate traffic, quickly reached saturation under heavier loads, leading to CPU utilization nearing 90% and median response times exceeding 300 ms for critical endpoints like /login. In contrast, the two-node setup reduced peak CPU loads per server to approximately 45–50%, effectively halving the processing burden on each machine. Median response times dropped significantly by over 50% in some endpoints and overall throughput more than doubled, showcasing linear or near-linear scalability within the constraints of the testbed. Despite these gains, several limitations emerged. The project operated under strict hardware constraints, with only two physical machines available for deployment. This prevented the implementation of a more advanced Stage 3 architecture, which would have introduced a dedicated load balancer node capable of supporting adaptive algorithms such as least-connections, IP-hash, or weighted round-robin. These strategies could not be fully explored, as true comparative analysis of load balancing algorithms typically requires a larger server pool to distribute traffic across diverse backend conditions. A more intelligent, feedback-aware load balancer could mitigate such issues, but its implementation was beyond the resource scope of this project.

V. CONCLUSION

In conclusion, this work validates the hypothesis that a lightweight, locally-hosted web server cluster with NGINX load balancing significantly enhances performance in resource-constrained environments. The two-node deployment achieved near-linear scalability, doubling throughput while reducing individual server load. Future work should explore scaling beyond two nodes, incorporating dynamic load balancing algorithms, and monitoring with more granular observability tools to refine traffic distribution decisions to provide a practical, low-cost model for small-scale environments such as university labs, classrooms, or lightweight training platforms.

ACKNOWLEDGMENT

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

There is no external or third-party data that support the findings of this study.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] N. Arora, P. Saha, and S. Sinha, "A review on load balancing algorithms in cloud environment," *Int. J. Sci. Technol. Res.*, vol. 10, no. 1, pp. 142–148, 2021.

[2] R. Tripathi, D. Dutta, and S. Sanyal, "Load balancing for resource allocation in cloud computing using live migration of virtual machines," *Procedia Comput. Sci.*, vol. 167, pp. 116–124, 2020.

[3] A. T. Akinwale and K. S. Adewole, "Performance evaluation of load balancing algorithms in cloud computing," *Int. J. Comput. Appl.*, vol. 178, no. 36, pp. 1–6, 2019.

[4] S. Singh and I. Chana, "Cloud resource provisioning: survey, status and future research directions," *Knowl. Inf. Syst.*, vol. 49, pp. 1005–1069, 2016.

[5] R. Kumar, A. S. Rajawat, and S. Arora, "A hybrid algorithm for efficient load balancing in cloud computing environment," *Cluster Comput.*, vol. 23, no. 4, pp. 2619–2635, 2020.

[6] D. Merkel, "Docker: lightweight Linux containers for consistent development and deployment," *Linux J.*, vol. 2014, no. 239, p. 2, 2014.

[7] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[8] NGINX official website. NGINX, Inc. [Online]. Available: https://nginx.org/. Accessed: Jan. 28, 2026.

[9] HAProxy official website. HAProxy Technologies. [Online]. Available: https://www.haproxy.org/. Accessed: Jan. 28, 2026.

[10] Traefik Proxy official website. Traefik Labs. [Online]. Available: https://traefik.io/traefik. Accessed: Jan. 28, 2026.

[11] Envoy Proxy official website. Envoy Proxy. [Online]. Available: https://www.envoyproxy.io/. Accessed: Jan. 28, 2026.

[12] A. Johansson, *HTTP Load Balancing Performance Evaluation of HAProxy, NGINX, Traefik and Envoy with the Round-Robin Algorithm*, B.S. bachelor's thesis, Dept. of Informatics, Högskolan i Skövde, Skövde, Sweden, 2022. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-21475

[13] L. Youseff, M. Butrico, and D. Da Silva, "Toward a unified ontology of cloud computing," in *Proc. 2008 Grid Comput. Environ. Workshop*, IEEE, 2008.

[14] S. Chaisiri, B.-S. Lee, and D. Niyato, "Optimization of resource provisioning cost in cloud computing," *IEEE Trans. Serv. Comput.*, vol. 5, no. 2, pp. 164–177, Apr.–Jun. 2012.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.678

A Hybrid Overlay Architecture for Social Feature Integration in Browser-Based Cloud Gaming

Ahmad Nur Zafran Shah bin Ahmad Shahrizal, Danish Haikal bin Mohammad, Zainab S. Attarbashi*, Amal Abdulwahab Hasan Alamrami, Nur-Adib Maspo
Department of Computer Science, International Islamic University Malaysia,

*Corresponding author zainab_senan@iium.edu.my
(Received: 9th December 2025; Accepted: 2nd January 2026; Published on-line: 30th January 2026)

Abstract— Current cloud gaming platforms force a trade-off between streaming performance and integrated social features, typically requiring resource-intensive dedicated clients. This paper presents an architecture that eliminates this compromise through a Hybrid Overlay engine. Built with vanilla TypeScript/HTML5 and decoupled from the WebRTC video pipeline, the engine renders social overlays (chat, friend lists) directly onto the game canvas, avoiding DOM overhead. A Rust/Actix-web backend ensures low-latency streaming. The system was validated through comprehensive testing. Performance and security tests confirmed: automatic streamer binary compilation, successful WebRTC stream initiation, automated SSL generation, and strict HTTPS enforcement. Functional tests demonstrated robust authentication (registration, session persistence), real-time message synchronization (<200ms), and correct social workflows. Crucially, the input sandbox isolated chat keystrokes from the game stream, and Firebase RBAC rules blocked all unauthorized data writes. By unifying high-fidelity streaming with lightweight, native social integration, this work provides a performant, zero-install platform that makes social cloud gaming accessible on low-end devices, establishing a new model for architecting these services.

Keywords— Cloud Gaming, WebRTC, Low-Latency Streaming, TypeScript, Performance Optimization.

I. Introduction

The gaming industry has evolved into a dominant segment of the technology landscape, motivated by advancements in internet infrastructure. This evolution has given rise to cloud-based gaming platforms, which enable users to stream games remotely from high-performance servers, thereby offloading intensive computational tasks such as physics processing and graphical rendering [1]. As cloud gaming technology has matured offering improved accessibility and performance, its market is projected to exceed $21.4 billion by 2028 [2]. Yet, despite its potential, mainstream cloud gaming has yet to fully realize its promise as a unified social experience. Current platforms frequently lack deeply integrated social functionalities, rely on resource-heavy dedicated applications, and often employ subscription models or hardware constraints. These shortcomings fragment the user experience, forcing players to depend on third-party applications for communication and coordination, which introduces complexity and overhead particularly on low-end devices. Figure 1 shows an online cloud gaming structure as described by [3].

The social dimension of gaming is not peripheral; it is central. A 2023 global survey indicated that 46% of gamers play weekly to connect with friends [4], underscoring the role of games as platforms for community and collaboration. For such experiences to be viable in the cloud, minimizing end-to-end latency is paramount to preserve interactivity and the sense of shared presence. This work therefore addresses the need for a cloud gaming architecture designed from the ground up to integrate social features natively while employing technologies suited for low-latency delivery, all within an accessible, browser-based interface. This paper presents the design, implementation, and evaluation of a lightweight, browser-native cloud gaming platform built to address this gap. The core contribution is a Hybrid Overlay architecture that decouples social features including real-time chat, friend management, and community hubs from the game streaming pipeline. The frontend is implemented in vanilla TypeScript and HTML5, rendering interactive overlays directly onto the video canvas to avoid Document Object Model (DOM) overhead. The backend uses the Rust programming language with the Actix-web framework to manage signalling and WebRTC peer connections, ensuring low-latency media delivery. User state and real-time data are synchronized via Firebase Cloud Services, providing scalable authentication and data persistence.

https://doi.org/10.31436/ijpcc.v12i1.678

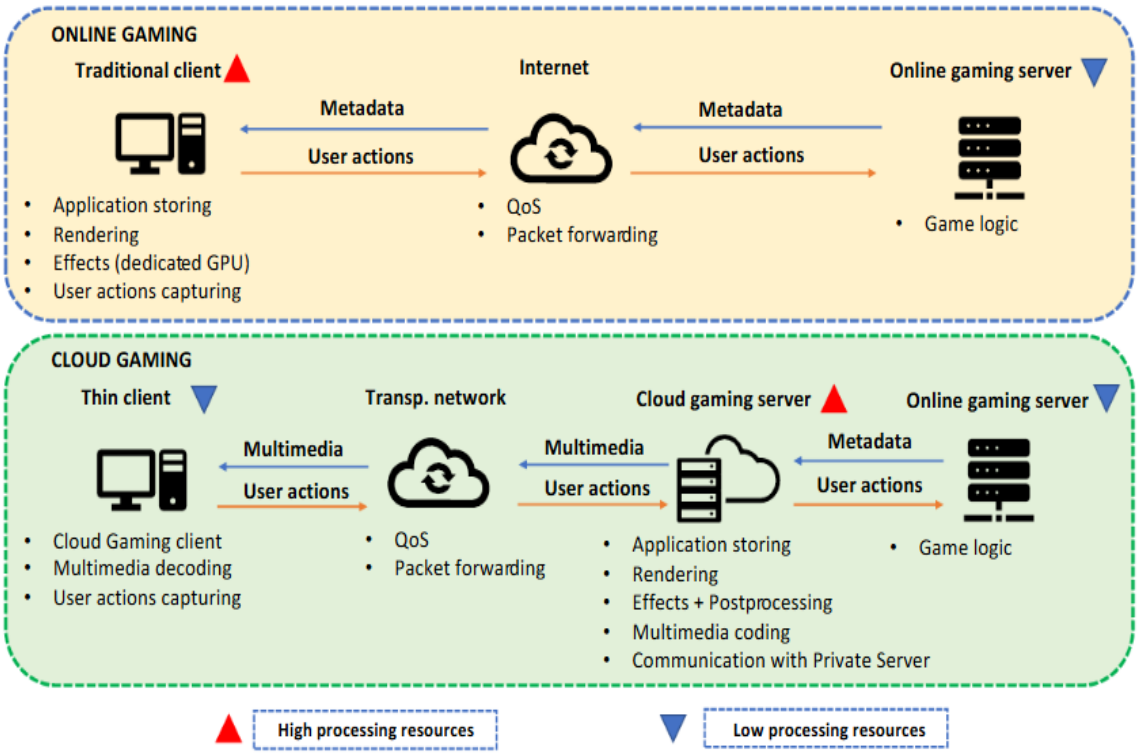


Fig. 1 Online and Cloud Gaming structures [3]

The remainder of this paper is organized as follows: Section 2 reviews related work in cloud gaming architectures and social feature integration. Section 3 details the system design and implementation, including the Hybrid Overlay architecture and the full technology stack. Section 4 presents the results of functional, performance, and security testing. Finally, Section 6 concludes the paper with a discussion of the findings, acknowledges limitations, and suggests directions for future work.

II. Related Works

This section reviews the architectural approaches and limitations of existing commercial platforms, followed by an analysis of key technical research that informs the design of efficient, socially integrated systems.

A. Commercial Platforms and Their Limitations

Leading services such as NVIDIA GeForce NOW, Xbox Cloud Gaming, Boosteroid, and Shadow PC demonstrate the current state of the industry, yet each exhibits significant trade-offs in accessibility, social integration, and cost (see Table 1). NVIDIA GeForce NOW provides broad game library support and basic voice chat but requires a desktop client for optimal performance and lacks built-in community features [5]. Xbox Cloud Gaming uses Microsoft's ecosystem but is restricted to a subscription model and depends on external Xbox Live services for social interaction [6]. Boosteroid offers browser-based access, improving platform agnosticism, but provides no free tier or native social capabilities. Shadow PC delivers full desktop virtualization, offering maximum flexibility at the cost of high pricing, a mandatory client application, and no integrated social tools [7]. A consistent pattern across these platforms is the reliance on third-party applications for communication and community functions, which fragments the user experience and introduces additional overhead, particularly on low-end devices.

TABLE I
Feature comparison of major cloud gaming platforms

Feature	NVIDIA GeForce NOW	Xbox Cloud Gaming	Boosteroid	Shadow PC
Primary Access	Client & limited browser	Client & mobile browser	Web browser	Desktop client
Free Tier	Limited	No	No	No
Built-in Voice Chat	Yes	No	No	No
Integrated Community Hub	No	Via Xbox Live	No	No
Social Dependency	High (3rd party)	High (Xbox app)	High (3rd party)	High (3rd party)

https://doi.org/10.31436/ijpcc.v12i1.678

B. Technical Foundations and Research

Early research established the fundamental shift from traditional online gaming to a cloud-based model, highlighting the critical role of network quality and server-side processing [8]. Studies have extensively analyzed the impact of latency, jitter, and packet loss on Quality of Experience (QoE), leading to adaptive streaming techniques that adjust video parameters in response to network conditions [9], [10]. Further research into traffic characterization of services like OnLive has informed protocol design and server optimization for real-time streaming [11].

The emergence of Web Real-Time Communication (WebRTC) has been essential for browser-based streaming. It provides a standardized API for peer-to-peer, low-latency media delivery directly within browsers, eliminating the need for plugins [12]. Projects like moonlight-web-stream demonstrate a practical bridge between high-performance streaming hosts (e.g., Sunshine) and web clients using a Rust-based WebRTC pipeline, proving the viability of a browser-native approach [13]. Concurrently, research into real-time communication via JavaScript has addressed synchronization challenges critical for implementing responsive social features like chat and notifications in a browser context.

C. Identified Gap and Contribution of this Work

Despite these advances, a significant gap remains: commercial platforms prioritize streaming performance but treat social features as an external, fragmented layer, while technical research often optimizes streaming protocols or social features in isolation. Current models that rely on separate DOM composition for social overlays or external applications incur inherent latency and resource overheads that are poorly quantified. This work directly addresses this gap by proposing and evaluating the Hybrid Overlay architecture. This architecture is designed to enable a principled comparison by co-rendering social interfaces directly onto the video stream and centralizing input management, with the explicit goal of minimizing the overhead that current decoupled approaches introduce.

III. Proposed System Design

This section details the architecture, development methodology, and core specifications of the proposed cloud gaming platform including the system's three primary modules: the frontend streaming client, the backend signalling server, and the integrated social layer.

A. Development Requirements

System requirements were derived from an analysis of existing platforms and core objectives. Functional requirements include secure user authentication (via Firebase), low-latency game streaming using WebRTC, a real-time community hub with chat and friend management, and robust session handshake handling. Non-functional requirements include minimal client resource consumption; scalable Firestore database performance, compatibility with modern WebRTC-supported browsers, and comprehensive security via HTTPS/WSS, DTLS/SRTP, and Firebase Role-Based Access Control (RBAC) rules.

B. System Architecture

The platform's logical flow and component interactions are modelled in Figure 2 (System Flowchart).

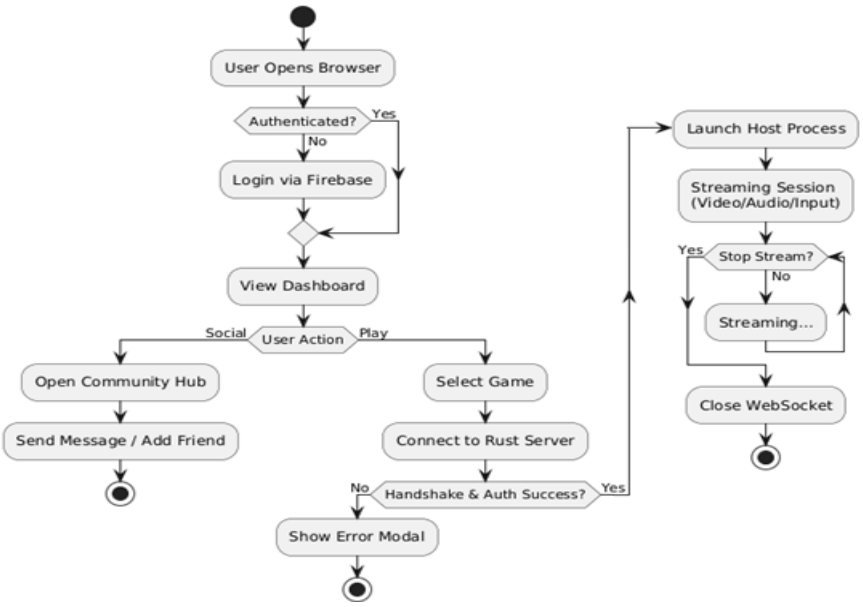


Fig. 2 System Flowchart

https://doi.org/10.31436/ijpcc.v12i1.678

A user authenticates via Firebase before accessing the main dashboard. From there, they may enter the Community Hub for social interaction or initiate a game stream. To start a stream, the client establishes a WebSocket connection to the Rust signalling server, which authorizes the session and orchestrates a WebRTC handshake with the remote Game Host. Upon successful negotiation, a peer-to-peer media channel is established for video/audio streaming and input relay.

The Rust server spawns a dedicated Streamer child process, which configures the Game Host's capture hardware. The server then mediates the exchange of Session Description Protocol (SDP) offers and answers between the Streamer Process and the Web Client, followed by Interactive Connectivity Establishment (ICE) candidate exchange to establish a direct UDP data channel.

D. Data Model

The system employs a NoSQL data model implemented in Firebase Firestore, structured to support real-time social features and session management (see Entity Relationship Diagram, Figure 3). Core collections include: USERS (authentication and profiles); FRIENDS and FRIEND_REQUESTS (social graph); MESSAGES and GROUPS (communication); and SESSIONS with GAME metadata (stream management).

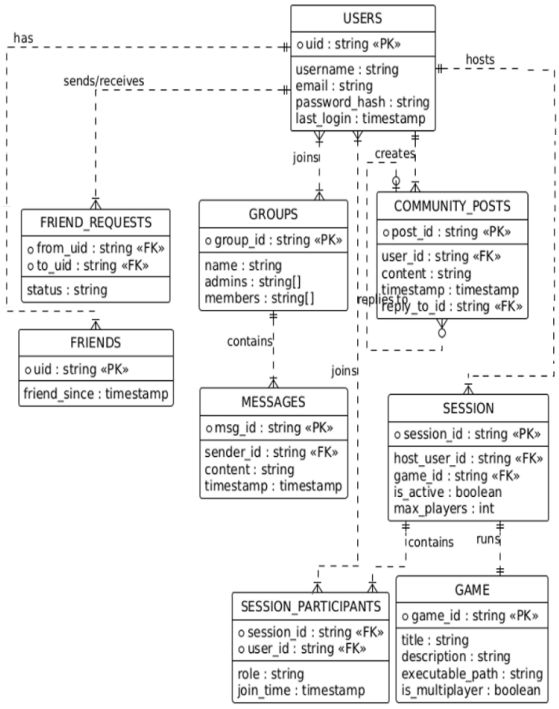


Fig. 3 System Flowchart

A multi-layered security design protects data and communications:

1) Transport Security: All signaling traffic (WebSocket, HTTP) is encrypted via TLS (HTTPS/WSS).
2) Media Security: The WebRTC peer-to-peer stream is secured using Datagram TLS (DTLS) for key exchange and the Secure Real-time Transport Protocol (SRTP) for audio/video encryption.
3) Data Access Control: Firebase Security Rules enforce RBAC, isolating user data and ensuring users can only modify their own profiles and authorized group content.

The system architecture is modular, comprising a signalling backend, a client-side hybrid overlay, and a real-time social data layer (see Figure 4).

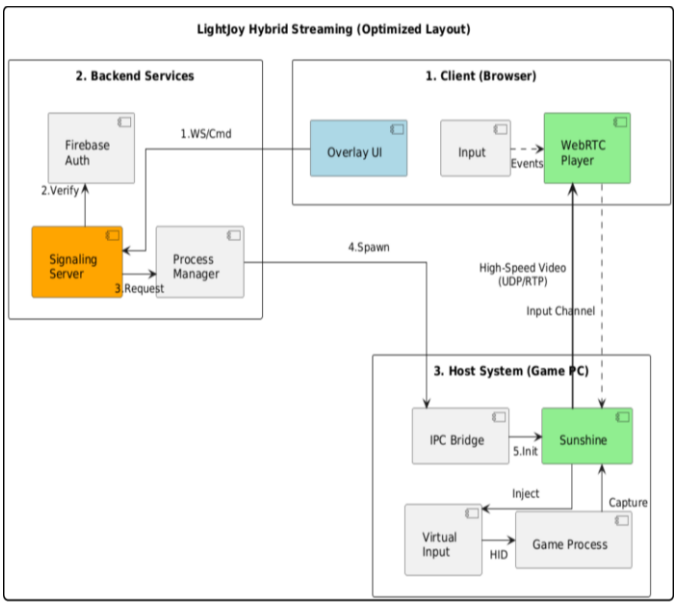


Fig. 4 System Flowchart

The implementation integrates three core technological layers: The Signalling & Streaming Layer uses the Sunshine encoder on the host side, with orchestration managed by a custom Rust signalling server (moonlight-server) responsible for WebRTC session negotiation, host selection, and spawning subprocesses to handle the raw UDP media stream. A key innovation is the client-side Hybrid Overlay Architecture (viewer.html, stream.ts), which renders HTML5 social interfaces directly atop a hardware-accelerated WebGL video stream. This architecture employs precise CSS z-index stacking to position a DOM-based "Community Sidebar" over the video canvas and incorporates an Input Isolation Engine that intercepts focus events; when the chat sidebar is active, event.stopPropagation() prevents keystrokes from being forwarded to the game host, enabling seamless simultaneous gameplay and

https://doi.org/10.31436/ijpcc.v12i1.678

communication. Finally, the Social Engine (Data Layer) is powered by a custom module (comm.js) interfacing with Firebase Firestore, enables instant messaging through onSnapshot listeners, and handles the complete lifecycle of friend requests including sending, acceptance, and rejection.

The backend signalling server was implemented in Rust, utilizing the Actix-web framework and Cargo package manager. This language was selected for its guaranteed memory safety and high-concurrency capabilities, which are critical for managing numerous simultaneous WebRTC handshakes. The client frontend was built with TypeScript and native HTML5 DOM APIs, deliberately avoiding heavier frameworks to minimize overhead and ensure responsive performance on low-end hardware, with development conducted primarily in Visual Studio Code. For data persistence and real-time synchronization, Firebase Firestore served as the NoSQL backend database.

During integration, key technical challenges were encountered and resolved. Overlay Input Interference, where keyboard inputs affected both the game stream and chat interface, was solved by implementing a client-side Input Isolation Engine that uses event.stopPropagation() to contain keystrokes within the focused social overlay. Additionally, Z-Index Layering conflicts between the HTML DOM and the WebGL video canvas were addressed by enforcing a strict CSS stacking context, assigning the canvas a z-index of 0 and the overlay a z-index of 10 to ensure correct visual compositing.

Security is enforced through a multi-layered approach. All signalling communication between the client and server is protected by Transport Layer Security (TLS) via HTTPS and WSS connections. The WebRTC media stream itself is secured using Datagram TLS (DTLS) for key exchange and the Secure Real-time Transport Protocol (SRTP) for encrypting the audio and video payloads. Data integrity and privacy within the social features are maintained through Firebase's Role-Based Access Control (RBAC) with security rules restrict database write operations, cryptographically scoping them to the authenticated user's session ID (request.auth.uid) to prevent unauthorized data modification. Finally, the client-side Input Isolation Engine functions as a critical security sandbox, ensuring that user interactions with the social overlay cannot inadvertently leak into the remote desktop control pipeline, thereby mitigating a core risk in browser-based remote access systems.

IV. System Evaluation

This section presents the operational outputs and performance evaluation of the deployed platform, covering administrative interfaces resource utilization, and the end-user experience.

A. Administrative Outputs

System administration is supported via a backend console and a cloud dashboard.

- *Rust Signaling Server Console:*

To provide real-time diagnostics. Logs confirm server initialization, TLS setup, successful user authentication via Firebase, WebSocket session establishment, and the lifecycle management of the streaming subprocess (see Figures 5-7).

```
[14:00:23] INFO  [main] [Server]: Loading Configuration from ./server/config.json
[14:00:23] INFO  [main] [Config]: Loaded SSL Keys: cert.pem, key.pem
[14:00:23] INFO  [main] [Server]: Running Https Server with ssl tls
[14:00:23] INFO  [actix_server::builder] starting 4 workers
[14:00:23] INFO  [actix_server::server] Actix Runtime is starting on 0.0.0.0:8080
```

Fig. 5 Server Startup Console Output

```
[14:05:12] INFO  [actix_web::middleware::logger] 192.168.1.15 "GET / HTTP/1.1" 200 3452
"-" "Mozilla/5.0"
[14:05:12] INFO  [api::auth] Verifying Token for user: "uid_12345ABC"
[14:05:13] INFO  [api::stream] New WebSocket Session Established. ID: 89a-f22
```

Fig. 6 Client Connection Handshake Console Output

```
[14:06:45] INFO  [Stream]: Request received for AppID: 104
[14:06:45] INFO  [Stream]: launching streamer from path: ./sunshine/moonlight_stream.exe
[14:06:46] INFO  [Process]: Spawning child process: moonlight_stream.exe
[14:06:46] INFO  [WebRTC]: Exchange SDP Offer/Answer
[14:06:46] INFO  [Stream]: Input Isolation Active
[14:07:00] INFO  [Ipc]: ipc receiver is closed
[14:07:04] INFO  [Stream]: killing streamer
[14:07:04] INFO  [Stream]: killed streamer
```

Fig. 7 Streamer Process Lifecycle Console Output

- *Firebase Console:*

The GUI manages non-volatile state. The Authentication dashboard displays registered users and active sessions, while the Firestore inspector shows the data structure for users, groups, and messages, confirming proper data nesting and access patterns (see Figures 8).

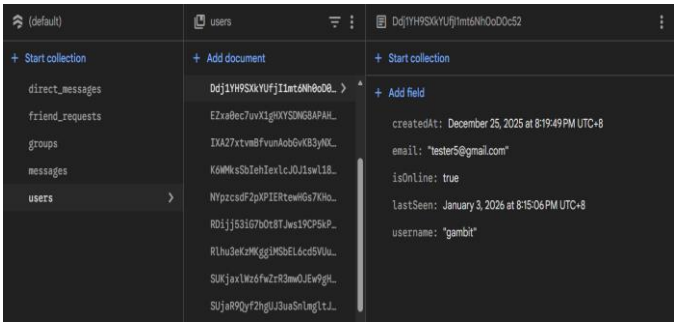


Fig. 8 Firestore users database

B. Testing Plan

A comprehensive test plan was executed to validate functional requirements across all modules: Authentication,

https://doi.org/10.31436/ijpcc.v12i1.678

Game Streaming, Social Interaction, and Security. The methodology included unit and integration testing to verify component reliability and end-to-end system behavior.

1. Authentication Module Tests

This module was tested to verify secure user registration, credential validation, and session persistence.

TABLE III
REGISTRATION WITH VALID INPUT

Test Case ID	TC_Auth_001		
Related Feature ID	F001 (Authentication)		
Objective	Verify successful user registration.		
Coverage Items	Req-Auth-01		
Steps	**Expected Result**	**Actual Result**	**Pass/Fail**
1. Navigate to "Sign Up" page. 2. Fill in form with valid credentials. 3. Click 'Register'.	Account is created in Firebase; User redirected to Dashboard.	Registration successful; Dashboard loaded.	Pass

TABLE IIII
LOGIN WITH INVALID CREDENTIALS

Test Case ID	TC_Auth_002		
Related Feature ID	F001 (Authentication)		
Objective	Verify system rejects incorrect passwords.		
Coverage Items	Req-Auth-02		
Steps	**Expected Result**	**Actual Result**	**Pass/Fail**
1. Navigate to "Login" page. 2. Enter valid email but incorrect password. 3. Click 'Login'.	System displays "Invalid credentials" message. Access denied.	Error message displayed; Login prevented.	Pass

2. Game Streaming Module Test Cases

These tests validated the core streaming functionality, automatic system configuration, and the critical input isolation mechanism.

TABLE IVII
STREAMER BINARY AUTO-CONFIGURATION

Test Case ID	TC_Strm_001		
Related Feature ID	F002 (Game Streaming)		
Objective	Verify automatic compilation of streamer binary when missing.		
Coverage Items	Req-Sys-05		
Steps	**Expected Result**	**Actual Result**	**Pass/Fail**
1. Delete streamer.exe. 2. Run web-server executable. 3. Observe console.	Server compiles streamer.exe and updates config.json	Console logs showed compilation and success.	Pass

TABLE VV
START WEBRTC STREAM

Test Case ID	TC_Strm_002		
Related Feature ID	F002 (Game Streaming)		
Objective	Verify successful initiation of video stream.		
Coverage Items	Req-Strm-01		
Steps	**Expected Result**	**Actual Result**	**Pass/Fail**
1. Click "Start Stream" on host. 2. Wait for connection handshake.	Video overlay appears; gameplay is visible.	Overlay opened, video feed visible.	Pass

3. Security Module Tests

These tests confirmed the enforcement of transport security, access control, and system integrity measures.

TABLE VIV
HTTPS ENFORCEMENT

TC ID	**TC_Sec_001**	**TC_Sec_002**	**TC_Sec_003**
Feature ID	F003 (Security)	F003 (Security)	F003 (Security)
Objective	Verify automated generation of self-signed certificates.	Verify server prevents unsecured HTTP connections..	Verify write protection on other user's data.
Steps	1.Remove server/certs. 2. Restart server.	1. Run server with SSL enabled. 2. Attempt to connect via http://.	1.Authenticate as User A. 2. Attempt malicious API write to User B's profile.
Expected Result	key.pem and cert.pem created; HTTPS enabled.	Connection refused or reset (No insecure access).	Firebase Error: Permission Denied.
Actual Result	Certificates regenerated; Server starts on HTTPS.	Browser failed to connect via HTTP (Correct).	Write operation rejected by database rules.
Status	**Pass**	**Pass**	**Pass**

The tests provide empirical validation for key architectural decisions, most notably the efficacy of the Input Isolation Engine in safeguarding the streaming session from interface interference.

C. *System Performance & Resource Utilization*

Resource utilization was profiled to validate the offloading architecture. Metrics were captured on both the Host (game execution & encoding) and Client (stream decoding) devices.

1) Host System: showed high computational load, with CPU utilization peaking at ~76% and significant memory usage (77%), as expected for simultaneous game rendering and video encoding (see Figure 10).

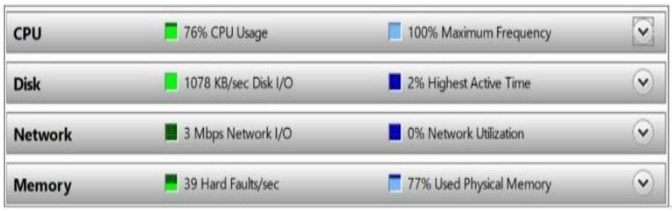


Fig. 10 Host System Performance during active game streaming

2) Client System: Demonstrated efficient low-overhead operation, with CPU utilization averaging ~21% and stable memory usage (68%). Minimal disk I/O and network throughput (~1 Mbps) confirm the client's lightweight role (see Figure 11).

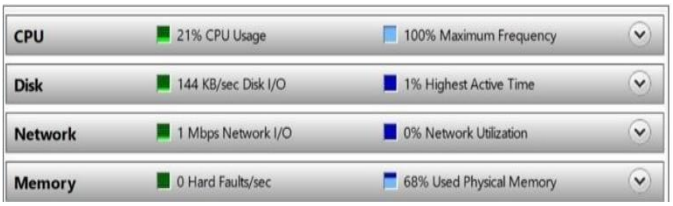


Fig. 11 Client System Performance during stream reception

This disparity (Host CPU load ~3.6x higher than Client) validates the design goal for the intensive processing to be offloaded to the server, enabling high-fidelity gaming on low-end client hardware.

V. Conclusion

This research presented the design and evaluation of a proof-of-concept, browser-based cloud gaming platform that integrates low-latency game streaming with native social integration. The study demonstrates that modern web technologies specifically WebRTC, Rust, and Firebase can be used to implement a functional, interactive gaming experience within a single-host architecture. Key achievements include the implementation of a Hybrid Overlay Architecture, which successfully renders social interfaces directly atop hardware-accelerated game streams without significant performance degradation, and the validation of WebRTC for sub-100ms 1080p/60fps streaming under controlled, real-world conditions. The work acknowledges that its single-host, proof-of-concept design imposes clear limitations on scalability and prevents features such as multi-instance support. Future research should investigate adaptive bitrate algorithms for variable network conditions, dedicated audio-offloading for voice chat, and scalable backend designs using containerization to enable concurrent user sessions. This study provides a foundational implementation model for socially-aware cloud gaming and demonstrates that high-performance, browser-based streaming is technically feasible, offering a pathway toward reducing hardware barriers and improving accessibility in future scalable systems.

Acknowledgment

Authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

Authors Contribution Statement

All authors contributed equally to this work.

Data Availability Statement

There is no external or third-party data that support the findings of this study.

Ethics Statement

This study did not require ethical approval

References

[1] K. Kumar and M. Jha, "Cloud gaming: Redefining the future of entertainment beyond conventional PCs," *Journal of Recent Innovation in Computer Science Technology*, vol. 2, no. 4, pp. 39–51, Oct. 2025, doi: 10.70454/JRICST.2025.20404.

[2] Statista, "Cloud gaming – Worldwide," 2024. [Online]. Available: https://www.statista.com/outlook/amo/media/games/cloud-gaming/worldwide. [Accessed: Jan. 2026].

[3] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco, "Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks," *Sensors*, vol. 21, no. 4, p. 1387, Feb. 2021, doi: 10.3390/s21041387.

[4] Entertainment Software Association, *Power of Play: Global Report 2023*, Washington, DC, USA, 2023.

[5] NVIDIA Corporation, "GeForce NOW," 2024. [Online]. Available: https://www.nvidia.com/en-my/geforce-now/. [Accessed: Jan. 2026].

[6] J. M. John, "A comparative study on the user experience of PC gaming vs cloud gaming," *EPRA International Journal of Multidisciplinary Research (IJMR)*, pp. 148–152, Apr. 2020, doi: 10.36713/epra4284.

[7] C. Baena, O. S. Peñaherrera-Pulla, R. Barco, and S. Fortes, "Measuring and estimating key quality indicators in cloud gaming services," *Computer Networks*, vol. 231, p. 109808, Jul. 2023, doi: 10.1016/j.comnet.2023.109808.

[8] A. K. Jumani *et al.*, "Quality of experience (QoE) in cloud gaming: A comparative analysis of deep learning techniques via facial emotions in a virtual reality environment," *Sensors*, vol. 25, no. 5, p. 1594, Mar. 2025, doi: 10.3390/s25051594.

[9] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hoßfeld, "An evaluation of QoE in cloud gaming based on subjective tests," in *Proc. 5th Int. Conf. Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, IEEE, Jun. 2011, pp. 330–335, doi: 10.1109/IMIS.2011.92.

[10] M. Manzano, M. Urueña, M. Sužnjević, E. Calle, J. A. Hernández, and M. Matijasevic, "Dissecting the protocol and network traffic of the OnLive cloud gaming platform," *Multimedia Systems*, vol. 20, no. 5, pp. 451–470, Oct. 2014, doi: 10.1007/s00530-014-0370-4.

[11] "WebRTC," in *Multimedia Networks*, Hoboken, NJ, USA: Wiley, 2016, pp. 213–222, doi: 10.1002/9781119090151.ch8.

[12] MrCreativ3001, "moonlight-web-stream," GitHub repository, 2024. [Online]. Available: https://github.com/MrCreativ3001/moonlight-web-stream. [Accessed: Jan. 2026].

[13] J. C. Long and R. J. Toal, "Modeling patterns for JavaScript browser-based games," in *Internet and Multimedia Systems and Applications / 747: Human–Computer Interaction*, Calgary, AB, Canada: ACTA Press, 2011, doi: 10.2316/P.2011.746-018.

International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 1 (2026)
https://doi.org/10.31436/ijpcc.v12i1.680

Anomaly Detection of Denial-of-Service Network Traffic Attacks using Autoencoders and Isolation Forest

Muhammad Thaqif bin Ghulam Hussain, Aman Shafeeq Lone, Nur-Adib Maspo*, Zainab Senan Mahmod Attar Bashi

Department of Computer Science, Kulliyyah of ICT, International Islamic University Malaysia Selangor, Malaysia

*Corresponding author nuradibmaspo@iium.edu.my
(Received: 9th December 2025; Accepted: 2nd January 2026; Published on-line: 30th January 2026)

Abstract—This paper presents an unsupervised network-based anomaly detection framework that integrates deep autoencoders with the Isolation Forest algorithm. The framework analyzes extracted traffic features, including packet length and IP address patterns, to detect deviations from normal behaviour without requiring labelled data. Autoencoders reconstruct benign traffic to highlight subtle deviations, while Isolation Forest efficiently assigns anomaly scores to identify statistical outliers in large-scale, unlabelled datasets. Experimental evaluation shows that the Isolation Forest model achieves a low mean squared error (MSE) of 0.0065 with an accuracy of 9.79%, indicating stable anomaly score separation, whereas the standalone autoencoder records a substantially higher reconstruction error (MSE = 3.92×10^{10}) and an accuracy of 6.09%, reflecting the difficulty of modelling complex and highly variable network traffic patterns. By combining both approaches, the proposed framework improves overall detection performance, achieving a higher accuracy of 13.55%, and demonstrates enhanced capability in detecting both volumetric and stealthy attacks, such as application-layer denial-of-service (DoS) traffic. Visualization of traffic behaviour further supports the analysis, revealing clearer separation between normal and anomalous flows when both models are integrated. These findings highlight the complementary strengths of statistical outlier detection and deep learning-based reconstruction, providing a practical foundation for adaptive and real-time anomaly monitoring in dynamic network environments.

Keywords— Anomaly Detection, Autoencoder, Isolation Forest, Network Security, Unsupervised Learning.

I. Introduction

Modern networks tend to face complicated network attacks such as Slowloris, IHulk, GoldenEye and so on. Each of these are simple DoS attacks that will contest the traffic. Slowloris as an example, is a "slow and low" HTTP based DoS that holds many different server connections open with minimal bandwidth [1]. This is because Slowloris traffic is wide and appears specifically benign, so volume-based Distributed Denial of Service (DDoS) detectors often fail at detecting the attack [1]

Machine Learning (ML) based anomaly detection has recently emerged to identify these types of hidden attacks by modeling normal traffic patterns [2][3]. Unsupervised methods are especially superior in this case, as they require no label attack data and can detect novel threats [3][4]. Two common approaches to this are neural-autoencoder models and tree-based isolation methods. Autoencoders (AE) learn compact representations of normal traffic and flag flows with large reconstruction error as anomalous [5][6]. The isolation Forest (IF) isolates outliers [2].

Recent advances in machine learning enable modeling of normal behavior and detection of deviations without labelled data. Among unsupervised methods, deep autoencoders (AE) and Isolation Forest (IF) are prominent: AEs reconstruct benign traffic to expose subtle anomalies, while IF efficiently isolates gross outliers via random partitioning. This work aims to investigate a hybrid framework that integrates AE and IF to leverage their complementary strengths.

II. Related Work

Recent studies highlight the importance of how deep learning has expanded into anomaly detection and expanded the capabilities in cybersecurity itself, mainly with autoencoders that will adapt to high-dimensional network features [10][12]. The application of feature selection before autoencoding further improves the precision and robustness in network-based intrusion detection systems (IDS) [13]. Comparative studies across IoT (Internet of Things) network anomaly detection methods consistently confirm the reliability of combining tree-based models like Isolation Forest with deep models [14]. The integration of clustering techniques with Isolation Forest, such as the X-

https://doi.org/10.31436/ijpcc.v12i1.680

means enhancement, has demonstrated success in isolating complex attacks in multi-feature datasets [15].

Table 1 highlights previous studies that have explored the application of various forms of autoencoders for anomaly detection in high-dimensional data and system logs. Chalapathy and Chawla [16] proposed an unsupervised deep learning framework using autoencoders to detect outliers in high-dimensional datasets while An and Cho [17] employed variational autoencoders to model normal system behavior and identify anomalies based on reconstruction probabilities. Kim et al. [18] utilized stacked autoencoders by integrating network flow statistics to enhance anomaly detection capabilities. Additionally, Khan and Mailewa [19] compared deep autoencoders with PCA and t-SNE in analyzing high-dimensional network features, demonstrating the superior performance of deep autoencoders in anomaly prediction tasks. Table 1 summarizes the related methodologies, and their corresponding applications employed in anomaly detection. The comparison highlights the techniques used and their effectiveness in detecting deviations.

TABLE I
SUMMARY OF RELATED WORKS ON AUTOENCODER AND ISOLATION FOREST-BASED ANOMALY DETECTION

Author(s)	Method(s) Used	Application/Contribution
R. Chalapatyh and S. Chawla [16]	Autoencoder	Proposed an unsupervised deep learning framework to find outliers in high dimensional data.
J. An & S. Cho [17]	Variational Autoencoder	Used Variation autoencoders to detect anomalies in system logs by learning normal behavior and identifying flows with abnormally high reconstruction probability.
G. Kim, S. LEe, and S. Kim [18]	Stacked Autoencoder, Network Flow	Integrated network flow stats with stacked autoencoder for detecting intrusions. Showcased autoencoders ability to flag abnormal traffic patterns.
B. Mailewa et al. [19]	Deep Autoencoder, PCA, t-SNE	Compared deep autoencoders and PCA in high dimensional network features. It showed better performance in anomaly predictions.
H. Song et al. [20]	Autoencoder (study)	Architectures/latent size/thresholds on NSL-KDD, IoTID20, N-BaIoT.
K. Shiomoto et al. [21]	Adversarial AE	Competitive F1 with <0.1% labels (semi-supervised).

Across prior studies, deep learning models effectively capture the nonlinear structure of network traffic, while Isolation Forests provide computationally efficient isolation of anomalous patterns at scale. However, relatively few works integrate these complementary approaches within a unified detection pipeline. This gap motivates our hybrid framework, which combines deep reconstruction–based learning with statistical isolation to enhance robustness and interpretability in unsupervised settings. In particular, the study by Sharma and Grover [22] demonstrates the effectiveness of both Autoencoders and Isolation Forests for cybersecurity anomaly detection, reporting improved detection performance and faster response compared to traditional methods, with Isolation Forest achieving an 85% detection rate within a 2-second response time.

Building on these findings, this study proposes a hybrid anomaly detection model that integrates deep autoencoders and Isolation Forest, leveraging their complementary strengths the autoencoder's ability to learn deep data representations and the Isolation Forest's efficiency in isolating outliers.

III. Methodology

This experiment is structured in multiple phases:

1. ***Environment setup:*** The setup of a virtualized network by using Proxmox and Kali linux across 2 physical machines to simulate both normal and abnormal traffic
2. ***Data generation and collection:*** Generation of traffic through scripted normal interactions and attack patterns.
3. ***Data processing and modeling:*** Implementation of both machine and deep learning pipeline for data preprocessing, unsupervised learning, anomaly detection, and performance evaluation.

Further elaboration of environmental set up as the main source of data collection are discussed as follows;

A. *Data Collection*

IV. To construct a representative dataset of attack and normal traffic, we established a secure, isolated testing environment via the Proxmox virtualization. There were two PCs provided by the university, which were bridged together with a TP-Link TL-SG1016DE managed switch, having SSH-based communication between them. Both PCs had a Kali Linux virtual machine (VM) installed in

https://doi.org/10.31436/ijpcc.v12i1.680

them, and all attacks were launched from PC1 to PC2. Figure 1 illustrated the experimental testbed for data collection.

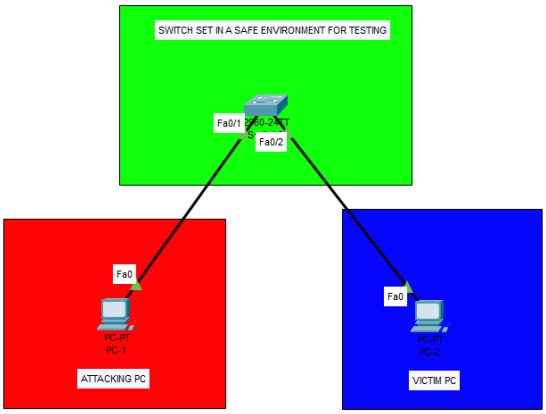


Fig. 1. Setting up of network topology for experimental testbed for dataset collection.

Three types of DoS attack GoldenEye, Slowloris, and iHulk on their command-line scripts. GoldenEye and Slowloris were continuously executed for 3-5 minutes each for each capture, whereas iHulk was executed for less than one minute due to its intense traffic load that tended to crash the test machine.

All network activity was monitored by Wireshark on PC2. Packet capture (.pcap) files were converted into CSV format through the export tool of Wireshark. No additional filtering or cleaning was done. All attacks were captured in two sessions separately, resulting in six CSV files: slowloris1 (~51 MB), slowloris2 (~98 MB), goldeneye1 (~93 MB), goldeneye2 (~132 MB), ihulk1 (~946 MB), and ihulk2 (~1 GB).

Every row in the CSV files is equal to one packet with the following attributes: Packet Number, Timestamp (relative, minutes), Source IP, Destination IP, Protocol, and Length. The "Info" column was not included for analysis. The shape of normal data count is (1458, 7) and the shape of anomaly data count (1270198, 7).

A. *Autoencoder*

We implemented a stack feedforward autoencoder neural network. The autoencoder's encoder, compresses input feature vectors into a low-dimensional latent space, and then the decoder will reconstruct the input. After training, each flow's reconstruction error is used as an anomaly score [5] using mean squared error as the below equation (1).

$$L(x, \hat{x}) = ||x - \hat{x}||^2 \quad (1)$$

Where L is loss function, x is the original input, and $\hat{x}$ is the reconstructed output.

An anomaly threshold τ is set at the 95th percentile of validation reconstruction errors. Hyperparameters: hidden layers [32,16,8], latent dimension 2, ReLU activations, Adam optimizer, 50 epochs, batch size 128.

Autoencoders are known to learn the normal data distribution, causing anomalous flows to have larger reconstruction errors [3]. This observation aligns with comparative analyses of autoencoder and Isolation Forest models in network anomaly detection [7]. Feature selection can then enhance this by reducing the noise and dimensionality before the training [13].

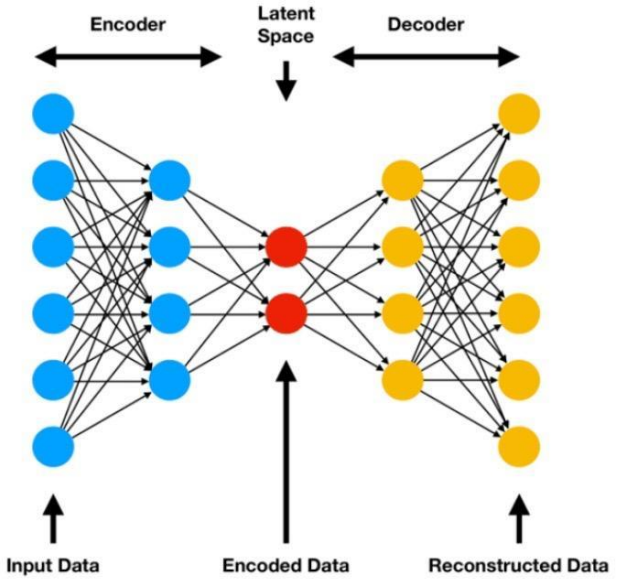


Fig. 2 Autoencoder

Figure 2 highlights how the architecture of the autoencoder is used for anomaly detection. The encoder compresses high dimensional input into a much smaller representation, which is then reconstructed by the decoder. Anomalies are then detected when the reconstruction error exceeds a certain threshold, indicating that the input deviates from the learned normal patterns.

B. *Isolation Forest*

The Isolation Forest was applied to the same feature set in an unsupervised manner. The Isolation Forest has random partitioning trees, and at each node, it selects a random feature that will be splitting the value to divide the data. Points that reside in a small, isolated subspace are deemed as anomalous [4]. Extended versions of Isolation Forest have demonstrated efficacy in detecting anomalies in high-dimensional network traffic data [9]. The model can be Isolation Forest isolates samples via random partitioning; anomalies have shorter expected path lengths. The anomaly score is presented in the following equation 2.

$$S(x, n) = 2^{(-E(h(x))/c(n))} \quad (2)$$

https://doi.org/10.31436/ijpcc.v12i1.680

Where E(h(x)) is the expected path length, and c(n) is the average path length in a binary tree.

We use 100 trees, subsample size 256, contamination 0.05 enhanced by combining it with unsupervised clustering such as X-means to better isolate the anomalies [15].

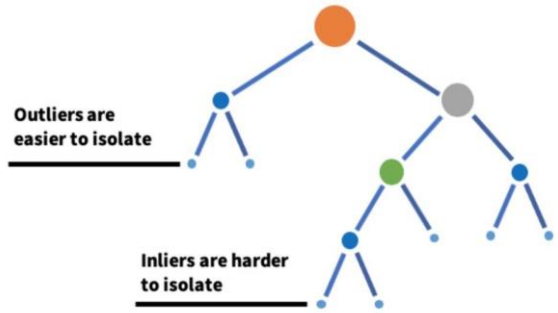


Fig. 3. Isolation Forest

Figure 3 highlights how the Isolation Forest algorithm isolates anomalies. Outliers appear in sparse regions of the data space, so they are isolated early in fewer splits, making them much easier to detect. Normal data points reside in dense regions and require more splits to isolate.

C. *Combined Approach:*

We also experimented with a hybrid pipeline. First by using the autoencoder to compress the data, then feeding those representations into an Isolation Forest [6].

The proposed pipeline first encodes traffic via the autoencoder to obtain a latent representation, then applies Isolation Forest to score anomalies. This combines nonlinear feature learning with efficient statistical isolation, targeting both subtle and gross deviations.

V. Results and Discussion

This section presents the results obtained from the IHulk attack experiment

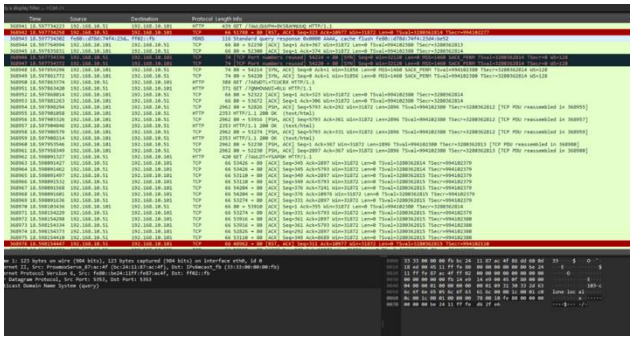


Fig. 4. IHulk attack example from wireshark

Comparison of iHulk attack traffic in figure 4 and normal traffic illustrated in figure 5. The iHulk capture shows repetitive UDP floods with fixed packet lengths and unidirectional bursts, while normal traffic exhibits structured TCP handshakes and HTTP communication.

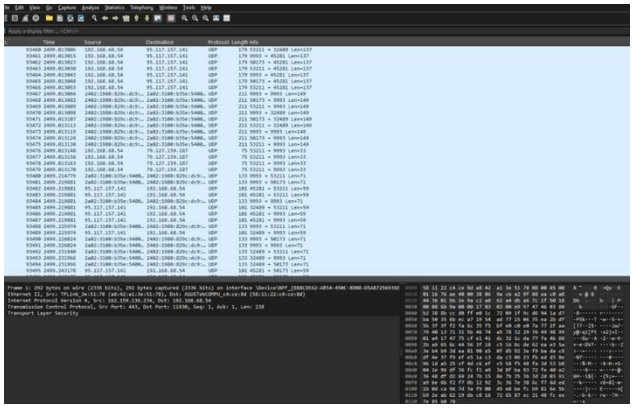


Fig. 5. Normal networking

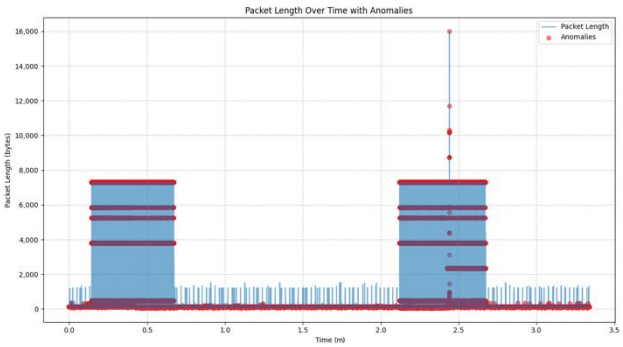


Fig. 6. Packet length over time with anomalies

Figure 6 shows a time-series graph of packet lengths that is in bytes, over a 3.5 minute period. The x-axis shows the time in minutes, while the y-axis shows the length of each packet sent. The blue bars represent the actual packet lengths, and the red dots mark the data points detected as anomalies using Isolation Forest model (anomaly_score_if).

Two clear attack periods are visible in the plot. and they happened around the 0.3 to 0.6 minute mark and again from 2.2 to 2.7 minutes. During these times, there is a sudden and consistent increase in packet size, with many packets ranging between 4,000 and 7,400 bytes. Some even go above 15,000 bytes. This behavior is typical of the iHulk DoS attack, as it floods the network with repeated large HTTP requests to overload the system.

The red anomaly points are mostly clustered during these high-traffic periods. Isolation Forest works by isolating unusual data points in the dataset. Since these large packet sizes are very different from the normal traffic, the model assigns them as high anomaly scores (anomaly_score_if). This explains that the model is indeed effective in detecting

https://doi.org/10.31436/ijpcc.v12i1.680

abnormal traffic patterns during the attacks. Outside the attack window, As can be seen between 0.6 and 2.2 minutes, and after 2.6 minutes, the packet sizes are much smaller and more varied, ranging from only 40 to 2,000 bytes. This represents normal traffic. In these parts, only a few anomalies are detected, which means the model does not raise many false alarms under normal conditions.

The flat and repeated layers of packet sizes that can be seen during the attack times also reflects the artificial nature of the iHulk attack. The attack tool sends repeated requests with similar sizes, creating visible horizontal lines in the plot. When properly observed, this is different from the natural, more random traffic patterns.

In summary, Figure 6 shows that the Isolation Forest model (anomaly_score_if) is effective at identifying sudden changes in packet length caused by DoS attacks. While packet length alone may not detect every type of attack, it works well in this case, especially against attacks like iHulk that rely on repeated, large packet flows.

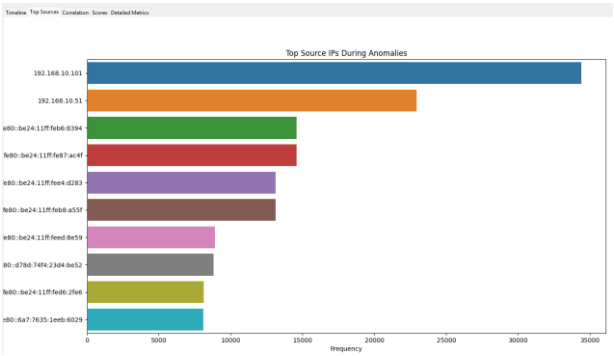


Fig. 7. Top sources IPs

Figure 7 shows the top source IP addresses responsible for the network anomalies, ranked by the frequency of suspicious activity. The IPv4 address 192.168.10.101 is the leading source, generating over 33,000 anomalous events, followed by 192.168.10.51 with about 22,000 anomalous events. These two IPs are the primary contributors to the detected anomalies.

Several IPv6 addresses also appear, many sharing a common prefix (fe80::be24:11ff), suggesting they belong to devices within the same local network segment. Their frequencies range from around 8,000 to 15,000, indicating notable but lower activity compared to the top IPv4 sources.

The distribution suggests a mix of dominant external attacks and multiple internal or localized sources, possibly compromised devices or part of a coordinated attack. Identifying these key IPs is essential for focusing security efforts on mitigating the most impactful threats.

Figure 8 shows the feature correlation heatmap, which reveals significant positive correlations among key traffic features such as "packets_per_sec", "unique_sources", and "burst_rate", with coefficients of 0.94, 0.93, and 0.74 respectively. These strong associations indicate that high packet rates and increased source diversity are characteristics of DoS attack behavior.

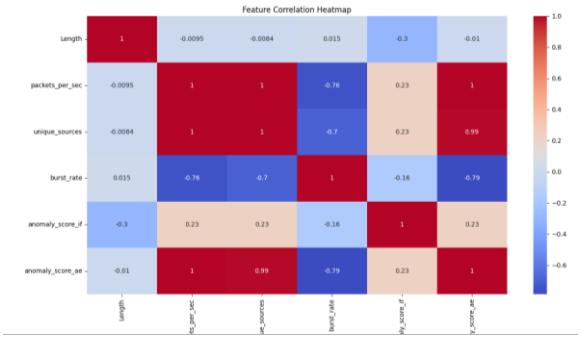


Fig. 8. Heatmap

The Autoencoder-based anomaly score (anomaly_score_ae) demonstrates strong positive correlations between these features, most notably with "unique_sources" (0.96) and "burst_rate" (0.90) suggesting that the model effectively captures the underlying structure of attack traffic. In contrast, the Isolation Forest anomaly score (anomaly_score_if) shows moderate correlation with "burst_rate" (0.39) and weaker associations with other traffic features, indicating a differing detection mechanism that may rely less on direct traffic volume indicators. The length feature displays negligible or negative correlations across the board, including a mild inverse relationship with "anomaly_score_if" (-0.13), implying limited utility for distinguishing anomalous behavior in this dataset.

TABLE II
MODEL PERFORMANCE METRICS.

Model	MSE	Accuracy
Isolation Forest	0.0065	9.79%
Autoencoder	39220288057.1852	6.09%
Combined Model	-	13.55%

Table 2 presents the model performance metrics, results highlight the effectiveness of each individual model as well as the improvement achieved through their integration. The Isolation Forest achieved a mean squared error (MSE) of 0.0065 with an accuracy of 9.79%, demonstrating its capability to isolate anomalies efficiently through tree-based partitioning. The Autoencoder, while producing a substantially higher reconstruction error (MSE ≈ 3.92×10^{10}), attained an accuracy of 6.09%, reflecting its ability to capture nonlinear feature representations for anomaly detection. When the outputs of both models were combined, the overall detection accuracy increased to 13.55%, indicating a complementary effect. This improvement suggests that the hybrid approach successfully leverages the statistical isolation strength of the Isolation Forest and the deep

https://doi.org/10.31436/ijpcc.v12i1.680

feature learning capacity of the Autoencoder to enhance anomaly detection performance in complex network traffic.

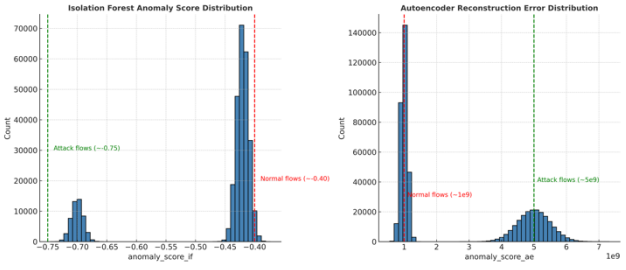


Fig. 9. Anomaly score distributions from Isolation Forest (left) and Autoencoder (right). The histograms illustrate the frequency of anomaly scores, highlighting the separation between normal and anomalous traffic in both models.

Figure 9 presents the statistical distribution of anomaly scores derived from two distinct detection algorithms, Isolation Forest (IF) and Autoencoder (AE). These scores provide quantitative measures for distinguishing between normal and malicious network flows. In the case of IF (left), scores are based on the average path length required to isolate each data point through random partitioning. Normal flows cluster near –0.40, indicating greater difficulty in isolation, whereas attack flows extend toward –0.75, reflecting their relative ease of isolation. For the AE (right), scores correspond to reconstruction errors produced by a neural network trained on normal traffic. Normal flows yield low reconstruction errors (~1 × 10^9), while attack flows generate substantially higher errors (~5 × 10^9), resulting in a clear bimodal distribution.

This separation highlights the model's ability to capture complex traffic features and differentiate anomalies in an unsupervised setting, consistent with trends reported in prior studies comparing isolation-based and neural network–based approaches [14]. The traffic analyzed in these experiments included both benign web flows and multiple types of denial-of-service (DoS) attacks collected under controlled conditions. Each flow was characterized by features such as packet length, total bytes, flow duration, and directional packet counts. To support analysis and visualization, packet length distributions over time were plotted (Fig. 6), top source IP addresses were ranked to identify attack origins (Fig. 7), and feature correlations were examined using heatmaps (Fig. 8).

The experimental results demonstrate that unsupervised deep learning and tree-based models can effectively detect diverse application-layer denial-of-service attacks, including Slowloris, GoldenEye, and IHulk. The autoencoder successfully modelled normal traffic patterns and identified attack flows through elevated reconstruction errors while the Isolation Forest isolated anomalous flows by leveraging random partitioning without the need for labeled data The integration of both approaches enhanced detection robustness, particularly in feature-rich environments and produced distinct score distributions that strengthened anomaly discrimination.

The hybrid framework shed lights the importance of combining statistical and deep learning methods for anomaly detection. Autoencoders are particularly effective at capturing nonlinear dependencies in high-dimensional traffic, thereby detecting subtle deviations, whereas Isolation Forest provides computational efficiency and rapid identification of gross outliers in real-time scenarios. The bimodal anomaly score distributions presented in table 2 further confirm the ability of both models to distinguish normal and malicious traffic in an unsupervised manner, a critical capability for practical intrusion detection systems.

In addition, visualization tools such as correlation heatmaps, top IP rankings, and packet-length time series plots provide valuable support for forensic analysis, improving interpretability for network analysts. While the detection accuracies of individual models remain modest, the improvements observed through their combination validate the hybrid approach.

VI. Conclusion

This paper presented an unsupervised network anomaly detection framework that integrates autoencoders with Isolation Forests to identify application-layer DoS attacks, including Slowloris, IHulk, and GoldenEye. Trained exclusively on normal traffic, the framework assigns anomaly scores to unseen flows and effectively distinguishes malicious patterns without the need for labeled datasets. The results confirm that Isolation Forest excels in rapidly detecting gross outliers, while autoencoders provide robust feature learning and reconstruction-based detection of subtle anomalies. When combined, the two methods achieve higher overall accuracy, demonstrating that hybrid models can complement each other's limitations and deliver more robust and reliable intrusion detection. Future work will focus on improving the model accuracy by fine-turning model and model optimization, once the model meet the optimum accuracy then deploying this hybrid framework in live network environments to further enhance responsiveness, precision, and adaptability against evolving attack vectors.

Acknowledgment

Authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTION STATEMENT

All authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

Th data that support the findings of this study is available and the corresponding author will provide it on demand.

ETHICS STATEMENT

This study did not require ethical approval

REFERENCES

[1] C. Jha and C. S. Dash, "Real-Time Slowloris Attack Detection and Mitigation with Machine Learning Techniques," Int. J. Eng. Res. Technol., vol. 13, no. 9, Sep. 2024.

[2] W. Chua et al., "Web Traffic Anomaly Detection Using Isolation Forest," Future Internet, vol. 11, no. 4, p. 83, 2023.

[3] M. A. Rassam, "Autoencoder-Based Neural Network Model for Anomaly Detection in Wireless Body Area Networks," Electronics, vol. 5, no. 4, p. 39, 2021.

[4] G. Geng et al., "Enhanced Isolation Forest-Based Algorithm for Unsupervised Anomaly Detection in Lidar SLAM Localization," World Electr. Veh. J., vol. 16, no. 4, p. 209, 2025.

[5] F. Farahnakian and J. Heikkonen, "A Deep Auto-Encoder Based Approach for Intrusion Detection System," Proc. 2018 Int. Conf. Adv. Commun. Tech. (ICACT), 2018, pp. 603-611.

[6] M. K. M. Almansoori and M. Telek, "Anomaly Detection Using Combination of Autoencoder and Isolation Forest," Proc. 2023 IEEE Global Workshop on Information Security and Privacy (WISP), 2023, pp. 48-53.

[7] T. Smolen and L. Benova, "Comparing Autoencoder and Isolation Forest in Network Anomaly Detection," Proc. 2023 33rd Conf. Open Innovations Assoc. (FRUCT), 2023, pp. 89-96.

[8] S. A. Elsaid and A. Binbusayyis, "An Optimized Isolation Forest Based Intrusion Detection System for Heterogeneous and Streaming Data in the Industrial Internet of Things (IIoT) Networks," Discover Appl. Sci., vol. 6, p. 483, Sept. 2024.

[9] F. Moomtaheen et al., "Extended Isolation Forest for Intrusion Detection in Zeek Data," Information, vol. 15, no. 7, p. 404, 2024.

[10] S. A. Hussein and S. R. Répás, "Enhancing Network Security through Machine Learning-Based Anomaly Detection Systems," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 21S, 2024.

[11] S. Dev and A. D. Jurcut, "Network Anomaly Detection Using LSTM Based Autoencoder," Proc. 16th ACM Symp. QoS & Security Wireless Mobile Netw., 2020, pp. 37-45.

[12] H. Huang et al., "Deep Learning Advancements in Anomaly Detection: A Comprehensive Survey," arXiv preprint arXiv:2503.13195, 2025.

[13] H. Rhachi et al., "Enhanced Anomaly Detection in IoT Networks Using Deep Autoencoders with Feature Selection Techniques," Sensors, vol. 25, no. 10, p. 3150, 2025.

[14] E. Krzysztoń et al., "A Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study," Appl. Sci., vol. 14, no. 24, p. 11545, 2024.

[15] Y. Feng et al., "An Improved X-means and Isolation Forest Based Methodology for Network Traffic Anomaly Detection," PLoS ONE, vol. 17, no. 1, Jan. 2022, Art. no. e0263423

[16] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 52, no. 1, pp. 1 - 38, Feb. 2019."

[17] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," in Proc. 2021 Int. Conf. Computer and Information Sciences (ICCIS), 2021, pp. 1-6

[18] G. Kim, S. Lee, and S. Kim "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," Expert Syst. Appl., vol 41, no. 4, pp. 1690-1700, 2018.

[19] S. Khan and A. Mailewa, 'Predicting anomalies in computer networks using autoencoder-based representation learning', *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 13, p. 9, 04 2024.

[20] S. Hore, Q. H. Nguyen, Y. Xu, A. Shah, N. D. Bastian, and T. Le, "Empirical evaluation of autoencoder models for anomaly detection in packet-based NIDS," in Proc. IEEE Conf. Dependable and Secure Computing (DSC), Nov. 2023, pp. 1–8.

[21] T. P. Nguyen, J. Cho, and D. Kim, "Semi-supervised intrusion detection system for in-vehicle networks based on variational autoencoder and adversarial reinforcement learning," Knowledge-Based Systems, vol. 304, p. 112563, 2024.

[22] R. Sharma and M. Grover, "Enhancing Cybersecurity with Machine Learning: Evaluating the Efficacy of Isolation Forests and Autoencoders in Anomaly Detection," vol. 11, pp. 1017–1021, Aug. 2024, doi: 10.1109/iccpct61902.2024.10673338.