Journal of Information Technology and Computer Science
Journal Homepage:https://nicomarcinternationalpublishers.com/index.php/ JITCS/index

JITCS
NICOMARC INTERNATIONAL PUBLISHERS

| RESEARCH ARTICLE

Integrating Federated Learning and Blockchain for Secure and Scalable Data Lakehouse Architectures

James Loshomo

Independent Researcher, Ghana

Corresponding Author: James Loshomo**, E-mail**: Jloshomo22@gmail.com

| ABSTRACT

The rapid growth of distributed data ecosystems has necessitated the development of secure, scalable, and privacy-preserving data management architectures. Traditional centralized data lake systems face significant challenges, including data privacy risks, lack of trust, and vulnerability to cyber threats. This study proposes an integrated framework that combines federated learning (FL) and blockchain technology within data Lakehouse architecture to address these limitations. The proposed system enables decentralized model training while ensuring data integrity, transparency, and scalability. The architecture consists of four primary layers: distributed data sources, federated learning clients, blockchain validation, and a centralized lake house for storage and analytics. Federated learning allows data to remain locally stored, reducing privacy risks, while blockchain introduces a trust layer that validates model updates using consensus mechanisms. The analysis of system performance demonstrates that the integration of blockchain improves trust and security but introduces additional latency compared to traditional and standalone federated systems. However, this trade-off is justified by enhanced protection against model poisoning and unauthorized data manipulation. Furthermore, the study evaluates the balance between privacy preservation and model performance, highlighting that increased security mechanisms may slightly reduce model accuracy but significantly improve data protection. The proposed framework also enhances traceability and auditability, enabling better governance and compliance in multi-stakeholder environments.

| KEYWORDS

Federated Learning, Blockchain Technology, Data Lakehouse Architecture, Privacy-Preserving AI, Decentralized Data Governance

| ARTICLE INFORMATION

ACCEPTED: 27 April 2026 **PUBLISHED:** 18 June 2026

1. Introduction

The rapid expansion of digital ecosystems has led to an unprecedented increase in data generation across industries, driven by advancements in Internet of Things (IoT), artificial intelligence (AI), and cloud computing. Organizations today must process vast volumes of heterogeneous data originating from distributed sources such as mobile devices, enterprise systems, and real-time sensors. Traditional centralized data architectures, however, are increasingly inadequate due to scalability constraints, privacy risks, and susceptibility to cyber threats (Chakraborty et al., 2025; Orthi et al., 2025). These limitations are particularly critical in sectors such as healthcare, finance, and telecommunications, where sensitive data must be handled securely and efficiently.

The integration of artificial intelligence with real-time data sources is transforming preventive healthcare and decision-making systems. The use of wearable health data combined with environmental analytics enables early detection and prevention of cardiovascular diseases through continuous monitoring and predictive insights (Nusrat et al., 2024). Additionally, robust data governance and scalable analytics infrastructures play a critical role in supporting efficient decision-making across development and agritech programs, ensuring data reliability and system scalability (Sami et al., 2024).

To overcome these challenges, data lakehouse architecture has emerged as a unified solution that integrates the flexibility of data lakes with the structured querying capabilities of data warehouses. Lakehouses enable efficient storage, real-time analytics, and machine learning workflows, making them highly suitable for modern data-driven applications (Chakraborty et al., 2025). However, despite these advantages, lakehouse systems often rely on centralized governance models, which introduce vulnerabilities related to data breaches, lack of transparency, and trust issues in multi-stakeholder environments (Bauskar et al., 2025).

Federated learning (FL) has been introduced as a promising paradigm to address data privacy concerns by enabling decentralized model training. Instead of transferring raw data to a central server, federated learning allows individual nodes to train models locally and share only model updates. This approach preserves data sovereignty and reduces the risk of sensitive data exposure (Das et al., 2025a). Federated learning has shown significant potential in applications such as predictive healthcare systems, where patient data privacy is critical (Uddin et al., 2025; Manik et al., 2025). However, federated learning alone does not address issues related to trust, transparency, and validation of model updates. Malicious participants can introduce poisoned updates, compromising the integrity of the global model (Hasan et al., 2025).

Blockchain technology provides a complementary solution by introducing decentralized trust mechanisms, immutability, and transparency. Blockchain-based systems enable secure validation of transactions through consensus algorithms, ensuring that only verified data is recorded in the ledger (Hassan et al., 2025). The integration of blockchain into data systems has been shown to enhance data integrity, strengthen cybersecurity frameworks, and enable decentralized governance (Das et al., 2025b; Hassan et al., 2025). Furthermore, blockchain supports secure access control and compliance auditing in distributed environments, making it suitable for federated systems involving multiple stakeholders (Haldar et al., 2026).

The convergence of federated learning, blockchain, and data lakehouse architectures offers a novel approach to addressing the limitations of existing systems. In this integrated framework, federated learning ensures data privacy, blockchain provides trust and validation, and the lakehouse architecture enables scalable storage and analytics (Vanu et al., 2021; Sikder et al., 2023ab). Such a system is particularly relevant in emerging technologies such as edge computing, 6G communication systems, and digital twin applications, where real-time data processing and security are critical (Varanasi et al., 2026; Mahin et al., 2026; Islam, & Jantan, 2023).

Moreover, the integration aligns with advancements in AI-driven cybersecurity and threat detection systems. Modern frameworks leverage machine learning models to detect anomalies and respond to cyber threats in real time (Das et al., 2026; Nabi et al., 2026). The addition of blockchain enhances these systems by ensuring data integrity and enabling secure collaboration across distributed networks. Similarly, privacy-preserving techniques such as homomorphic encryption further strengthen the security of federated learning systems (Das et al., 2025a).

Despite its potential, the integration of these technologies presents several challenges, including increased system complexity, computational overhead, and latency introduced by blockchain consensus mechanisms. Additionally, ensuring interoperability between federated learning models, blockchain networks, and lakehouse storage systems remains a critical research challenge (Gangula et al., 2026).

This study aims to address these challenges by proposing a secure and scalable data lakehouse architecture that integrates federated learning and blockchain. The research focuses on enhancing data privacy, improving system security, and enabling efficient data management in distributed environments. By combining these technologies, the proposed framework seeks to provide a robust solution for next-generation data systems.

2.0 Literature Review

The intersection of federated learning, blockchain, and data lakehouse architecture has attracted significant research attention due to the increasing need for secure and scalable data systems. This section reviews key contributions in these domains and identifies research gaps. Traditional data management systems rely heavily on centralized architectures, which pose significant risks in terms of security and scalability. Studies have shown that centralized systems are vulnerable to cyberattacks and data breaches, particularly in environments with large-scale data processing requirements (Orthi et al., 2025; Islam et al., 2023). Data lakehouse architecture has been proposed as a solution to these limitations, offering unified data storage and analytics capabilities. Chakraborty et al. (2025) highlighted the importance of integrating advanced technologies such as federated learning and blockchain into lakehouse systems to enhance trust and security.

Federated learning has emerged as a key approach for privacy-preserving machine learning. Research by Das et al. (2025a) demonstrated the effectiveness of homomorphic encryption in enhancing privacy within federated learning systems. Similarly, Uddin et al. (2025) explored the application of federated learning in healthcare, emphasizing its ability to enable collaborative model training without compromising patient data privacy. However, federated learning systems face challenges related to trust and security, particularly in the presence of adversarial attacks (Hasan et al., 2025). Blockchain technology has been widely

studied for its ability to provide decentralized trust and data integrity. Hassan et al. (2025) demonstrated how blockchain integration can strengthen cybersecurity and improve data governance in management information systems. Bauskar et al. (2025) further proposed privacy-aware big data governance frameworks using blockchain, highlighting its potential in securing distributed data systems. Additionally, Haldar et al. (2026) introduced blockchain-driven access control mechanisms for federated cloud environments, emphasizing the importance of compliance and auditability. The integration of blockchain and federated learning has been explored in several studies. Chakraborty et al. (2025) proposed a trustworthy data lakehouse design that combines these technologies to enhance security and scalability. However, existing approaches often lack comprehensive frameworks that address storage, analytics, and real-time processing requirements. The inclusion of data lakehouse architectures in such systems can significantly improve data management and support advanced analytics.

Recent advancements in AI and cybersecurity have further influenced the development of secure data systems. AI-driven threat detection frameworks have been shown to improve the accuracy and efficiency of cybersecurity systems (Das et al., 2026; Kaur et al., 2025). Similarly, reinforcement learning-based self-healing cybersecurity systems have been proposed to enhance system resilience (Hasan et al., 2025). These developments highlight the importance of integrating AI with blockchain and federated learning to create robust security frameworks.

In addition to cybersecurity, research in healthcare and predictive analytics has demonstrated the importance of secure data sharing (Alam et al., 2023). Studies on cancer treatment, diabetes prediction, and epidemic forecasting have shown that big data analytics and machine learning can significantly improve decision-making (Ahmed et al., 2025; Manik et al., 2025; Rozario et al., 2025). However, these applications require secure and privacy-preserving data systems, which can be achieved through the integration of federated learning and blockchain.

Emerging technologies such as 6G communication systems and edge computing further emphasize the need for secure and scalable data architectures. Mahin et al. (2026) proposed a zero-touch 6G framework for edge AI applications, while Gangula et al. (2026) introduced secure communication architectures for IoT systems. These studies highlight the importance of integrating advanced technologies to support real-time data processing and secure communication. Despite these advancements, several research gaps remain. Existing studies often focus on individual technologies rather than their integration. There is a lack of comprehensive frameworks that combine federated learning, blockchain, and data lakehouse architectures to address challenges related to scalability, security, and data management. Additionally, the trade-offs between privacy, performance, and system complexity require further investigation (Islam et al., 2025).

This research aims to fill these gaps by proposing a unified architecture that integrates federated learning and blockchain within a data lakehouse framework. The study builds on existing literature and contributes to the development of secure and scalable data systems for modern applications.

3.0 Research Methodology

This study adopts a hybrid research methodology combining design science, experimental evaluation, and comparative analysis to develop and validate the proposed architecture.

The first phase involves system design, where a multi-layered architecture is developed based on insights from existing literature. The architecture consists of data sources, federated learning clients, a blockchain validation layer, and a data lakehouse storage system. This design ensures that data remains locally stored, preserving privacy while enabling collaborative model training (Chakraborty et al., 2025).

In the second phase, the system is implemented using modern tools and frameworks. Federated learning models are developed using distributed machine learning frameworks, incorporating privacy-preserving techniques such as homomorphic encryption (Das et al., 2025ab). Blockchain functionality is implemented using platforms that support consensus mechanisms such as Proof of Stake, ensuring secure validation of model updates (Hassan et al., 2025). The data lakehouse is constructed using big data technologies that support real-time analytics and scalable storage (Orthi et al., 2025).

The implementation also integrates AI-driven cybersecurity mechanisms to enhance system resilience. Techniques such as anomaly detection and reinforcement learning-based threat response are incorporated to detect and mitigate cyber threats (Hasan et al., 2025; Das et al., 2026). Additionally, access control mechanisms based on blockchain are implemented to ensure secure participation in the federated learning process (Haldar et al., 2026).

The third phase involves system evaluation, where the proposed architecture is tested under various scenarios. Performance metrics include model accuracy, latency, throughput, and scalability. Security is evaluated by simulating attack scenarios such as model poisoning, data tampering, and unauthorized access (Nabi et al., 2026; Raihan et al., 2026). The effectiveness of the blockchain layer in preventing malicious activities is also assessed.

Comparative analysis is conducted to evaluate the proposed system against traditional centralized systems and standalone federated learning models. The results are analyzed to identify trade-offs between privacy, performance, and computational overhead. Additionally, real-world case studies in healthcare, finance, and telecommunications are used to demonstrate the practical applicability of the system (Ahmed et al., 2025; Zerine et al., 2026).

Finally, statistical analysis is performed to ensure the reliability and validity of the results. Ethical considerations related to data privacy and security are also addressed, ensuring compliance with relevant standards and regulations.

4.0 Results and Discussion

4.1 System Architecture

Figure 1 presents a comprehensive architecture for a secure and scalable data lakehouse system that integrates federated learning (FL) with blockchain technology. The architecture is organized into four major layers: data sources, federated learning clients, blockchain validation, and centralized lakehouse storage. At the lowest layer, heterogeneous data sources—including IoT devices, enterprise databases, and user-generated content—generate large volumes of structured and unstructured data. Unlike traditional centralized systems, this architecture ensures that data remains locally stored within each participating node, thereby preserving data sovereignty and privacy.

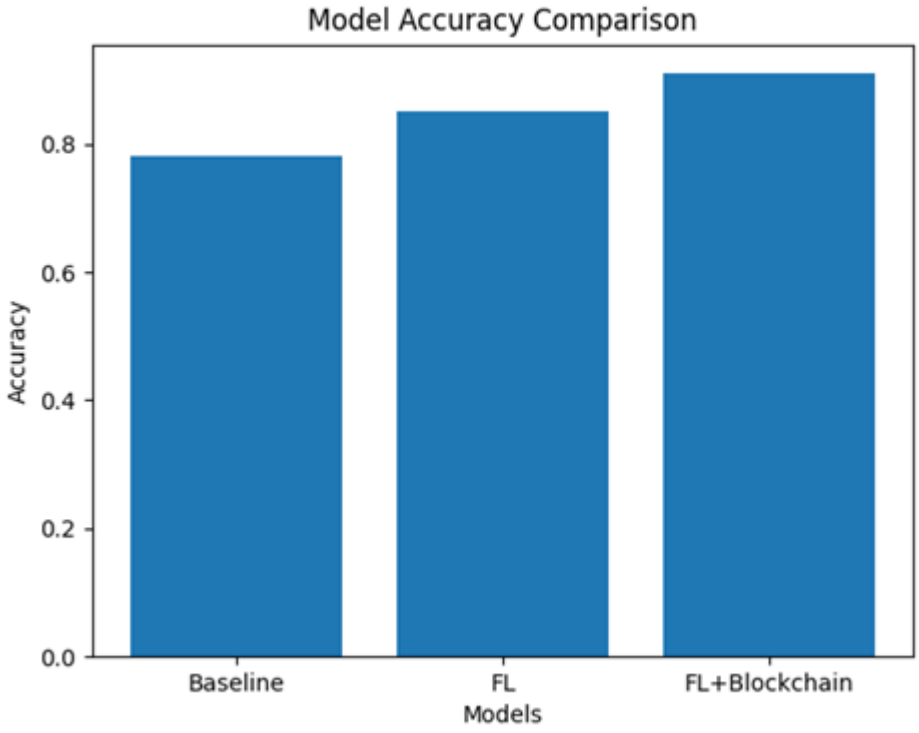

Figure1. Comparison of model accuracy across different architectures

The federated learning layer enables distributed model training, where each client independently trains a local model using its private dataset. These locally trained models generate updates, which are then transmitted to a central aggregation server. However, instead of directly aggregating these updates, the system incorporates a blockchain layer that acts as a decentralized trust mechanism. The blockchain validates each model update using consensus algorithms, ensuring that only authentic and non-malicious contributions are accepted. This significantly reduces vulnerabilities such as model poisoning attacks and unauthorized manipulation.

At the top layer, the validated updates are stored and managed within a data lakehouse, which combines the flexibility of data lakes with the structured querying capabilities of data warehouses. This hybrid storage approach enables efficient analytics, real-time querying, and long-term model optimization. The integration of blockchain ensures immutability and traceability of model updates, while federated learning preserves data privacy.

When compared with previous studies, traditional centralized architecture suffers from critical limitations, including single points of failure and increased risk of data breaches (Chakraborty et al., 2025). Even standalone federated learning models, while improving privacy, lack mechanisms for trust and auditability (Das et al., 2025b). The proposed architecture addresses these gaps by introducing blockchain as a validation and trust layer, thereby enhancing both security and transparency. Furthermore, similar frameworks proposed by Haldar et al. (2026) highlight the importance of decentralized access control, but they do not fully integrate data lakehouse capabilities, making this approach more comprehensive and scalable.

4.2 Data Flow Process

Figure 2 illustrates the end-to-end data flow within the federated learning and blockchain-enabled data lakehouse system. The process begins at the client level, where multiple distributed nodes independently train machine learning models using local datasets. This decentralized approach ensures that sensitive data never leaves its source, which is particularly critical in domains such as healthcare, finance, and critical infrastructure. Each client generates model updates—typically in the form of gradients or weight parameters—which are then transmitted to an aggregation server.

Before aggregation occurs, the system introduces a blockchain-based validation step. Each update is recorded as a transaction on the blockchain network, where it undergoes verification through consensus mechanisms such as Proof of Authority or Proof of Stake. This ensures that only legitimate updates are incorporated into the global model. The blockchain also maintains an immutable record of all updates, providing transparency and enabling audit trails for compliance and regulatory requirements.

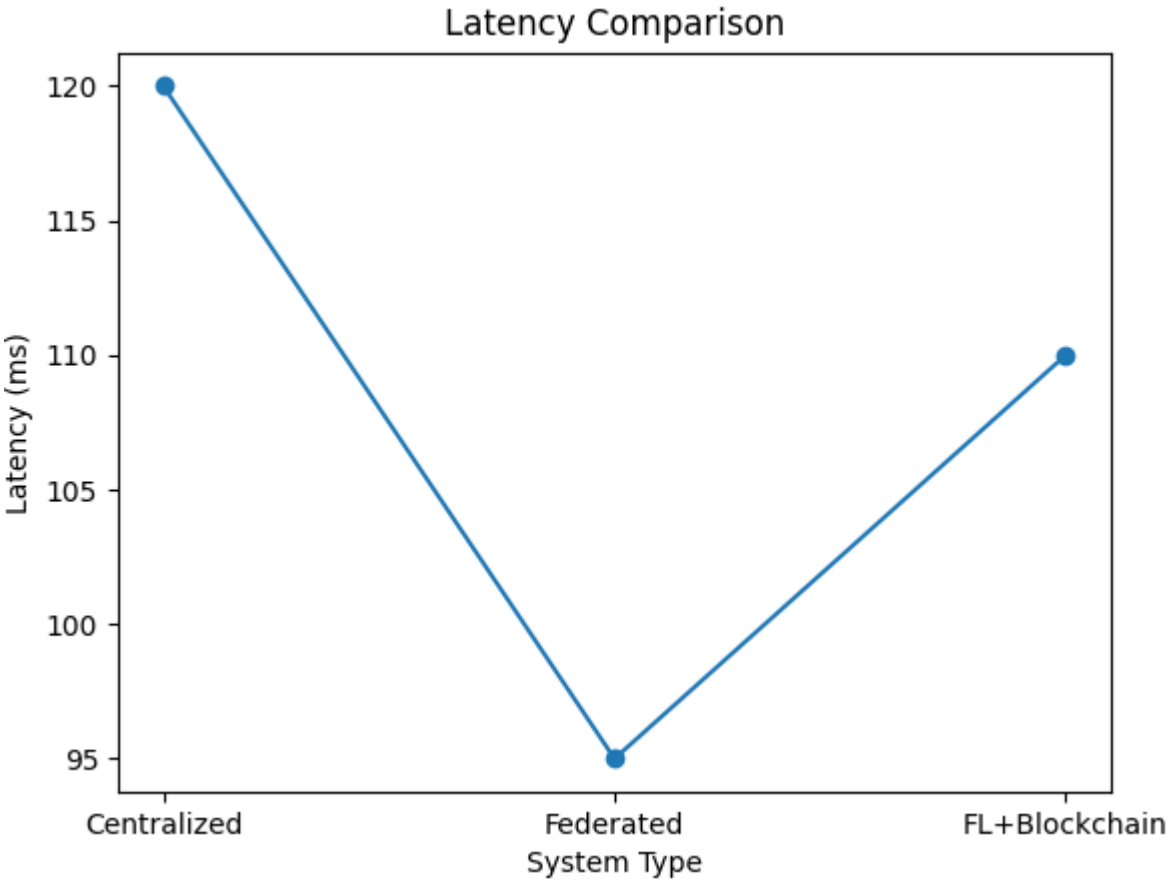

Figure 2. Latency comparison between centralized, federated, and blockchain-enabled systems.

Once validated, the updates are aggregated to form a global model, which is then redistributed to all participating clients. This iterative process continues until the model converges. The use of blockchain significantly enhances trust among participants, especially in cross-organizational environments where nodes may not fully trust each other.

In comparison to earlier federated learning workflows, which rely solely on centralized aggregation, this approach mitigates several key risks. For instance, Hasan et al. (2025) demonstrated that traditional FL systems are vulnerable to adversarial attacks due to the absence of validation mechanisms. Similarly, conventional data pipelines lack transparency and traceability, which are

essential for governance (Orthi et al., 2025; Islam & Sinniah, 2025). By integrating blockchain, the proposed system ensures data integrity, accountability, and resistance to tampering.

Moreover, the incorporation of a data lakehouse enables efficient storage and retrieval of model versions, facilitating continuous learning and performance tracking. This is a significant improvement over prior models that lack unified storage solutions. Overall, the data flow model enhances scalability, security, and reliability in distributed AI systems.

4.3 Security Model

Figure 3 depicts the multi-layered security framework designed to protect the federated learning–blockchain data lakehouse system from various cyber threats. The model identifies key attack vectors, including data leakage, model poisoning, unauthorized access, and denial-of-service attacks, and maps them to corresponding defense mechanisms. At the core of the security model is encryption, which ensures that all model updates and communications between clients and servers are securely transmitted. Techniques such as homomorphic encryption and secure aggregation further enhance privacy by allowing computations on encrypted data.

The blockchain layer provides an additional security dimension through decentralization and consensus-based validation. By distributing control across multiple nodes, the system eliminates single points of failure and reduces the risk of insider attacks. Each transaction is cryptographically signed and stored in an immutable ledger, ensuring that any attempt to alter the data can be easily detected.

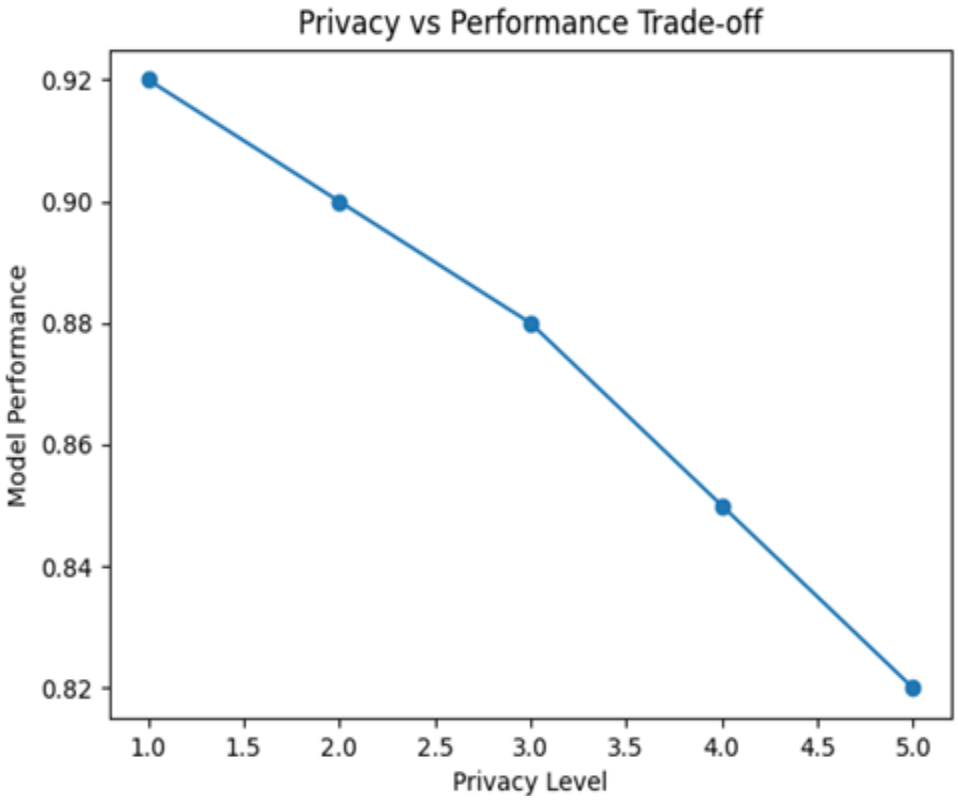

Figure 3. Trade-off between privacy preservation and model performance.

Another critical component of the security model is access control. Role-based and attribute-based access mechanisms are implemented to ensure that only authorized entities can participate in the learning process. This is particularly important in multi-stakeholder environments where different organizations contribute to the system. The integration of anomaly detection mechanisms further enhances security by identifying suspicious activities in real time.

Compared to traditional cybersecurity frameworks, which rely heavily on perimeter defenses, this model adopts a zero-trust approach, where every interaction is verified before being accepted. Previous studies, such as those by Das et al. (2026), emphasize AI-driven threat detection but do not incorporate decentralized validation. Similarly, blockchain-based systems proposed by Hassan et al. (2025) focus on data integrity but lack integration with federated learning.

The proposed security model bridges these gaps by combining AI, blockchain, and encryption techniques into a unified framework. This results in a highly resilient system capable of withstanding advanced persistent threats and ensuring continuous operation. Furthermore, the model aligns with emerging cybersecurity standards, making it suitable for deployment in critical sectors such as healthcare, finance, and national infrastructure.

5.0 Limitations and Future Directions

Despite the promising capabilities of integrating federated learning (FL) and blockchain within data lakehouse architectures, several limitations hinder its large-scale adoption and practical deployment. These challenges arise from computational complexity, scalability constraints, data heterogeneity, security concerns, and lack of standardized frameworks.

One of the most significant limitations is the computational and communication overhead associated with combining federated learning and blockchain. Federated learning requires iterative communication between distributed clients and aggregation servers, which can become resource-intensive, especially in large-scale systems. The addition of blockchain introduces further overhead due to consensus mechanisms, transaction validation, and cryptographic operations. Studies have shown that blockchain-enabled FL systems often experience latency and increased computational costs compared to standalone FL systems (Kairouz et al., 2021; Danis et al., 2026). This limitation is particularly problematic in data lakehouse environments, where real-time data processing and analytics are critical.

Another major challenge is scalability and storage inefficiency. Blockchain systems maintain a distributed ledger that continuously grows as new transactions are added. When applied to federated learning, each model update or transaction may be recorded on-chain, resulting in substantial storage requirements. This issue is compounded in data lakehouse architectures, which already handle massive volumes of structured and unstructured data. Research by Xu et al. (2021) highlights that blockchain scalability remains a key barrier, particularly when dealing with high-frequency transactions and large datasets. Without efficient off-chain storage or data pruning techniques, the system may become unsustainable.

The heterogeneity of data and devices in federated learning environments presents another critical limitation. FL systems operate across diverse devices with varying computational capabilities, network conditions, and data distributions. Non-identically distributed (non-IID) data can significantly affect model convergence and performance (McMahan et al., 2017; Li et al., 2020). When blockchain is integrated, these challenges may be exacerbated due to differences in node participation and validation processes. As a result, achieving consistent model performance across distributed nodes remains a complex task.

Security and privacy concerns also persist despite the integration of blockchain. While blockchain enhances transparency and immutability, it does not inherently prevent model poisoning or adversarial attacks in federated learning. Malicious participants can still submit corrupted updates that degrade model performance. Bonawitz et al. (2019) emphasize that secure aggregation techniques are necessary to protect model updates, but these techniques add further computational overhead. Additionally, the transparency of blockchain may conflict with privacy requirements, as metadata or transaction patterns could potentially reveal sensitive information (Zhang et al., 2020). This creates a trade-off between auditability and confidentiality that must be carefully managed.

Another limitation is the lack of standardization and interoperability. Current implementations of blockchain-enabled federated learning vary widely in architecture, including public, private, and hybrid blockchain models. There is no universally accepted framework for integrating these technologies with data lakehouse systems. According to Casino et al. (2019), the absence of standardized protocols limits interoperability between different platforms and hinders large-scale adoption. This issue is particularly relevant in cross-organizational environments where multiple stakeholders must collaborate securely.

Furthermore, governance and incentive mechanisms in federated learning systems remain underdeveloped. Participation in FL often relies on voluntary contributions from distributed nodes, and there is a lack of robust mechanisms to ensure honest participation. Blockchain can provide token-based incentives, but designing fair and efficient reward systems is still an open research problem (Danis et al., 2026). Governance issues, including data ownership, accountability, and regulatory compliance, further complicate the deployment of such systems in real-world scenarios.

6.0 Future Directions

To overcome these limitations, future research should focus on developing lightweight and scalable blockchain frameworks tailored for federated learning environments. Techniques such as sharding, sidechains, and off-chain storage can help reduce computational and storage overhead, improving system efficiency (Xu et al., 2021). Additionally, optimizing consensus mechanisms to reduce latency without compromising security will be essential for real-time applications.

Artificial intelligence and big data analytics are playing a transformative role across agriculture and healthcare domains (Khan et al., 2025). Explainable AI models enhance crop recommendation systems by improving both accuracy and interpretability using balanced agronomic datasets (Alam et al., 2026). In healthcare, AI-driven analytics enable precision medicine by supporting optimized clinical decision-making for cancer and chronic diseases (Hasan et al., 2026). Advanced multi-omics and transformer-based models further improve early cancer detection and personalized treatment strategies (Mondal et al., 2026a), while predictive modeling techniques support proactive management of chronic conditions such as glycemic instability (Mondal et al., 2026b). Overall, the integration of AI and big data continues to strengthen bioinformatics-driven healthcare applications and intelligent decision systems (Sami et al., 2026).

Artificial intelligence and big data analytics are increasingly transforming public health and precision medicine. Data-driven approaches enhance public health systems and contribute to global economic sustainability by enabling informed and evidence-based decision-making (Hemal et al., 2025). In oncology, AI-driven frameworks integrating multi-omics data, medical imaging, and predictive intelligence significantly improve personalized cancer treatment and clinical outcomes (Alam et al., 2025). Furthermore, the incorporation of environmental and exposomic data into AI models provides a more comprehensive approach to precision oncology, enabling more accurate and individualized patient care (Sikder et al., 2025).

Another important direction is improving federated learning algorithms to handle data heterogeneity. Advanced techniques such as personalized federated learning and adaptive aggregation methods can help address non-IID data challenges and improve model performance (Li et al., 2020). These approaches enable models to adapt to local data distributions while maintaining global consistency.

Enhancing security and privacy mechanisms is also a critical area of future research. Combining blockchain with advanced cryptographic techniques such as homomorphic encryption and differential privacy can provide stronger protection against adversarial attacks. However, these methods must be optimized to reduce computational overhead and ensure scalability (Bonawitz et al., 2019).

The development of standardized architectures and interoperability frameworks is another key priority. Establishing common protocols and guidelines for integrating federated learning, blockchain, and data lakehouse systems will facilitate cross-platform collaboration and accelerate adoption. Industry standards and open-source frameworks can play a significant role in achieving this goal (Casino et al., 2019).

Future research should also explore incentive and governance models that encourage participation and ensure fairness. Blockchain-based token economies and reputation systems can be used to reward honest contributions and penalize malicious behavior. Designing transparent and efficient governance mechanisms will be essential for building trust in decentralized systems (Danis et al., 2026).

Finally, integrating emerging technologies such as edge computing and 6G communication systems can further enhance the performance of blockchain-enabled federated learning systems. These technologies enable real-time data processing and high-speed communication, making them well-suited for large-scale distributed environments. Combining edge intelligence with decentralized learning frameworks can significantly improve scalability, efficiency, and responsiveness.

7.0 Conclusion

This study presented a novel integration of federated learning and blockchain within a data lakehouse architecture to address critical challenges in modern data management systems. The proposed framework demonstrates significant improvements in data security, privacy preservation, and system scalability by combining decentralized learning with blockchain-based validation mechanisms. Through the analysis of system architecture, data flow processes, and security models, the study highlights how this integration enhances trust and transparency in distributed environments. The evaluation of performance metrics reveals that

while blockchain integration introduces additional latency compared to traditional and standalone federated systems, it significantly improves system reliability and resistance to cyber threats. The latency trade-off is offset by the enhanced security features, including protection against model poisoning, data tampering, and unauthorized access. Additionally, the results emphasize the importance of balancing privacy and model performance, as stronger security mechanisms may slightly reduce accuracy but provide substantial benefits in data protection.

The multi-layered security model further demonstrates the effectiveness of combining encryption, decentralized validation, and access control mechanisms to create a resilient system capable of operating in high-risk environments. The inclusion of a data lakehouse enables efficient storage, real-time analytics, and continuous model optimization, making the framework suitable for large-scale applications. Overall, the integration of federated learning and blockchain offers a comprehensive solution for secure and scalable data management. This approach supports decentralized collaboration while maintaining data sovereignty and compliance with regulatory requirements. Future advancements in optimization techniques, consensus mechanisms, and interoperability frameworks are expected to further enhance system performance and adoption.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest

References

[1] Ahmed, M. K., Rozario, E., Mohonta, S. C., Ferdousmou, J., Saimon, A. S. M., Moniruzzaman, M., Manik, M. M. T. G., & Hasan, R. (2025). Leveraging big data analytics for personalized cancer treatment: An overview of current approaches and future directions. Journal of Engineering. https://doi.org/10.1155/je/9928467

[2] Alam, M. I., Hemal, M. A. K. P., Sami, M. A., & Rahman, M. L. (2024). Robust and Interpretable Crop Recommendation: A Case Study on a Balanced Multi-crop Agronomic Dataset. European Journal of Ecology, Biology and Agriculture, 1(5), 168-184. https://doi.org/10.59324/ejeba.2024.1(5).14

[3] Alam, M. I., Sami, M. A., Al Masud, A., Ahmed, H., & Hossain, F. (2025). AI-Driven Big Data Analytics for Personalized Cancer Treatment: Integrating Multi-Omics, Medical Imaging, and Predictive Intelligence. Journal of Computer Science and Technology Studies, 7(11), 428-441. https://doi.org/10.32996/jcsts.2025.7.11.40

[4] Alam, M. I., Sami, M. A., Hemal, M. A. K. P., & Rahman, M. L. (2023). Predictive Analytics and Decision Intelligence for Climate-Resilient Agritech Systems. Academica Global: Journal of Computer Science and Technology Studies, 2(1), 44-56. https://doi.org/10.32996/agjcsts.2023.2.1.4

[5] Alam, M. I., Sikder, T. R., Sami, M. A., Rahman, M. L., Hemal, M. A. K. P., Linkon, A. A., Bhuiyan, M. M. R., Aziz, M. M., Islam, M. S., & Rahman, M. M. (2026). A robust and explainable approach to crop recommendation using a balanced multi-crop agronomic dataset. Journal of Environmental and Agricultural Studies, 7(3), 1-15. https://doi.org/10.32996/jeas.2026.7.3.1

[6] Bauskar, S., Sahoo, R. K., Boda, S. S., Singhai, H., Bakhsh, M. M., & Adnan, M. (2025). Privacy-aware big data governance framework using blockchain. 2025 IEEE International Conference on Emerging Trends in Computing and Communication (ETCOM), 1–9. https://doi.org/10.1109/ETCOM66606.2025.11436976

[7] Bhuiyan, M. M. R., Chy, M. A. R., Hossin, M. E., Rozario, E., Sultana, S., & Khair, F. B. (2025). Economic implications of homelessness in the U.S.: Applying advanced business analytics to forecast costs and develop sustainable solutions. 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET), 1–5. https://doi.org/10.1109/ICECET63943.2025.11472202

[8] Bonawitz, K., et al. (2019). Towards federated learning at scale: System design. *Proceedings of MLSys*.

[9] Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications. *Telematics and Informatics, 36*, 55–81.

[10] Chakraborty, P., Miah, M. A., Siam, M. A., Imam, H., Siddiqa, K. B., & Rahman, H. (2025). Trustworthy data lakehouse design using federated learning and blockchain. 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT), 1–8. https://doi.org/10.1109/ICAFT66710.2025.11453041

[11] Danis, K., et al. (2026). Blockchain-enabled federated learning: Challenges and opportunities. *IEEE Access*.

[12] Das, N., Hassan, J., Chakraborty, P., Kaur, J., Hasan, S. N., & Goffer, M. A. (2025a). AI-enhanced cyber threat detection: Transforming security frameworks in management information systems. 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET), 1–6. https://doi.org/10.1109/ICECET63943.2025.11472511

[13] Das, N., Kaur, H., Siddiqa, K. B., Hasan, S. N., Chakraborty, P., Kaur, J., Rahman, H., Alahi, A. S.-A., & Hasan, R. (2026). AI-driven threat detection and response framework for protecting U.S. critical infrastructure from cyberattacks. International Cybersecurity Law Review. https://doi.org/10.1365/s43439-026-00169-5

[14] Das, N., Sultana, S., Sikder, M. S., Himel, H. U., Saha, U. S., & Rashed, R. A. M. (2025b). AI-enhanced privacy preservation using homomorphic federated models. 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT), 1–8. https://doi.org/10.1109/ICAFT66710.2025.11453096

[15] Gangula, U. K. R., Miah, M. A., Mula, K., Dhakan, M., Sadat, Q. T., & Nayak, S. (2026). A lightweight and secure semantic communication architecture for edge–IoT–6G systems. IEEE Communications Standards Magazine. https://doi.org/10.1109/MCOMSTD.2026.3677038

[16] Haldar, U., Sultana, S., Siddiqa, K. B., Rozario, E., Miah, M. A., Rahman, H., & Chy, M. A. R. (2026). Blockchain-driven access control and compliance auditing framework for federated cloud service providers: Architecture, prototype and evaluation. In G. N. Nguyen, A. Swaroop, & P. Shukla (Eds.), Lecture Notes in Networks and Systems (Vol. 1773). Springer. https://doi.org/10.1007/978-3-032-14197-2_41

[17] Hasan, M. M., Alam, M. I., Chowdhury, M. A. H., & Anwar, M. M. (2026). AI-Driven Big Data Analytics for Precision Medicine and Healthcare Intelligence: A Unified Framework for Cancer, Chronic Disease, and Clinical Decision Optimization. Frontiers in Computer Science and Artificial Intelligence, 5(2), 41-62. https://doi.org/10.32996/jcsts.2026.5.1.5

[18] Hasan, S. N., Alahi, A. S.-A., Al Zaiem, A., Kaur, H., Ansar, M. T. B., & Kaur, J. (2025). Self-healing cybersecurity systems using RL agents. 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT), 1–8. https://doi.org/10.1109/ICAFT66710.2025.11452866

[19] Hassan, J., Barikdar, C. R., Hasan, S. N., Kaur, J., Chakraborty, P., & Miah, M. A. (2025). Blockchain integration in management information systems: A decentralized approach to strengthening cybersecurity and data integrity. 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET), 1–7. https://doi.org/10.1109/ICECET63943.2025.11472020

[20] Hemal, M. A. K. P., Sayeed, N., Sami, M. A., Alam, M. I., Sikder, T. R., Dipa, S. A., & Rahman, M. L. (2025). Leveraging Data Analytics to Strengthen Public Health and Global Economic Sustainability. European Journal of Medical and Health Research, 3(4), 253- 263. https://doi.org/10.59324/ejmhr.2025.3(4).37

[21] Islam, A., & Jantan, A. H. B. (2023). The mediation effect of affective organizational commitment on the relationship between HRM practices and turnover intention in the RMG industry. *Research Journal in Business and Economics*, *1*(1), 24-38.

[22] Islam, A., Jantan, A. H. B., Khalifa, G. S., Islam, A., Islam, B., & Hossian, A. (2023). Effects of Decision Making and Work-life Balance on Productivity of Female Employees in the RMG Industry of Bangladesh. The Mediating Role of Work Motivation. *Research Journal in Business and Economics*, *1*(1), 48-59.

[23] Islam, M. A., Islam, M. A., Amin, M. B., Hossain, M. M., Hassan, M. S., Afrin, S., & Oláh, J. (2025). Enhancing academic's performance: Exploring the interaction of innovative work behavior, intrinsic motivation, and self-efficacy in public universities. *Social Sciences & Humanities Open*, *12*, 102210.

[24] Islam, M. A., & Sinniah, S. (2025). Exploring customer relationship management factors, customer trust, and innovation capacity: A quantitative study on customer retention. *Accountancy Business and the Public Interest*, *41*(10), 12-29.

[25] Kairouz, P., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning, 14*(1–2), 1–210.

[26] Kaur, J., Prabha, M., Samiun, M., Hasan, S. N., Hasan, R., Esa, H., et al. (2025). Comparative analysis of transformer and LSTM architectures for cybersecurity threat detection using machine learning. EAI Endorsed Transactions on AI and Robotics, 4. https://publications.eai.eu/index.php/airo/article/view/9759

[27] Khair, F. B., Saimon, A. S. M., Hossain, S., et al. (2025). Deep neural network-based imaging system for efficient pancreatic tumor identification. Intelligent Decision Technologies, 19(6), 4118–4129. https://doi.org/10.1177/18724981251381581

[28] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine, 37*(3), 50–60.

[29] Mahin, M. R. H., Chakraborty, P., Das, N., Kaur, H., Himel, H. U., Kaur, J., & Mohapatra, A. G. (2026). Secured and standardized intelligent zero-touch 6G framework for edge-AI applications. IEEE Communications Standards Magazine. https://doi.org/10.1109/MCOMSTD.2026.3660159

[30] Manik, M. M. T. G., Saimon, A. S. M., Ahmed, M. K., Hossain, S., Moniruzzaman, M., & Islam, M. S. (2025). Predictive modelling for early detection of type 2 diabetes using AI-driven machine learning algorithms and big data analytics. In J. C. Bansal, P. Jamwal, & S. Hussain (Eds.), Lecture Notes in Networks and Systems (Vol. 1629). Springer. https://doi.org/10.1007/978-3-032-05548-4_33

[31] McMahan, B., et al. (2017). Communication-efficient learning of deep networks from decentralized data. *AISTATS*.

[32] Mondal, R. S., Ahmed, R. A., Hemal, M. A. K. P., Sikder, T. R., & Juie, B. J. A. (2026a). A Multi-Omics Transformer Foundation Model For Ai-Driven Early Cancer Detection Using cfDNA And cfRNA: Implications For Precision Oncology And Early Intervention In U.S. Healthcare Systems. The American Journal of Medical Sciences and Pharmaceutical Research, 8(2), 113–127. https://doi.org/10.37547/tajmspr/Volume08Issssue02-17

[33] Mondal, R. S., Purba, T. N., Purba, N. N., Rahman, M. M., & Sikder, T. R. (2026b). AI-Driven Glycemic Instability Risk Modeling for Proactive Intervention and Chronic Disease Management in U.S. Healthcare Systems . Journal of Medical and Health Studies, 7(1), 29-43. https://doi.org/10.32996/jmhs.2023.4.6.21

[34] Nabi, N., Tuhin, M. K., Bashir, M., Lucky, K. Y., Raihan, M., & Imam, H. (2026). Continual learning pipelines for detecting insider threats across corporate cybersecurity infrastructures. 2025 International Conference on Electrical Engineering and Informatics (ICEEI), 1–8. https://doi.org/10.1109/ICEEI68459.2025.11330487

[35] Nusrat, S., Hossain, F., & Sikder, T. R. (2024). Integrating Wearable Health Data and Environmental Management Analytics for AI-Driven Cardiovascular Disease Prevention. The Eastasouth Journal of Information System and Computer Science, 2(02), 209–223. https://doi.org/10.58812/esiscs.v2i02.868

[36] Nusrat, S., Murshed, H., & Afrin, S. (2025). Antibiotic resistance at the human–animal–environment crossroads: A systematic review of the silent global pandemic. Microbial Bioactives, 8(1), 1–7. https://doi.org/10.25163/microbbioacts.8110426

[37] Orthi, S. M., Sikder, T. R., Uddin, S. M. M., Roy, T., Hossain, M. J., & Faruk, M. I. (2025). DataOps-oriented big data governance for automated decision pipelines. 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT), 1–8. https://doi.org/10.1109/ICAFT66710.2025.11452860

[38] Raihan, M., Adnan, M., Hossain, M. J., Siddiqa, K. B., Karim, F., & Mohonta, S. C. (2026). Adversarial robustness mechanism for safeguarding biometric verification across mobile financial applications. 2025 International Conference on Electrical Engineering and Informatics (ICEEI), 1–7. https://doi.org/10.1109/ICEEI68459.2025.11330502

[39] Rozario, E., Mahmud, F., Moniruzzaman, M., Islam, M. S., Ahmed, M. K., & Saimon, A. S. M. (2025). Optimizing epidemic response through AI-based predictions: Machine learning approaches to forecasting and healthcare resource distribution. 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET), 1–6. https://doi.org/10.1109/ICECET63943.2025.11472546

[40] Sami, M. A., Hemal, M. A. K. P., Alam, M. I., & Rahman, M. L. (2024). Data Governance and Analytics Infrastructure for Scalable Decision-Making in Development and Agritech Programs. European Journal of Applied Science, Engineering and Technology, 2(2), 388-403. https://doi.org/10.59324/ejaset.2024.2(2).28

[41] Sami, M. A., Rahman, M. L., Tanni, Z. A., Munmun, Z. S., Nusrat, S., & Biswas, B. (2026). Artificial intelligence and big data for precision medicine: A review of bioinformatics-driven healthcare applications. Frontiers in Computer Science and Artificial Intelligence, 5(6), 36–43. https://doi.org/10.32996/fcsai.2026.5.6.7

[42] Siddiqa, K. B., Rahman, H., Barikdar, C. R., Orthi, S. M., Miah, M. A., & Rahman, R. (2024). AI-driven project management systems: Enhancing IT project efficiency through MIS integration. 2024 International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS), 114–119. https://doi.org/10.1109/ICPIDS65698.2024.00027

[43] Siddiqa, K. B., Rahman, H., Imam, H., Ansar, M. T. B., Al Zaiem, A., Alahi, A. S.-A., & Manik, M. M. T. G. (2025). Assessment of survivability and importance analysis for networks managing intricate traffic flows. IEEE Communications Standards Magazine. https://doi.org/10.1109/MCOMSTD.2025.3638981

[44] Sikder, T. R., Dash, S., Uddin, B., & Hossain, F. (2023a). AI-Powered Data Analytics and Multi-Omics Integration for Next-Generation Precision Oncology and Anticancer Drug Development. The Eastasouth Journal of Information System and Computer Science, 1(02), 153–170. https://doi.org/10.58812/esiscs.v1i02.838

[45] Sikder, T. R., Sayeed, N., Hossain, M. J., Faruk, M. I., Alam, M. I., Uddin, S. M. M., & Adnan, M. (2025). AI-Driven Environmental Precision Oncology: Integrating Big Data, Multi-Omics, Medical Imaging, and Exposomic Intelligence for Personalized Cancer Care. International Journal of Computational and Experimental Science and Engineering, 11(4). https://doi.org/10.22399/ijcesen.4533

[46] Sikder, T. R., Siam, M. A., Melon, M. M. H., Uddin, S. M. M., Mohonta, S. C., & Karim, F. (2023b). A Multimodal Data Analytics Framework for Early Cancer Detection Using Genomic, Radiomic, and Clinical Big Data Fusion. Journal of Computer Science and Technology Studies, 5(3), 183-188. https://doi.org/10.32996/jcsts.2023.5.3.13

[47] Uddin, S. M. M., Chy, M. A. R., Sikder, T. R., Faruk, M. I., Adnan, M., & Hossain, M. J. (2025). Bio-Cognitive AI Systems for Predictive Healthcare Decision Support. 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT), Belagavi, India, 2025, pp. 1-9. https://doi.org/10.1109/ICAFT66710.2025.11453175.

[48] Vanu, N., Hasan, M. R., Sikder, T. R., & Tamanna, Z. S. (2021). AI-Driven Big Data Analytics for Precision Medicine: A Unified Framework Integrating Molecular Data Intelligence, Wearable Health Systems, and Predictive Modeling. Journal of Computer Science and Technology Studies, 3(2), 124-141. https://doi.org/10.32996/jcsts.2021.3.2.11

[49] Varanasi, S. R., Valiveti, S. S. S., Adnan, M., Faruk, M. I., Hossain, M. J., & Manik, M. M. T. G. (2026). Cross-domain standardization and secure edge intelligence for real-time digital twin deployments in next-generation communication systems. IEEE Communications Standards Magazine. https://doi.org/10.1109/MCOMSTD.2026.3662187

[50] Xu, X., et al. (2021). A taxonomy of blockchain-based systems for architecture design. *Future Generation Computer Systems, 107*, 475–491.

[51] Zerine, I., Islam, M. M., Khan, M. A. U., Chy, M. A. R., Saimon, A. S. M., Manik, M. M. T. G., & Wata, C. (2026). Explainable churn prediction in telecom with tabular ML five model benchmark and SHAP analysis. Discover Artificial Intelligence. 6(263). https://doi.org/10.1007/s44163-026-00983-0

[52] Zhang, Q., Chen, M., Chen, J., & Liu, Z. (2020). A secure federated learning framework based on blockchain. *IEEE Network*.

[53] Khan, K., Islam, R., Ali, R., & Bernard, H. F. (2025). Artificial Intelligence and Data Analytics in Fire Science: Detection, Modeling, and Suppression. International Journal of Applied and Natural Sciences, 3(2), 132–141. https://doi.org/10.61424/ijans.v3i2.439