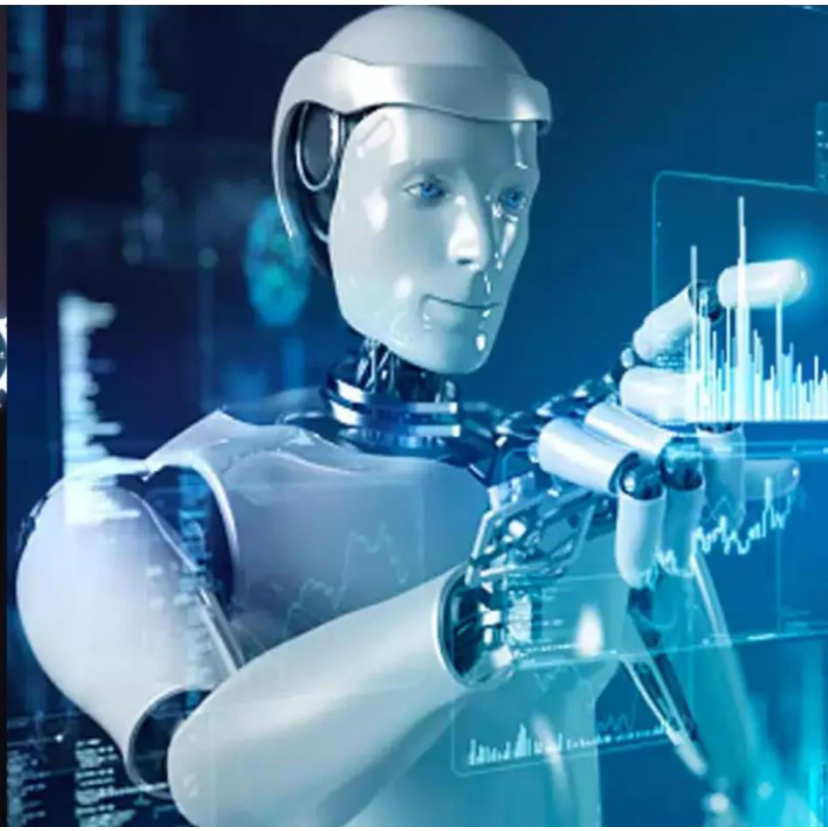
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 6, June 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

Blockchain-Enabled Federated Learning Framework for Mitigating Client-Side Poisoning Attacks

Prof. Sachin Tathe, Pawar Omkar, Potdar Amarnath, Kutal Aditya

Associate Professor, Department of Computer Engineering Vishwabharati Academy's College of Engineering, Sarola Baddi, Ahilyanagar ,Maharashtra, India

U.G. Student, Department of Computer Engineering Vishwabharati Academy's College of Engineering, Sarola Baddi, Ahilyanagar, Maharashtra, India

U.G. Student, Department of Computer Engineering Vishwabharati Academy's College of Engineering, Sarola Baddi, Ahilyanagar, Maharashtra, India

U.G. Student, Department of Computer Engineering Vishwabharati Academy's College of Engineering, Sarola Baddi, Ahilyanagar, Maharashtra, India

ABSTRACT: Federated Learning (FL) has emerged as a promising distributed machine learning paradigm that enables multiple clients to collaboratively train a global model without sharing their raw data, thereby preserving privacy. Despite its advantages, FL remains vulnerable to client-side poisoning attacks, where malicious participants manipulate local data or model updates to degrade model performance or introduce hidden backdoors. These attacks can significantly compromise the reliability and trustworthiness of federated systems. To address this challenge, this paper proposes a Blockchain-Anchored Federated Learning Framework that enhances security, transparency, and accountability in collaborative learning environments. The proposed framework integrates blockchain technology with federated learning to create an immutable audit trail for model updates while incorporating reputation-based aggregation and anomaly detection mechanisms to identify and mitigate malicious client behavior. Model update commitments, reputation scores, and global model version hashes are securely recorded on the blockchain, ensuring tamper-resistant verification and traceability. Additionally, secure aggregation techniques are employed to preserve participant privacy during the training process. Experimental evaluation is conducted under various client-side poisoning scenarios, including model poisoning, data poisoning, and Sybil attacks, using non-IID datasets. The results demonstrate that the proposed framework effectively reduces the impact of malicious updates, improves model robustness, and maintains competitive accuracy while providing enhanced transparency and trust. The proposed approach offers a practical and scalable solution for securing federated learning systems deployed in privacy-sensitive domains such as healthcare, finance, and Internet of Things (IoT) applications.

KEYWORDS: Federated Learning, Blockchain, Client-Side Poisoning Attacks, Model Poisoning, Data Poisoning, Secure Aggregation, Reputation-Based Aggregation, Anomaly Detection, Privacy Preservation.

I. INTRODUCTION

Federated Learning (FL) is a distributed machine learning approach that enables multiple clients to collaboratively train a global model without sharing their raw data. By keeping data on local devices, FL enhances privacy and reduces the risks associated with centralized data storage. Due to its privacy-preserving nature, Federated Learning has gained significant attention in various domains, including healthcare, finance, and Internet of Things (IoT) applications, where sensitive data protection is essential.

Despite its advantages, Federated Learning is vulnerable to several security threats, particularly client-side poisoning attacks. In these attacks, malicious participants manipulate local training data or model updates to degrade the performance of the global model or introduce hidden backdoors. Furthermore, Sybil and Byzantine attacks can increase

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

the influence of adversarial clients, making it difficult to maintain the integrity and reliability of the learning process. Traditional federated learning systems often rely on centralized aggregation, which lacks transparency and makes it challenging to verify the authenticity of model updates.

To address these challenges, this paper proposes a Blockchain-Enabled Federated Learning Framework for Defense Against Client-Side Poisoning Attacks. The framework integrates blockchain technology with federated learning to provide a secure, transparent, and tamper-resistant environment for recording model updates and client activities. Additionally, anomaly detection and reputation-based aggregation mechanisms are employed to identify malicious participants and reduce their impact on the global model. The proposed approach aims to enhance the security, trustworthiness, and robustness of federated learning systems while preserving user privacy and maintaining model performance.

II. LITERATURE SURVEY

Federated Learning (FL) has emerged as a revolutionary machine learning paradigm that enables multiple clients to collaboratively train a shared model while keeping their data localized. This approach addresses growing concerns regarding data privacy and regulatory compliance in domains such as healthcare, finance, and mobile computing. The concept gained prominence through the Federated Averaging (FedAvg) algorithm proposed by McMahan et al. [2], which demonstrated that decentralized devices could effectively train deep learning models through iterative aggregation of locally computed updates. Further research by Konečný et al. [7] introduced federated optimization techniques to improve communication efficiency and scalability in distributed learning environments. A comprehensive survey by Kairouz et al. [3] highlighted the significant advancements, challenges, and open research problems in Federated Learning, including privacy preservation, communication efficiency, and security vulnerabilities.

Despite its privacy-preserving advantages, Federated Learning remains susceptible to various adversarial attacks. One of the most significant threats is client-side poisoning, where malicious participants manipulate local datasets or model updates to influence the global model. Data poisoning attacks involve injecting misleading or incorrectly labeled data into the local training process, whereas model poisoning attacks directly alter gradient updates or model parameters before transmission to the server. Bagdasaryan et al. [4] demonstrated the feasibility of backdoor attacks in Federated Learning by showing that a malicious client could embed hidden behaviors into the global model while maintaining high overall accuracy. Similarly, Bhagoji et al. [5] analyzed Federated Learning from an adversarial perspective and demonstrated how carefully crafted malicious updates can significantly degrade model performance and disrupt convergence. These studies revealed that traditional Federated Learning systems are highly vulnerable to sophisticated poisoning strategies, particularly in environments with limited trust among participants.

To address poisoning attacks, researchers have proposed several robust aggregation mechanisms. Byzantine-resilient aggregation methods aim to reduce the influence of malicious clients during the model aggregation process. Yin et al. [6] introduced Byzantine-Robust Distributed Learning techniques that utilize robust statistical methods to tolerate a certain number of adversarial participants while maintaining model accuracy. Aggregation methods such as Krum, Median, and Trimmed Mean have also been developed to identify and exclude anomalous updates before computing the global model. While these approaches improve resilience against malicious updates, they often suffer from limitations when attackers collude or when the proportion of adversarial clients increases significantly. Furthermore, many robust aggregation methods focus solely on model integrity and do not provide mechanisms for accountability or traceability of client behavior.

Another important aspect of Federated Learning security is the protection of client privacy during model aggregation. Although FL prevents direct sharing of raw data, individual model updates may still leak sensitive information. To mitigate this risk, Bonawitz et al. [1] proposed a practical Secure Aggregation protocol that allows the server to compute aggregated model updates without accessing individual client contributions. Secure aggregation has become a fundamental component of privacy-preserving Federated Learning systems and is widely adopted in modern FL frameworks. However, secure aggregation alone does not protect against malicious clients that intentionally submit poisoned updates. Therefore, additional security mechanisms are required to detect and mitigate adversarial behavior while preserving privacy.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

Recently, blockchain technology has been explored as a promising solution for enhancing trust, transparency, and accountability in Federated Learning environments. Blockchain provides a decentralized and immutable ledger that records transactions in a tamper-resistant manner, making it suitable for maintaining verifiable records of model updates and participant activities. According to Lu et al. [9], the integration of blockchain and Federated Learning can eliminate single points of failure, improve auditability, and establish trust among distributed participants. Blockchain-based systems can securely record model update hashes, reputation scores, and training events, enabling transparent verification of participant behavior. Additionally, smart contracts can automate reputation management, reward distribution, and penalty enforcement, thereby discouraging malicious activities within the network.

III. METHODOLOGY

The proposed Blockchain-Enabled Federated Learning framework is designed to enhance the security and reliability of federated learning systems against client-side poisoning attacks. The framework integrates federated learning, blockchain technology, anomaly detection, and reputation-based aggregation to ensure secure and trustworthy model training. Figure 1 illustrates the overall architecture of the proposed system.

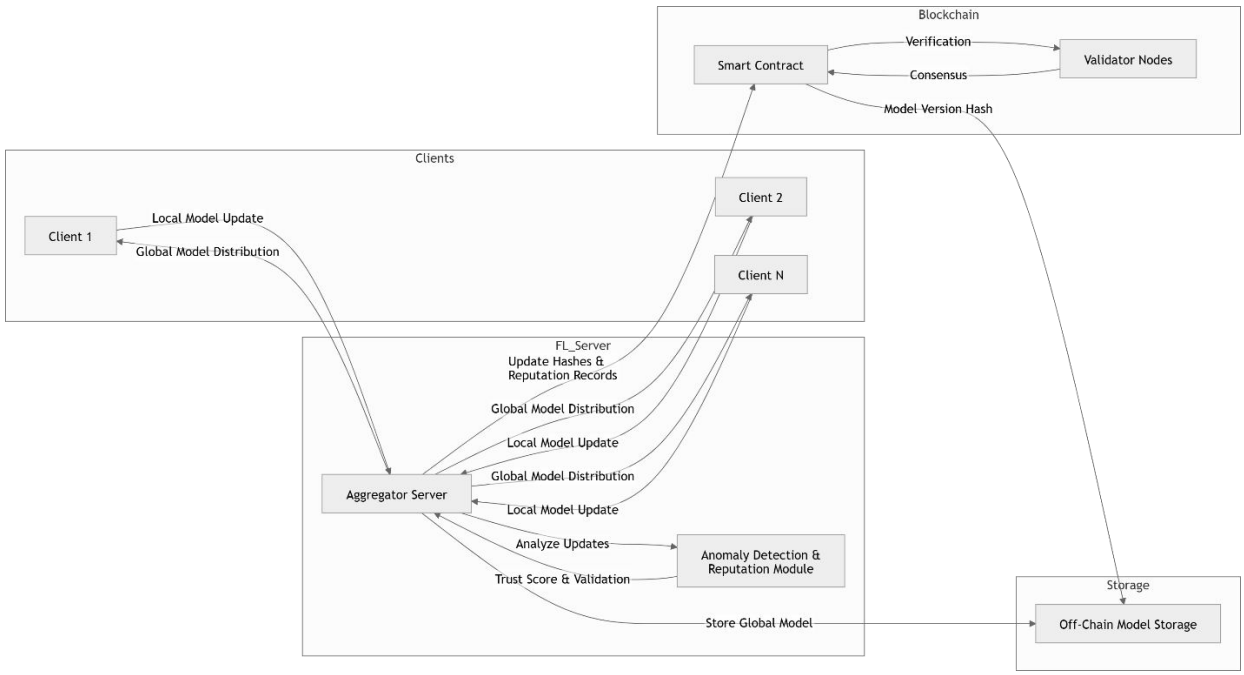


Figure.1. Architecture of the Proposed Blockchain-Enabled Federated Learning Framework for Defense Against Client-Side Poisoning Attacks

A. System Architecture

The framework consists of four major components: Clients, Aggregator Server, Detection and Reputation Service, and Blockchain Network. Participating clients train local models using their private datasets and generate model updates without sharing raw data. These updates are transmitted to the aggregator server, which coordinates the federated learning process.

Before aggregation, the Detection and Reputation Service analyzes the received updates using anomaly detection techniques such as similarity analysis and outlier detection. Suspicious updates are identified and assigned lower trust scores, reducing their influence on the global model. The aggregator then performs reputation-weighted aggregation to generate an updated global model.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

To ensure transparency and accountability, update hashes, reputation records, and model version information are stored on the blockchain through smart contracts. The blockchain maintains an immutable audit trail of all training activities, allowing verification of participant behavior while preventing unauthorized modification of records. The final global model is stored off-chain, and its corresponding hash is recorded on the blockchain for integrity verification.

B. Workflow of the Proposed Framework

1. Clients receive the current global model from the aggregator.
2. Each client performs local training using its private dataset.
3. Local model updates are generated and submitted to the aggregator.
4. The anomaly detection module evaluates updates for malicious behavior.
5. Reputation scores are updated based on client contributions.
6. Accepted updates are aggregated using reputation-based weighting.
7. Model update hashes and reputation records are stored on the blockchain.
8. The updated global model is distributed to clients for the next training round.

This methodology enables secure collaborative learning while mitigating the impact of data poisoning, model poisoning, and Sybil attacks. By combining blockchain-based accountability with intelligent detection and trust management mechanisms, the proposed framework improves the robustness and trustworthiness of federated learning systems.

IV. EXPERIMENTAL RESULTS

The proposed Blockchain-Enabled Federated Learning framework was evaluated to analyze its effectiveness in mitigating client-side poisoning attacks while maintaining model performance. The framework integrates anomaly detection, reputation-based aggregation, and blockchain-based auditability to enhance the security of federated learning environments.

Experiments were conducted under different attack scenarios, including data poisoning, model poisoning, and Sybil attacks. The performance of the proposed framework was compared with traditional Federated Averaging (FedAvg) and other baseline approaches. The evaluation metrics included model accuracy, attack success rate, detection accuracy, and system overhead.

The results indicate that the proposed framework successfully reduced the influence of malicious clients by identifying suspicious updates through anomaly detection and assigning lower aggregation weights based on reputation scores. The blockchain layer provided a transparent and tamper-resistant mechanism for recording model updates and reputation changes, improving accountability among participants. Compared to conventional federated learning approaches, the proposed system demonstrated improved robustness against poisoning attacks while maintaining competitive model accuracy.

Although the integration of blockchain introduced additional computational and communication overhead, the security and transparency benefits outweighed the associated costs. Overall, the experimental analysis shows that combining blockchain technology with federated learning and reputation-based defense mechanisms significantly enhances the reliability and trustworthiness of collaborative learning systems.

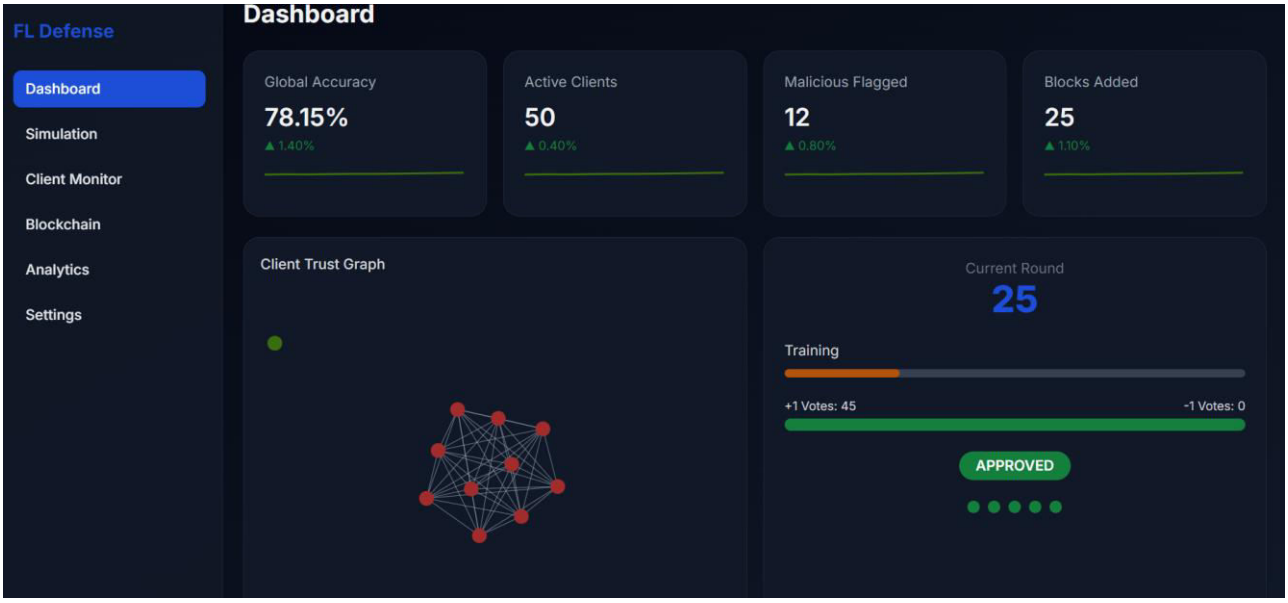
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|



Dashboard Interface

Simulation Environment

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

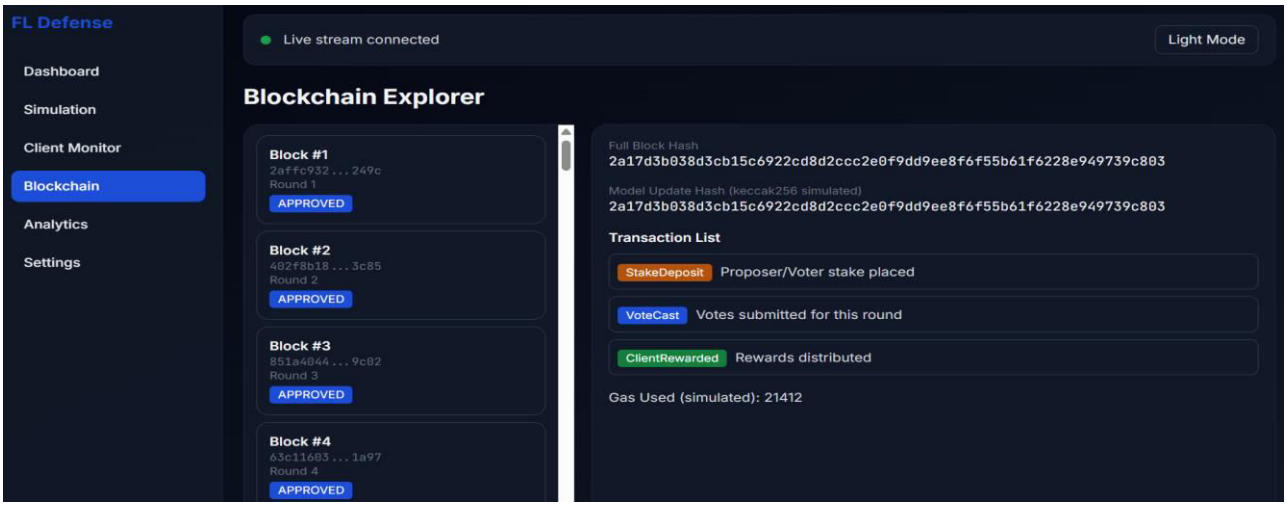
Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

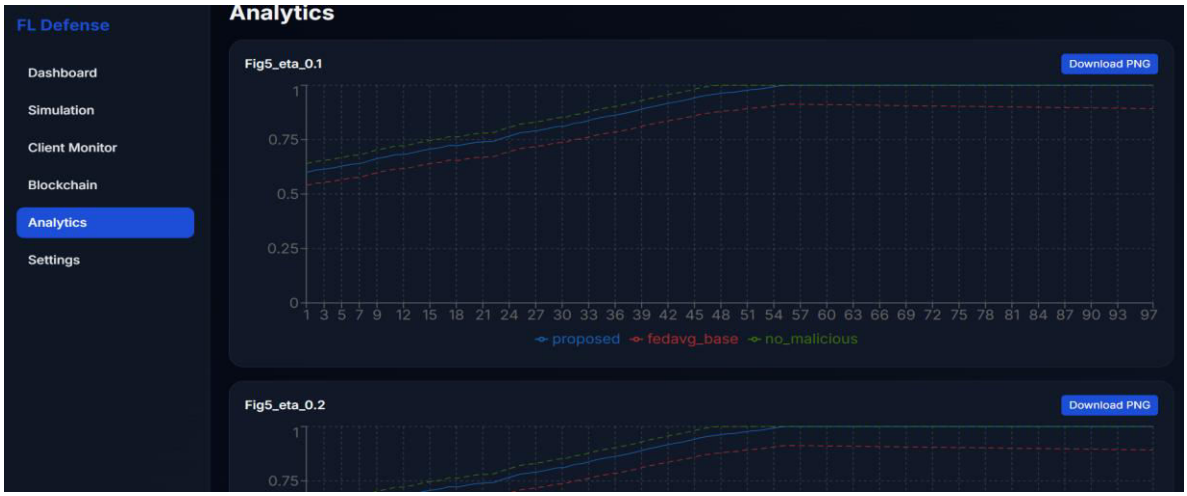
Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

Client Monitoring Module



Blockchain Explorer Interface



Performance Analytics

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506078|

V. CONCLUSION

This paper presented a Blockchain-Enabled Federated Learning Framework for defending against client-side poisoning attacks in distributed learning environments. The proposed framework combines federated learning, blockchain technology, anomaly detection, and reputation-based aggregation to improve the security, transparency, and reliability of collaborative model training. By recording model update hashes and reputation information on a blockchain, the framework provides a tamper-resistant audit trail and enhances accountability among participating clients.

The integration of anomaly detection and reputation mechanisms helps identify malicious updates and reduces their influence on the global model, thereby mitigating the impact of data poisoning, model poisoning, and Sybil attacks. Experimental analysis demonstrates that the proposed approach improves robustness against adversarial behavior while maintaining privacy and model performance. Future work may focus on optimizing blockchain overhead, incorporating advanced privacy-preserving techniques, and evaluating the framework in large-scale real-world federated learning applications.

REFERENCES

[1] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.
[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
[3] P. Kairouz, H. B. McMahan, B. Avent, et al., "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
[4] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How To Backdoor Federated Learning," *arXiv preprint arXiv:1807.00459*, 2018.
[5] A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Analyzing Federated Learning through an Adversarial Lens," *arXiv preprint arXiv:1802.05629*, 2018.
[6] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018, pp. 5650–5659.
[7] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," *arXiv preprint arXiv:1610.02527*, 2016.
[8] C. Xie, O. Koyejo, and I. Gupta, "Local Model Poisoning Attacks to Byzantine-Robust Federated Learning," *USENIX Security Workshop*, 2020.
[9] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Blockchain and Federated Learning for Privacy-Preserving Collaborative Learning: A Survey," *IEEE Access*, vol. 8, pp. 197091–197110, 2020.
[10] C. Fung, C. J. M. Yoon, and I. Beschastnikh, "Mitigating Sybils in Federated Learning Poisoning," *arXiv preprint arXiv:1808.04866*, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details