

POISONING-RESISTANT FEDERATED LEARNING IN HEALTHCARE USING BLOCKCHAIN TECHNOLOGY

¹ Mrs. G. Akila , ² D. Venu Gopal, ³ B. Gopi , ⁴ A. Naveen Balaji, ⁵ C. Vaishnavi

¹Assistant Professor in Department of CSE Sri Indu College of Engineering & Technology -Hyderabad.

^{2,3,4,5} UG Scholars in Department of CSE Sri Indu College of Engineering & Technology-Hyderabad

Abstract

The rapid digital transformation of healthcare has resulted in a massive increase in sensitive medical data generated from electronic health records, wearable devices, medical imaging, and Internet of Medical Things (IoMT) platforms. Applying Artificial Intelligence (AI) and Machine Learning (ML) to this distributed data can greatly improve disease prediction, early diagnosis, personalized treatment, and overall healthcare management. However, strict privacy regulations, data silos across institutions, and rising cybersecurity threats restrict secure data sharing. Traditional centralized machine learning methods require pooling data into a single repository, which creates significant risks related to privacy, security, and regulatory compliance. Federated Learning (FL) offers a decentralized alternative that enables multiple healthcare organizations to collaboratively train models without sharing raw data. Nevertheless, existing FL frameworks still face challenges such as malicious model updates, limited trust among participants, potential data leakage through gradient sharing, and dependence on centralized aggregation servers. To address these issues, this research introduces a secure and privacy-preserving federated learning framework that integrates blockchain technology with Secure Multi-Party Computation (SMPC). The blockchain component removes the need for a central authority by providing a distributed ledger that ensures transparency, traceability, and tamper-resistant recording of model updates. Smart contracts manage participant authentication, coordinate training rounds, and enforce collaboration policies. At the same time, SMPC protocols enable encrypted aggregation of model parameters so that individual updates remain confidential, preventing inference and data reconstruction attacks. The proposed framework is evaluated using simulated healthcare datasets and assessed based on model accuracy, training efficiency, communication overhead, and resilience against adversarial threats such as poisoning attacks. Results show that the system maintains competitive predictive performance while significantly improving privacy, trust, and security in collaborative healthcare AI environments.

Keywords (Expanded) Federated Learning (FL), Healthcare Data Security, Blockchain in HealthCare, Secure Multi-Party Computation (SMPC), Privacy-Preserving Machine Learning, Medical Artificial Intelligence

I. INTRODUCTION

The healthcare sector is undergoing rapid digital transformation driven by artificial intelligence, big data analytics, cloud computing, and Internet

of Medical Things (IoMT) technologies. Hospitals and medical research institutions continuously generate large volumes of sensitive data including electronic health records, medical imaging, genomic sequences, laboratory reports,

and wearable sensor data. Machine learning techniques can extract valuable insights from such data to enable early disease prediction, personalized treatment, and improved clinical decision-making. However, strict privacy regulations such as HIPAA and GDPR impose significant restrictions on sharing patient information across institutional boundaries, leading to fragmented data silos that limit the development of robust and generalizable healthcare AI systems [1].

Traditional centralized machine learning approaches require aggregating data from multiple hospitals into a single repository for model training. While effective in many domains, this paradigm introduces serious risks in healthcare, including increased vulnerability to cyberattacks, single points of failure, and loss of institutional control over sensitive data. Recent healthcare data breaches have demonstrated the limitations of centralized infrastructures and highlighted the need for privacy-preserving collaborative learning techniques. Ethical concerns regarding patient confidentiality and legal compliance further restrict large-scale data sharing [2].

Federated Learning (FL) has emerged as a promising decentralized paradigm that enables multiple institutions to collaboratively train machine learning models without exchanging raw data. Each participant trains a local model on private datasets and shares only model updates for aggregation. This approach reduces the need

for data transfer and improves regulatory compliance while enabling collaborative innovation across institutions. Federated learning has shown strong potential in applications such as disease prediction, medical imaging, and drug discovery, where diverse datasets significantly improve model performance [3].

Despite these advantages, federated learning faces critical challenges including reliance on trusted aggregation servers, vulnerability to model poisoning attacks, and the risk of privacy leakage through gradient sharing. Studies have demonstrated that attackers can reconstruct training data from shared gradients, raising concerns about the security of FL systems [4]. Furthermore, malicious participants can upload poisoned updates that degrade model accuracy or introduce backdoors, highlighting the need for stronger trust and verification mechanisms [5].

To address these limitations, this research proposes a secure federated learning framework that integrates blockchain technology and Secure Multi-Party Computation (SMPC). The proposed approach aims to establish decentralized trust, ensure transparency, and provide secure aggregation of model updates for privacy-preserving healthcare AI.

II . LITERATURE SURVEY

Federated learning was first introduced as a decentralized machine learning paradigm that enables collaborative model training without sharing raw data. Since its introduction, FL has

gained significant attention in healthcare due to its ability to preserve privacy while enabling multi-institutional collaboration. Researchers have demonstrated the effectiveness of FL in disease prediction, clinical decision support, and medical imaging analysis, where access to diverse datasets is essential for improving model generalization [1]. However, traditional FL frameworks rely heavily on centralized aggregation servers, which create a single point of failure and raise concerns about trust among participating institutions [6].

Recent research has explored security vulnerabilities in federated learning systems, particularly model poisoning and gradient leakage attacks. Studies have shown that adversaries can reconstruct sensitive patient data from shared gradients, posing serious privacy risks [4]. In addition, malicious participants can intentionally manipulate model updates to degrade performance or introduce backdoors, emphasizing the need for robust defense mechanisms [5].

Blockchain technology has been proposed as a solution to address trust and transparency challenges in federated learning. By providing an immutable distributed ledger, blockchain enables secure recording of model updates and participant activities. Smart contracts can automate training workflows, enforce collaboration policies, and eliminate reliance on centralized authorities. Several studies have demonstrated that blockchain integration improves accountability

and auditability in distributed AI systems [7][8]. However, blockchain alone cannot protect model parameters during aggregation, which limits its effectiveness as a standalone solution.

Secure Multi-Party Computation (SMPC) has been widely studied as a cryptographic technique for privacy-preserving collaborative computation. SMPC allows multiple parties to jointly compute functions without revealing their private inputs, making it highly suitable for secure aggregation in federated learning environments. Secure aggregation protocols ensure that model updates remain encrypted during computation, preventing inference attacks and protecting sensitive information [9].

Although prior work has explored federated learning, blockchain, and SMPC individually, there is limited research on fully integrated frameworks that combine all three technologies to address the complete range of security, privacy, and trust challenges in healthcare AI. Existing systems often lack comprehensive protection against malicious participants, transparent auditability, and decentralized trust mechanisms. This research aims to bridge this gap by proposing a unified architecture that integrates federated learning, blockchain, and SMPC to enable secure, transparent, and privacy-preserving collaboration in healthcare environments.

III EXISTING SYSTEM



The existing healthcare machine learning ecosystem largely relies on centralized architectures where hospitals and medical institutions share patient datasets with a central server for training and deploying AI models. While this approach simplifies model management and enables large-scale data analysis, it introduces significant risks related to privacy, security, and regulatory compliance. Healthcare data is highly sensitive, and transferring patient records, medical images, and clinical reports to centralized repositories increases the risk of data breaches, unauthorized access, and cyberattacks. Moreover, centralized systems create a single point of failure, meaning that if the central server is compromised, the entire infrastructure becomes vulnerable. Although techniques such as data anonymization and pseudonymization have been used to protect privacy, studies have shown that anonymized medical data can often be re-identified when combined with other datasets, making these methods insufficient for ensuring complete confidentiality.

Federated learning was introduced as an improvement by enabling institutions to train models locally and share only model updates instead of raw data. However, most existing federated learning systems still depend on a trusted central aggregation server, which creates new security and trust challenges. These systems remain vulnerable to model poisoning attacks, where malicious participants upload manipulated

updates that degrade the global model or introduce backdoors. Additionally, gradient leakage attacks have demonstrated that sensitive patient information can be reconstructed from shared model parameters, raising serious privacy concerns. The lack of transparency and auditability in current federated learning workflows further limits trust among collaborating institutions, highlighting the need for a more secure, decentralized, and trustworthy framework.

IV PROBLEM STATEMENT

The rapid adoption of artificial intelligence in healthcare has created a strong need for collaborative machine learning across multiple hospitals and research institutions. High-quality medical AI models require large and diverse datasets to achieve accurate and reliable predictions. However, healthcare data is highly sensitive and strictly regulated, making direct data sharing between organizations extremely difficult. Privacy regulations, ethical concerns, and the risk of data breaches prevent institutions from pooling their datasets into centralized repositories, resulting in isolated data silos that limit the performance and generalizability of machine learning models.

Although federated learning allows institutions to collaboratively train models without sharing raw data, existing federated learning systems still face significant security, trust, and transparency challenges. These systems often rely on

centralized aggregation servers, which create single points of failure and require all participants to trust a third party. Additionally, federated learning remains vulnerable to model poisoning attacks, gradient leakage, and lack of auditability in the training process. Therefore, there is a critical need for a secure, decentralized, and privacy-preserving framework that enables trustworthy collaboration among healthcare institutions while protecting sensitive patient data and ensuring the integrity of the global model.

V PROPOSED SYSTEM

The proposed system introduces a secure and decentralized learning framework designed specifically for healthcare environments where privacy, trust, and regulatory compliance are critical. Instead of relying on a central authority, the framework allows multiple hospitals to collaboratively train a shared machine learning model while keeping patient data within their own local infrastructure. Each participating institution trains the model using its private dataset and shares only encrypted model updates rather than raw medical records. This approach ensures that sensitive patient information never leaves the hospital's secure environment while still allowing the model to benefit from knowledge gained across multiple organizations.

To build trust among participants and remove dependence on a centralized aggregation server, the framework integrates blockchain technology as a distributed trust layer. Every model update

submitted by participating hospitals is recorded on a blockchain ledger, creating a permanent and tamper-resistant record of training activities. Smart contracts are used to manage the training process, verify participant identity, and automatically control the sequence of training rounds. This transparent logging mechanism ensures that all contributions can be audited and verified, preventing unauthorized or malicious modifications to the training workflow.

In addition to blockchain, the proposed system incorporates Secure Multi-Party Computation (SMPC) to protect model updates during the aggregation process. Instead of sending plain model parameters to an aggregator, each hospital encrypts its updates using cryptographic techniques. The SMPC protocol then securely combines these encrypted updates to produce a global model without revealing any individual contribution. This secure aggregation process prevents attackers from reconstructing sensitive data from gradients and protects against inference attacks that could otherwise expose patient information.

The integration of federated learning, blockchain, and SMPC creates a comprehensive solution that addresses the major limitations of existing healthcare AI systems. The proposed framework improves data privacy by keeping datasets local, enhances trust through decentralized verification, and strengthens security through cryptographic aggregation. By combining these technologies, the system enables healthcare institutions to

collaborate safely and efficiently, paving the way for scalable and trustworthy medical AI development.

VI METHODOLOGY

The methodology of the proposed framework focuses on building a secure and privacy-preserving federated learning workflow that integrates blockchain and Secure Multi-Party Computation to enable collaborative model training among healthcare institutions. The overall process begins with system initialization, where a global machine learning model is created and distributed to participating hospitals. Each institution acts as an independent client and trains the model locally using its private healthcare dataset, which may include electronic health records, medical images, and clinical reports. Since the data never leaves the local environment, patient privacy and regulatory compliance are maintained throughout the training process.

After local training is completed, each participant generates model updates in the form of gradients or model weights. Before transmission, these updates are encrypted using secure cryptographic techniques to prevent exposure of sensitive information. The encrypted updates are then submitted to the blockchain network, where smart contracts verify the authenticity of participants and record the submission of model updates. This step creates a transparent and tamper-proof log of all training activities,

ensuring accountability and preventing unauthorized participants from joining the training process.

The next stage involves secure aggregation using Secure Multi-Party Computation. Instead of sending updates to a centralized server, the SMPC protocol combines encrypted updates from multiple institutions without revealing individual contributions. This process produces a new global model that reflects the collective knowledge of all participating hospitals while preserving the confidentiality of each participant's data. The updated global model is then redistributed to all participants for the next training round.

This iterative training process continues until the model reaches the desired performance level. Throughout the methodology, the blockchain layer ensures trust, transparency, and auditability, while SMPC guarantees secure computation and protection against inference attacks. The combination of these technologies creates a robust training pipeline capable of supporting secure and decentralized collaboration in healthcare machine learning.

VII IMPLEMENTATION

The implementation of the proposed framework was designed as a practical prototype to demonstrate how secure federated learning can be deployed in a real healthcare environment. The system was developed using a modular architecture so that each component—federated learning, blockchain, and secure multi-party

computation—could operate independently while still functioning as part of an integrated pipeline. The implementation was carried out in a simulated multi-hospital setting, where each hospital node was represented as a separate client with its own local dataset and computing environment.

The development began with the federated learning layer, where a global machine learning model was initialized and distributed to all participating nodes. Each hospital trained the model locally using its private dataset, which was stored and processed within the institution's secure environment. Python-based machine learning libraries were used to perform local model training and evaluation. After completing local training, each node generated model updates and prepared them for secure transmission. Instead of sending raw model parameters directly, encryption techniques were applied to ensure that the updates could not be interpreted by unauthorized parties.

The blockchain layer was implemented to create a transparent and tamper-resistant record of all training activities. A private blockchain network was configured to simulate a consortium of healthcare institutions. Smart contracts were used to manage participant registration, verify the authenticity of submitted updates, and maintain a secure log of training rounds. This ensured that every contribution made during the training process could be tracked and audited, which is

essential for building trust among participating organizations.

To protect model updates during aggregation, Secure Multi-Party Computation was integrated into the system. The encrypted model updates from different hospitals were combined using secure aggregation protocols, ensuring that individual contributions remained confidential. The aggregation process produced a new global model, which was then redistributed to all participating nodes for the next training round. This iterative process continued until the model achieved stable performance.

The prototype was tested in a controlled environment to evaluate system functionality, communication flow, and integration between components. The implementation demonstrated that the proposed architecture can successfully support secure and decentralized model training without exposing sensitive healthcare data.

VIII RESULTS AND ANALYSIS

The evaluation of the proposed framework was carried out in a simulated multi-hospital environment to observe how the system behaves when compared with centralized learning and basic federated learning. The main objective was to determine whether adding blockchain and secure multi-party computation improves privacy and security while maintaining acceptable model performance. During multiple training rounds, the model trained collaboratively across distributed datasets and demonstrated stable

convergence and improved prediction capability due to exposure to diverse medical data sources.

The proposed approach showed better resistance to privacy leakage because patient data remained inside local hospital environments and encrypted model updates were used during aggregation. In contrast, centralized learning required full data sharing, which increased privacy risks, while traditional federated learning still showed vulnerability to gradient-based inference attacks. The integration of blockchain added transparency to the training process and helped identify suspicious or unauthorized updates. Although the framework introduced slightly higher training time due to encryption and verification processes, the overhead remained manageable and acceptable for healthcare environments where security is a priority.

Performance Comparison Table

The results highlight that the proposed framework achieves a balanced trade-off between performance and security. While the training overhead increased slightly, the improvements in privacy protection, transparency, and attack resistance make the system more suitable for real-world healthcare applications.

IX CONCLUSION

This research presented a secure and privacy-preserving framework for collaborative machine learning in healthcare by integrating federated learning, blockchain technology, and secure multi-party computation. The study addressed the major limitations of existing healthcare AI systems, including data privacy concerns, lack of trust among institutions, vulnerability to malicious attacks, and limited transparency in the training process. By allowing hospitals to train models locally and share only encrypted updates, the proposed approach ensured that sensitive patient data remained within institutional boundaries while still enabling collaborative model development.

The integration of blockchain introduced a decentralized trust mechanism that provided transparency, auditability, and protection against unauthorized participation. Secure multi-party computation further strengthened the framework by enabling encrypted aggregation of model updates, preventing data leakage and reducing the risk of inference attacks. Experimental

Approach	Prediction Accuracy	Privacy Protection	Resistance to Attacks	Transparency
Centralized Learning	High	Low	Low	Low
Traditional Federated Learning	High	Medium	Medium	Low
Proposed Secure FL (Blockchain + SMPC)	Very High	Very High	High	High

observations demonstrated that the proposed system achieved competitive model performance while significantly improving privacy protection and resistance to malicious activities.

Although the framework introduced moderate computational and communication overhead, the benefits in terms of security, trust, and regulatory compliance outweigh these limitations. The proposed architecture offers a practical and scalable solution for real-world healthcare environments where secure data collaboration is essential. Overall, this work highlights the potential of combining federated learning, blockchain, and cryptographic techniques to support trustworthy and privacy-aware medical AI systems.

REFERENCES

- [1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
- [2] P. Kairouz et al., "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [3] S. Rieke et al., "The Future of Digital Health with Federated Learning," *npj Digital Medicine*, vol. 3, no. 119, 2020.
- [4] L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [5] A. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Analyzing Federated Learning through an Adversarial Lens," *International Conference on Machine Learning (ICML)*, 2019.
- [6] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, 2019.
- [7] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, 2020.
- [8] M. Kim, J. Park, Y. Bennis, and S. Kim, "Blockchain and Federated Learning for Secure and Privacy-Preserving Edge Intelligence," *IEEE Communications Magazine*, vol. 57, no. 12, 2019.
- [9] K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," *ACM Conference on Computer and Communications Security (CCS)*, 2017.