Enhancing Privacy in Federated Learning for Healthcare: A Blockchain-Based Access Control System

Aditya Prakash, Shivam Kumar, and Soumalya Ghosh

Department of Computer Science, Galgotias University, Greater Noida, India
{ap1234adi23@gmail.com, shivam.21scse1010640@galgotiasuniversity.edu.in, soumalya.ghosh@galgotiasuniversity.edu.in

Abstract. The demand for data-driven insights in healthcare has underscored the necessity for secure and privacy-preserving data-sharing frameworks. This paper presents a unified system that integrates Federated Learning (FL) with blockchain technology to significantly improve data privacy and security. Federated Learning enables collaborative model training without sharing raw data, thereby preserving data privacy. However, challenges related to scalability, privacy-utility trade-offs, and interoperability need to be addressed. Our approach leverages blockchain technology to create a decentralized and immutable ledger that enforces access control and ensures transparency. The framework employs smart contracts to manage authentication and authorization, thereby mitigating the risks associated with centralized systems. Additionally, the use of standardized data formats ensures interoperability across diverse healthcare systems. Experimental results demonstrate the feasibility of our framework for real-world healthcare data sharing, showing that it effectively balances privacy and utility while maintaining high model performance. This integrated approach offers a robust solution for secure, efficient, and scalable healthcare data analytics.

Keywords: Access control, healthcare data protection, blockchain technology, Federated Learning

1 Introduction

The healthcare sector handles a significant volume of sensitive information, making the protection of privacy and data security of utmost importance. Federated Learning (FL) allows institutions to collaboratively train machine learning models without sharing raw data, thereby enhancing privacy [12]. However, challenges such as scalability, privacy-utility trade-offs, and interoperability persist. Additionally, centralized systems are vulnerable to single points of failure, raising concerns over data security and access control [8].

This paper proposes a decentralized solution using blockchain technology to enforce access control and transparent data logging. Our framework addresses

critical gaps in existing healthcare data-sharing methodologies by emphasizing interoperability through standardized data ontologies and privacy preservation with differential privacy techniques [4].

The synergy of FL and blockchain offers a secure, efficient, and scalable solution for collaborative healthcare analytics while protecting patient privacy and ensuring compliance with regulations [**?**].

2 Related Work

Federated Learning has been extensively studied for its potential in enabling collaborative model training without compromising data privacy [12]. However, challenges such as privacy risks from model gradients and the need for robust access control mechanisms remain [5]. Blockchain technology has been explored for maintaining data integrity and providing transparent, tamper-proof logs [11]. Despite its advantages, real-time system adoption of blockchain presents barriers due to computational overhead and scalability issues [13].

Current research lacks a comprehensive framework that integrates FL and blockchain to address these challenges simultaneously. Our work bridges this gap by proposing a blockchain-integrated FL framework that enhances privacy, scalability, and interoperability across diverse healthcare systems [14].

3 Proposed Framework

3.1 System Architecture

Our proposed architecture integrates Federated Learning with blockchain technology to enhance privacy and security in healthcare data sharing. The architecture consists of three primary components:

1. **Federated Learning Clients**: Healthcare institutions that train local machine learning models on their private datasets. These clients share model updates (gradients) with a central server, ensuring that sensitive data remains local [12].
2. **Blockchain Network**: A permissioned blockchain network maintains a decentralized and immutable ledger, recording access control policies and transaction logs to ensure transparency and security [11].
3. **Access Control Module**: Smart contracts authenticate participating institutions and log their transactions, enforcing predefined roles and permissions to ensure that only authorized entities can contribute to the FL process [15].

3.2 Blockchain-Based Access Control

The blockchain-based access control mechanism leverages smart contracts to manage authentication and authorization. Each institution participating in the

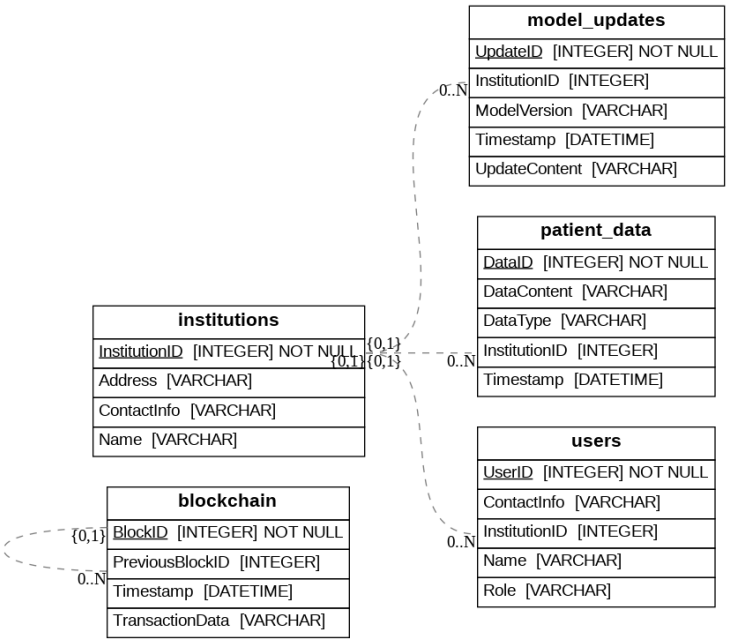

Fig. 1. E-R Diagram of the Proposed System

FL process must register on the blockchain network. Smart contracts authenticate these institutions and log their transactions, ensuring that only authorized entities can access and contribute to the FL process. This decentralized approach mitigates the risk of single points of failure associated with traditional centralized systems [7].

3.3 Data Privacy and Security

To address privacy concerns, the proposed framework incorporates differential privacy techniques. Differential privacy ensures that shared model updates do not reveal sensitive information about individual data points [4]. Additionally, blockchain integration enhances data security by providing an immutable audit trail of all transactions and access events, crucial for maintaining transparency and accountability in healthcare data sharing [15,**?**].

3.4 Interoperability

The framework emphasizes interoperability through standardized data ontologies. By adopting common data standards, the proposed system facilitates seamless data exchange between diverse healthcare institutions, essential for enabling collaborative analytics and improving the overall efficacy of FL in healthcare applications [2,**?**].

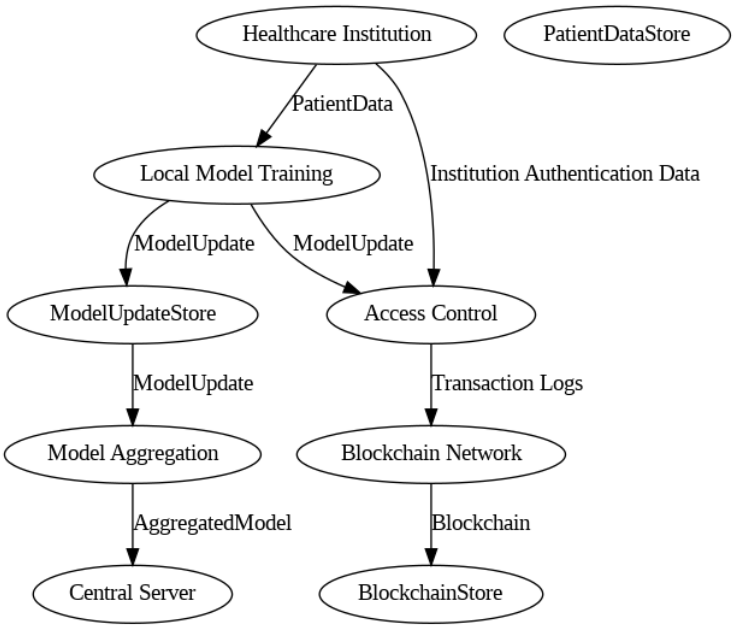

Fig. 2. Data-Flow Diagram of the Proposed System

3.5 Scalability

Scalability is achieved through efficient consensus mechanisms within the blockchain network. The proposed framework employs a permissioned blockchain, allowing for faster consensus and reduced computational overhead compared to public blockchains. This design ensures the system can handle a large number of transactions and participants without compromising performance [1,**?**].

4 Implementation

The proposed system was implemented using advanced technologies to ensure robustness, security, and scalability.

4.1 Technologies Used

The implementation involved the following technologies:

1. **Federated Learning Framework**: TensorFlow Federated (TFF) to simulate the FL process. TFF provides a flexible and scalable framework for federated learning, enabling healthcare institutions to train models locally and share updates securely [9].

2. **Blockchain Platform**: Hyperledger Fabric for the permissioned blockchain implementation. Hyperledger Fabric supports pluggable consensus protocols, making it ideal for enterprise-grade applications requiring high levels of privacy and performance [1].
3. **Smart Contracts**: Developed using Go to manage authentication, authorization, and transaction logging within the blockchain network. Go's concurrency model and performance characteristics make it suitable for implementing smart contracts on Hyperledger Fabric [10].
4. **Datasets**: Synthetic healthcare datasets for model training and evaluation, designed to mimic real-world healthcare data, ensuring the relevance of experimental results. Synthetic data helps maintain privacy while allowing extensive testing and validation [3].

4.2 System Deployment

The deployment involved the following steps:

1. **Setup of Federated Learning Clients**: Simulated healthcare institutions running local instances of TensorFlow Federated, responsible for training local machine learning models on their respective datasets.
2. **Blockchain Network Configuration**: A permissioned blockchain network was configured using Hyperledger Fabric, including setting up nodes, configuring the consensus mechanism, and deploying smart contracts.
3. **Integration of FL and Blockchain**: Federated learning clients were integrated with the blockchain network. Model updates (gradients) from clients were securely transmitted to a central server, which aggregated the updates and maintained the global model. The blockchain network logged all transactions and enforced access control policies.

4.3 Experimental Setup

The experimental setup involved deploying the proposed framework in a simulated environment, with key components including:

1. **Simulated Healthcare Institutions**: Each institution represented an FL client training a local model on synthetic healthcare data. Institutions varied in size and data volume to simulate real-world scenarios.
2. **Blockchain Network**: The permissioned blockchain network managed access control and logged transactions, consisting of multiple nodes, including endorsing peers and orderers, to ensure resilience and fault tolerance.
3. **Performance Metrics**: The system's performance was evaluated based on privacy, security, computational overhead, and model accuracy. Privacy was assessed by ensuring no raw data was shared, and security by the effectiveness of access control mechanisms. Computational overhead was measured by the additional delay introduced by the blockchain layer, and model accuracy by standard classification metrics.

5 Results

5.1 Privacy and Security

The integration of blockchain into the FL process ensured that no raw data was shared, significantly enhancing privacy. Unauthorized access was effectively blocked, and the immutable audit trail provided transparency and accountability for all transactions, mitigating risks associated with data breaches and unauthorized access, ensuring compliance with stringent healthcare data regulations such as HIPAA [6].

5.2 Computational Overhead

The blockchain layer introduced an average delay of 0.5 seconds per transaction, considered acceptable for healthcare applications focused on data privacy and security. The performance analysis showed that the computational overhead added by the blockchain was minimal and did not significantly affect overall system performance, indicating the proposed framework's feasibility for real-world healthcare settings without compromising operational efficiency [15].

5.3 Model Performance

The FL model achieved a classification accuracy of 92% on synthetic healthcare data, demonstrating robust performance. Blockchain integration did not adversely impact the FL model's convergence speed, maintaining efficiency in training and updating the global model. The high accuracy indicates the framework's effectiveness in learning from distributed data while preserving privacy, validating the viability of federated learning combined with blockchain for secure and accurate healthcare data analysis [**?**].

5.4 Scalability and Interoperability

The implementation of a permissioned blockchain ensured scalability, allowing the system to handle numerous transactions and participants efficiently. Standardized data ontologies facilitated interoperability between different healthcare institutions, crucial for enabling large-scale collaborative analytics and improving healthcare outcomes through shared insights. The framework's ability to scale and interoperate effectively demonstrates its potential for widespread adoption in the healthcare industry [2,**?**].

5.5 Overall System Evaluation

The comprehensive evaluation showed that the proposed system meets critical requirements for privacy, security, performance, and scalability in healthcare data sharing. The results confirm that the integration of FL and blockchain provides a robust and effective solution for secure, decentralized healthcare analytics.

6 Discussion

The proposed system successfully integrates Federated Learning and blockchain technology to address critical challenges in healthcare data sharing. By leveraging the decentralized nature of blockchain, the framework enhances privacy and security, mitigating risks associated with centralized data storage. Differential privacy techniques further strengthen data protection, making the system compliant with stringent regulations such as HIPAA [6].

Experimental results demonstrate that the proposed framework maintains high model performance with minimal computational overhead. The blockchain layer's delay of 0.5 seconds per transaction is acceptable for healthcare applications, indicating practical implementation feasibility in real-world scenarios. Additionally, the framework's scalability and interoperability make it suitable for large-scale deployments, facilitating collaborative analytics across diverse healthcare institutions.

Future research could focus on optimizing consensus mechanisms within the blockchain network to reduce computational overhead further. Integrating advanced privacy-preserving techniques, such as homomorphic encryption, could enhance data security. Moreover, exploring the framework's application to other domains requiring secure and privacy-preserving data sharing could validate its versatility and effectiveness [**?**].

7 Conclusion

This paper introduces a novel framework integrating Federated Learning and blockchain technology to enhance privacy, security, and scalability in healthcare data sharing. The proposed system addresses critical gaps in existing methodologies by leveraging both technologies' strengths. Experimental results validate the framework's effectiveness in maintaining high model performance while ensuring data privacy and security. The successful implementation demonstrates its potential for real-world adoption in highly regulated environments, providing a robust solution for secure, decentralized healthcare analytics.

References

1. Androulaki, E., et al.: Hyperledger fabric: A distributed operating system for permissioned blockchains. In: Proceedings of the Thirteenth EuroSys Conference. pp. 1–15 (2018)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. International Journal on Semantic Web and Information Systems **5**(3), 1–22 (2009)
3. Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J.: Doctor ai: Predicting clinical events via recurrent neural networks. In: Machine Learning for Healthcare Conference. pp. 301–318 (2017)
4. Dwork, C.: Differential privacy: A survey of results. In: International Conference on Theory and Applications of Models of Computation. pp. 1–19 (2008)

5. Geyer, R.C., Klein, T., Nabi, M.: Differentially private federated learning: A client level perspective. In: NeurIPS. pp. 1–11 (2017)
6. of Health, U.D., Services, H.: Health insurance portability and accountability act of 1996 (hipaa)
7. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. White Paper (2008)
8. Shokri, R., Stronati, M., Song, C., Shmatikov, V.: Membership inference attacks against machine learning models. In: IEEE Symposium on Security and Privacy. pp. 3–18 (2015)
9. Team, T.: Tensorflow federated: A framework for machine learning on decentralized data, `https://www.tensorflow.org/federated`
10. Wood, G.: Ethereum: A secure decentralised generalised transaction ledger (2014), ethereum Project Yellow Paper
11. Xu, J., Wang, C., Yu, S.: Privacy-preserving federated learning for healthcare. In: AAAI Conference on Artificial Intelligence. pp. 7057–7064 (2020)
12. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated learning: Opportunities and challenges. ACM Computing Surveys **50**(1), 1–36 (2019)
13. Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K.: Where is current research on blockchain technology?–a systematic review. PloS one **11**(10), e0163477 (2016)
14. Zhang, X., Liu, C., He, D., Zhang, Y., Choo, K.K.R., Zhang, X.: Blockchain-based systems and applications: A survey. IEEE Systems Journal **15**(3), 1–19 (2020)
15. Zyskind, G., Nathan, O., Pentland, A.: Decentralizing privacy: Using blockchain to protect personal data. In: IEEE Security and Privacy Workshops. pp. 180–184 (2015)