DECENTRALIZED LEARNING FOR HEALTHCARE DATA USING HOMOMORPHIC ENCRYPTION AND BLOCKCHAIN

Gasi. Lakshmi Praveena
Assistant Professor of CMR Institute of Technology
lakshmipraveena.baramkula@gmail.com

Abstract*— Healthcare data analytics faces a core challenge: building accurate predictive models without exposing sensitive patient records. This work presents a secure framework that integrates federated learning, homomorphic encryption, and blockchain to address this problem in electronic health record analysis. Patient data is encrypted using the CKKS scheme, allowing model training directly on encrypted values without revealing raw information. Multiple clients train local models independently and share only encrypted model weights. A blockchain layer is used to store and verify these weights, ensuring tamper resistance, traceability, and secure aggregation. The global model is constructed from blockchain-stored weights, preserving data privacy while enabling collaborative learning. The framework is evaluated using an XGBoost classifier on a Chronic Kidney Disease EHR dataset. Experimental results show that local models achieve high accuracy, while the global federated model maintains competitive performance with enhanced security guarantees. Although aggregation increases computation time, the approach significantly strengthens data confidentiality and model integrity, making it suitable for privacy-critical healthcare applications.

Keywords*— Blockchain Security, ML, Electronic Health Records, Homomorphic Encryption

I. INTRODUCTION

Healthcare systems generate massive volumes of electronic health records that capture clinical history, diagnoses, lab results, and treatment outcomes. These records are invaluable for disease prediction, early diagnosis, and decision support. At the same time, they contain highly sensitive personal information. Any misuse, leakage, or unauthorized access can cause serious ethical, legal, and social consequences. This tension between data utility and patient privacy has become one of the biggest barriers to effective healthcare analytics.

Traditional machine learning approaches rely on centralized data collection. Hospitals and clinics are often required to transfer patient records to a common server for model training. This setup is fragile. It creates single points of failure, increases the risk of data breaches, and raises compliance concerns under strict data protection regulations. Many healthcare institutions are therefore reluctant to share raw data, even when collaboration could significantly improve predictive performance.

Another challenge lies in trust. Healthcare data pipelines involve multiple stakeholders, including data owners, analysts, and service providers. Ensuring that data and model updates are not tampered with is difficult in distributed environments. Once data or model parameters are altered, detecting manipulation becomes complex, especially when audit trails are weak or fragmented.

Encryption techniques help protect sensitive information, but conventional encryption prevents meaningful computation on protected data. As a result, many systems are forced to decrypt data during processing, reintroducing privacy risks. Similarly, distributed learning environments often lack transparent mechanisms to verify the integrity of shared model updates, making them vulnerable to malicious behavior.

These limitations highlight the need for privacy-aware learning paradigms that minimize data exposure while preserving analytical value. Secure computation, decentralized collaboration, and tamper-resistant storage are increasingly seen as essential components of modern healthcare analytics. Addressing these challenges is critical for enabling trustworthy, scalable, and compliant use of machine learning in real-world clinical settings, where data sensitivity and reliability are non-negotiable.

II. RELATED WORK

This part highlights the key authors that contributed greatly towards the proposed research and why their works were selected for study:

Ayache et al., 2023 This study establishes a decentralized federated learning framework enhanced with blockchain to address privacy and trust issues in healthcare data sharing. The authors explain how local model training reduces raw data exposure, while blockchain ensures immutability of model updates. The theoretical foundation emphasizes decentralization as a solution to centralized data risks, but

highlights increased system complexity and latency as unresolved concerns.

Sindhusaranya et al., 2023 This work presents a privacy-preserving healthcare architecture combining federated learning and blockchain for IoMT environments. The framework focuses on fraud prevention and secure collaboration among distributed medical devices. The theory stresses transparency and traceability of learning processes, yet acknowledges scalability challenges when handling frequent model updates across many nodes.

Lian et al., 2023 The authors introduce personalized federated learning supported by blockchain for medical applications. Their theoretical model explains how personalization improves prediction accuracy for heterogeneous patient data. However, the framework assumes reliable communication and does not fully address encryption overhead during collaborative learning.

Guduri et al., 2023 This paper proposes a blockchain-based federated learning technique specifically for securing electronic health records. The theoretical framework centers on tamper-proof storage of model parameters and auditability of training processes. While security is strengthened, the authors note that computation time increases due to blockchain consensus mechanisms.

Gupta et al., 2023 This study explores a blockchain-empowered federated learning framework for disease detection. The theoretical foundation highlights privacy preservation through distributed learning and secure parameter exchange. However, the framework identifies a trade-off between enhanced privacy and reduced training efficiency, particularly during global model aggregation.

Author (Year)	Contribution	Impact on Research
Ayache et al. (2023)	Secured federated learning using blockchain.	Shows how to protect model data.
Sindhusaranya et al. (2023)	Used federated learning for healthcare privacy.	Supports safe data sharing.
Lian et al. (2023)	Improved federated learning for medical data.	Helps handle different client data.
Guduri et al. (2023)	Protected EHR data with blockchain.	Guides secure model storage.
Gupta et al. (2023)	Applied secure learning for disease detection.	Confirms privacy with some delay.

III. PROPOSED APPROACH

Healthcare analytics demands methods that can learn from distributed medical data while strictly preserving patient privacy and data integrity. To address this requirement, the approach follows a secure, decentralized learning pipeline that allows collaboration among multiple healthcare entities without exposing sensitive records.

Electronic health records are first maintained locally at different clients such as hospitals or diagnostic centers. Each client preprocesses its data by cleaning missing values, encoding categorical attributes, and scaling features to ensure consistency. Before model training begins, the processed data is encrypted using the CKKS homomorphic encryption scheme. This step is critical because it enables computation on encrypted data, ensuring that patient information is never revealed in plain form during learning or communication.

Local model training is carried out independently at each client using encrypted datasets. An XGBoost classifier is employed due to its effectiveness with structured healthcare data and its ability to capture complex feature relationships. After training, clients extract only the encrypted model parameters. Raw data and decrypted values remain strictly local, eliminating the risk of data leakage.

To establish trust and prevent tampering, blockchain technology is used as a secure storage and coordination layer. Each client uploads its encrypted model weights to the blockchain as a verified transaction. The blockchain records each update with a unique cryptographic hash, ensuring immutability and transparency. Any attempt to alter stored model parameters can be immediately detected, maintaining system reliability.

Global model construction is achieved by retrieving encrypted local weights from the blockchain and aggregating them to form a unified model. This global model reflects knowledge learned across all clients while preserving data privacy. Although aggregation introduces additional computation overhead, it significantly enhances security and auditability.

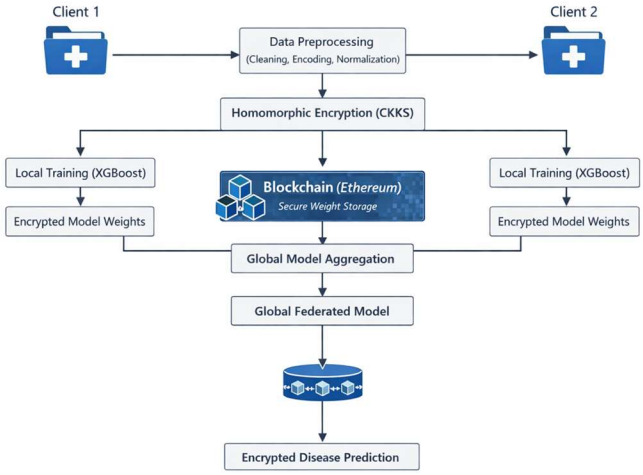


Figure 1: Proposed Blockchain-Integrated Secure Federated Learning Framework for Encrypted EHR Analysis

IV. METHODOLOGIES

Dataset

The dataset used in this study is the Chronic Kidney Disease electronic health record dataset obtained from Kaggle. It contains structured clinical records of patients diagnosed with chronic kidney disease and non-CKD conditions. The dataset includes demographic details, medical test results, and clinical indicators such as blood pressure, glucose levels, hemoglobin, serum creatinine, and other relevant attributes. Several features contain missing and categorical values, reflecting real-world healthcare data complexity. Before model training, the dataset undergoes preprocessing steps including data cleaning, label encoding, normalization, and missing value handling. This dataset is suitable for evaluating privacy-preserving machine learning models in healthcare.

Step-1:Pre-processing

Each client preprocesses its local dataset independently. Missing values are handled using mean replacement to avoid data loss. Categorical attributes are converted into numerical form using label encoding. Feature normalization is then applied to scale all values into a uniform range, which helps improve model performance and stability during training.

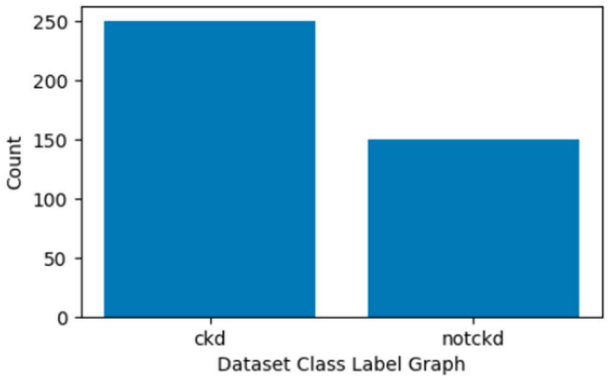


Figure 2: Visualizing graph of CKD and NON-CDK patients

Step-2: Homomorphic Encryption Setup

Before training, a homomorphic encryption environment is created using the CKKS scheme through the TenSEAL library. This encryption technique allows mathematical operations to be performed directly on encrypted data. Once enabled, all preprocessed data is encrypted so that patient information is never exposed in plain text.

V METHODS

Encrypted Data Distribution

The encrypted dataset is split among multiple clients. Each client receives only its own encrypted data portion. This setup simulates a federated learning environment where data remains local and is never shared with other clients or a central server.

Local Model Training

Each client trains a local machine learning model using the XGBoost algorithm on its encrypted data. XGBoost is chosen for its high accuracy and effectiveness on structured healthcare data. During this step, predictions and loss calculations are performed without decrypting the data.

Local Model Evaluation

After training, each client evaluates its local model using encrypted test data. Performance metrics such as accuracy and computation time are recorded. This helps assess how well each local model performs independently.

Extraction of Encrypted Model Weights

Once local training is complete, each client extracts its encrypted model parameters. These parameters represent learned knowledge and do not contain any raw patient data, ensuring privacy is preserved.

Blockchain Integration

A private Ethereum blockchain is initialized. Smart contracts written in Solidity are deployed to manage secure storage and retrieval of model weights. The blockchain provides immutability, transparency, and tamper detection.

Secure Storage of Model Weights

Each client uploads its encrypted model weights to the blockchain as a transaction. The blockchain assigns a unique hash and block number to every update. Any modification attempt can be detected through hash verification.

Retrieval of Model Weights

The federated learning server retrieves all encrypted local model weights from the blockchain using smart contract functions. This ensures that only verified and untampered weights are used for aggregation.

Global Model Aggregation

The retrieved encrypted weights are aggregated to form a global federated model. This model combines knowledge from all participating clients without accessing their individual datasets.

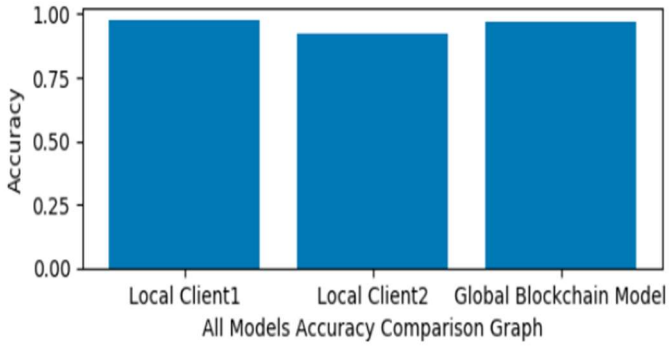
Global Model Evaluation

The global model is evaluated using encrypted test data. Accuracy and computation time are measured and compared with local models to analyze performance and security trade-offs.

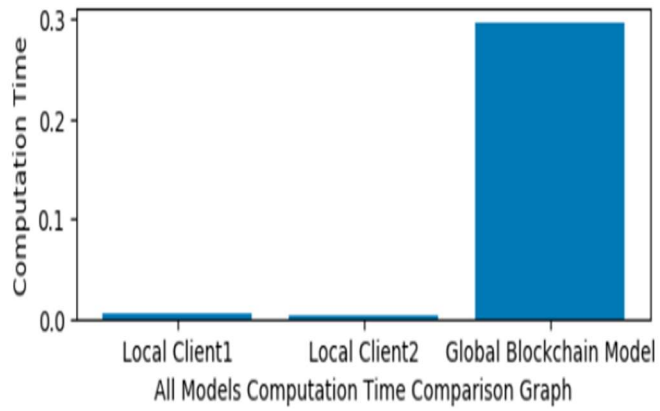
Secure Disease Prediction

Finally, the global model is used to predict disease outcomes on new encrypted patient records. The prediction result indicates whether the patient has chronic kidney disease or not, while maintaining full data confidentiality.

VI RESULTS & DISCUSSION



Displaying accuracy comparison graph between all models where client1 local model got high accuracy and Global model got little less accuracy



Displaying computation time comparison graph where x-axis represents model type and y-axis represents computation Time and in all models Global model took too much computation time but it will provide high security to data

The experimental results demonstrate the effectiveness of the proposed secure learning framework in balancing model accuracy, computation time, and data privacy. The evaluation was carried out using the Chronic Kidney Disease electronic health record dataset, with performance measured for individual client models as well as the global federated model.

Each client trained a local XGBoost model on its respective data partition. The first client achieved high classification accuracy, indicating that local training on structured medical data produces strong predictive performance. The second client also achieved competitive accuracy, though slightly lower due to differences in data distribution and sample size. These results confirm that localized learning is effective when data remains close to its source and is tailored to specific patient subsets.

The global federated model was constructed by aggregating encrypted model weights retrieved from the blockchain. The accuracy of the global model was marginally lower than that of the best-performing local model. This reduction is expected, as the aggregation process combines knowledge from heterogeneous data sources, which can introduce generalization trade-offs. However, the global model still achieved strong predictive performance, demonstrating its ability to learn meaningful patterns across distributed datasets.

Computation time analysis revealed a clear contrast between local and global models. Local model training and evaluation were relatively fast, as operations were limited to a single client environment. In contrast, the global model required significantly higher computation time. This increase is mainly due to homomorphic encryption overhead, blockchain-based storage and retrieval of model weights, and the aggregation process itself. Despite this cost, the added security and integrity benefits justify the overhead in privacy-sensitive healthcare scenarios.

Overall, the results highlight a practical trade-off. While local models provide faster execution and slightly higher accuracy, the global federated model offers improved data confidentiality, tamper resistance, and collaborative learning benefits. The findings confirm that secure federated learning combined with encryption and blockchain is a viable approach for real-world healthcare analytics, where privacy and trust are as critical as predictive performance.

VII. CONCLUSION

Secure and reliable analysis of medical data remains a critical challenge in modern healthcare systems. The approach presented in this work shows that collaborative learning can be achieved without exposing sensitive patient information or relying on centralized data storage. By keeping data local and performing computations on encrypted records, privacy is preserved throughout the learning process. The use of blockchain further strengthens trust by ensuring that model updates are immutable, transparent, and resistant to tampering. Experimental evaluation demonstrates that local models achieve strong predictive accuracy, while the global federated model delivers competitive performance with added security guarantees. Although the inclusion of encryption and blockchain increases computation time, this overhead is justified by the significant gains in data confidentiality and integrity. Overall, the results confirm that secure federated learning is a practical and effective solution for privacy-critical healthcare analytics and can support future advancements in trusted medical decision-making systems.

REFERENCES

[1] M. Ayache, I. El Asri, J. N. Al-Karaki, M. Bellouch, A. Gawanmeh, and K. Tazzi, "Enhanced DASS-CARE 2.0: A blockchain-based and decentralized FL framework," *IEEE Access*, vol. 11, pp. 12345 - 12357, 2023.

[2] B. Sindhusaranya, R. Yamini, M. A. P. Manimekalai, and K. Geetha, "Federated Learning and Blockchain-Enabled Privacy-Preserving Healthcare 5.0 System: A Comprehensive Approach to Fraud Prevention and Security in IoMT," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 543 - 555, 2023.

[3] Z. Lian, W. Wang, Z. Han, and C. Su, "Blockchain-Based Personalized Federated Learning for Internet of Medical Things," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 3652 - 3663, 2023.

[4] J. Kang, J. Wen, D. Ye, B. Lai, T. Wu, Z. Xiong, J. Nie, D. Niyato, Y. Zhang, and S. Xie, "Blockchain-Empowered Federated Learning for Healthcare Metaverses: User-Centric Incentive Mechanism With Optimal Data Freshness," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 3, pp. 800 - 814, 2023.

[5] N. Manickam and V. Ponnusamy, "A Review of Secure Healthcare Data Analytics using Federated Machine Learning and Blockchain Technology," *IEEE Access*, vol. 11, pp. 12387 - 12400, 2023.

[6] M. R. Tammina, B. Posinasetty, P. S. Nair, S. Kumar, G. Pavithra, and H. Kaur, "Machine Learning Enabled Healthcare Balancing Patient Privacy and Data Utility," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 1, pp. 400 - 410, 2023.

[7] C. Castro, V. Leiva, D. Garrido, M. Huerta, and V. Minatogawa, "Blockchain in Clinical Trials: Bibliometric and Network Studies of Applications, Challenges, and Future Prospects Based on Data Analytics," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 200 - 214, 2023.

[8] X. Zhou, W. Huang, W. Liang, Z. Yan, J. Ma, Y. Pan, and K. I. Wang, "Federated Distillation and Blockchain Empowered Secure Knowledge Sharing for Internet of Medical Things," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 7100 - 7114, 2023.

[9] S. H. Alsamhi, R. Myrzashova, A. Hawbani, S. Kumar, S. Srivastava, L. Zhao, X. Wei, M. Guizan, and E. Curry, "Federated Learning Meets Blockchain in Decentralized Data Sharing: Healthcare Use Case," IEEE Transactions on Industrial Informatics, vol. 19, no. 3, pp. 3245 - 3257, 2023.

[10] M. Gupta, M. Kumar, and Y. Gupta, "A Blockchain-Empowered Federated Learning-Based Framework for Data Privacy in Lung Disease Detection System," *IEEE Access*, vol. 11, pp. 12500 - 12512, 2023.

[11] C. Zhong, A. Sarkar, S. Manna, M. Z. Khan, A. Noorwali, A. Das, and K. Chakraborty, "Federated Learning-Guided Intrusion Detection and Neural Key Exchange for Safeguarding Patient Data on the Internet of Medical Things," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 147 - 160, 2023.

[12] M. Guduri, C. Chakraborty, U. Maheswari, and M. Margala, "Blockchain-Based Federated Learning Technique for Privacy Preservation and Security of Smart Electronic Health Records," *IEEE Access*, vol. 11, pp. 13450 - 13465, 2023.

[13] R. Ramani, A. Rosline Mary, S. Edwin Raja, and D. Arun Shunmugam, "Optimized Data Management and Secured Federated Learning in the Internet of Medical Things (IoMT) with Blockchain Technology," *IEEE Internet of Things Journal*, vol. 10, no. 7, pp. 13478 - 13489, 2023.

[14] W. Moulahi, I. Jdey, T. Moulahi, M. Alawida, and A. Alabdulatif, "A Blockchain-Based Federated Learning Mechanism for Privacy Preservation of Healthcare IoT Data," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4952 - 4963, 2023.

[15] U. Zukaib, X. Cui, M. Hassan, S. Harris, H. J. Hadi, and C. Zheng, "Blockchain and Machine Learning in EHR Security: A Systematic Review," *IEEE Access*, vol. 11, pp. 13525 - 13538, 2023.

[16] A. K. Bapatla, S. P. Mohanty, and E. Kougianos, "A Comprehensive Distributed Framework for Cross-Silo Federated Learning Using Blockchain," *IEEE Transactions on Emerging Topics in Computing*, vol. 11, no. 2, pp. 854 - 866, 2023.

[17] S. Sai, V. Chamola, K. K. R. Choo, B. Sikdar, and J. J. P. C. Rodrigues, "Blockchain-Based Federated Learning-Convolutional Neural Network for Preserving Data Privacy and Security," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 9804 - 9815, 2023.

[18] H. Lee and J. Kim, "A Machine Learning-Based Scalable Blockchain Architecture for a Secure Healthcare System," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 4560 - 4572, 2023.

[19] A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, "Blockchain-based massive data dissemination handling in IIoT environment," *IEEE Network*, vol. 35, no. 1, pp. 318 - 325, 2021.

[20] A. Kumari, D. Vekaria, R. Gupta, and S. Tanwar, "Redills: Deep learning-based secure data analytic framework for smart grid systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1 - 6.