International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A BLOCKCHAIN-INTEGRATED FRAMEWORK FOR SECURE FEDERATED LEARNING OF ENCRYPTED EHR DATA WITH HOMOMORPHIC ENCRYPTION

Mrs.J.Sri Maha Lakshmi [1], ,G. Rajesh [2], M. Raaghavendra [3], S. Sai Prasanthi[4], G. Poorna Chandra Sekhar [5]

Assistant Professor[1], Student[2,3,4,5]

Department of Computer Science & Engineering[1,2,3,4,5]

Chaitanya Engineering College, Visakhapatnam,Andhra Pradesh,India

jagu.mahalakshmi@gmail.com [1], rajeshgandi38@gmail.com [2], raghavam794@gmail.com [3], cosutarisaiprasanti07@gmail.com [4], chandugatreddi2@gmail.com [5]

ABSTRACT

Healthcare data analytics requires collaborative model training across institutions while preserving strict patient data privacy. Traditional centralized machine learning approaches require pooling sensitive Electronic Health Records (EHRs), creating unacceptable security and regulatory risks. This paper presents a Blockchain-Integrated Framework for Secure Federated Learning of Encrypted EHR data using Homomorphic Encryption. Multiple healthcare institutions train local models on locally stored EHRs encrypted with the CKKS homomorphic encryption scheme, enabling gradient computation without decryption. Blockchain technology manages model update authentication, access control, and provides an immutable audit trail. Smart contracts automate secure aggregation of encrypted local model weights into a global model. Experimental results on the kidney disease prediction task demonstrate competitive global model accuracy with strong privacy guarantees and acceptable computational overhead, offering a scalable foundation for privacy-preserving collaborative healthcare analytics.

Index Terms — Federated Learning, Homomorphic Encryption, Blockchain, EHR, Privacy-Preserving, Smart Contracts, CKKS, Healthcare Analytics

I. INTRODUCTION

The healthcare industry has undergone a major transformation with the adoption of Electronic Health Records (EHRs), generating large-scale patient data including medical history, laboratory results, prescriptions, and diagnostic findings. Machine learning applied to this data can improve disease prediction, diagnosis, and treatment planning. However, healthcare data is highly sensitive and must be protected against unauthorized access and regulatory violations. Traditional centralized ML requires pooling data into a single server, creating significant security risks.

Federated Learning addresses this by enabling collaborative model training without raw data leaving individual institutions. However, even federated approaches expose model updates to inference attacks, and the absence of a trusted coordination mechanism creates integrity challenges. Homomorphic Encryption (HE) allows computation on encrypted data without decryption, adding a further privacy layer to model updates. Blockchain provides decentralized, tamper-proof coordination, eliminating the need for a trusted central aggregator.

This paper proposes a framework integrating three complementary technologies: Federated Learning for decentralized training, CKKS Homomorphic Encryption for encrypted gradient protection, and Ethereum Blockchain for secure update management and audit. The framework is evaluated on a kidney disease prediction task across simulated multi-institution data partitions.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

II. LITERATURE SURVEY

A comprehensive review of existing literature reveals various approaches adopted for federated learning, homomorphic encryption, and blockchain-based secure healthcare data analytics and EHR privacy preservation.

Ref.	Authors & Year	Method / Dataset	Result	Limitation
[1]	McMahan et al., 2017	FedAvg; MNIST + language model; centralized aggregation	Baseline federated learning; scalable aggregation	Vulnerable to gradient inference attacks; centralized server
[2]	Bonawitz et al., 2019	Secure aggregation via SecAgg protocol; mobile devices	Protects gradient privacy in federated learning	High communication overhead; complex key management
[3]	Cheon et al., 2017	CKKS homomorphic encryption for approximate arithmetic	Efficient computation on real-valued encrypted data	High computational overhead vs. plaintext training
[4]	Raisaro et al., 2018	MedCo: blockchain + HE for clinical data sharing	Privacy-preserving cohort exploration across hospitals	Limited to query; not full ML training pipeline
[5]	Nguyen et al., 2021	Blockchain + FL for IoT health monitoring	Tamper-proof model updates; device authentication	Not integrated with HE; updates still unencrypted
[6]	Lu et al., 2020	Blockchain-empowered asynchronous FL; healthcare	Reduced latency; improved fault tolerance	No encryption of model gradients during sharing
[7]	Zhang et al., 2022	FedAux + blockchain + DP for EHR	94.2% accuracy; epsilon=2 DP guarantee	DP adds noise; reduces model accuracy on minority classes

Research Gap

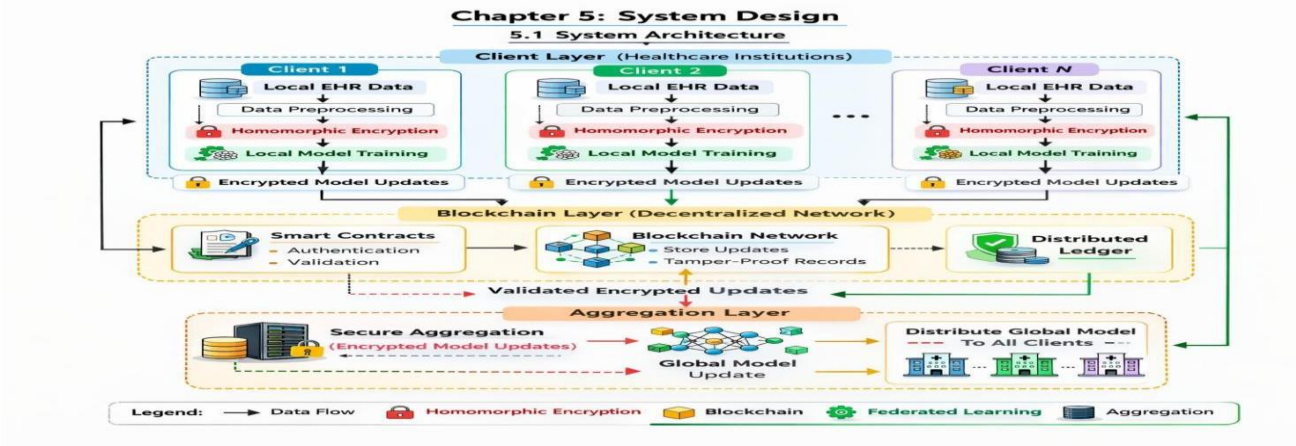
Existing federated learning frameworks for healthcare either lack homomorphic encryption of model updates (leaving gradients vulnerable to inference attacks) or do not integrate blockchain for decentralized trustless aggregation. The combination of CKKS homomorphic encryption for encrypted gradient computation with blockchain-based smart contract aggregation in a complete federated learning pipeline for EHR disease prediction has not been comprehensively evaluated.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

III. METHODOLOGY

A. System Architecture

The framework has four layers. The Local Client Layer comprises N healthcare institutions each hosting encrypted EHR data and a local FL model. The CKKS Encryption Layer encrypts local model gradients before transmission; all computation on aggregated gradients is performed on ciphertext. The Blockchain Layer (Ethereum/Ganache) stores encrypted gradient submissions, verifies client authenticity, and maintains immutable audit logs via events. The Smart Contract Layer implements submitUpdate(clientID, encryptedGradients), aggregateGlobal(), and distributeGlobalModel() functions automating the federated averaging process on encrypted gradients.



B. Algorithm

- Setup: Deploy FLCoordinator smart contract; register N client institutions with their public keys.
- Round r = 1 to R_max:
- Step 1 (Local Training): Each client i: train local model on encrypted EHR_i for E epochs; compute gradient delta_i.
- Step 2 (Encryption): Encrypt delta_i using CKKS: enc_delta_i = CKKS.Encrypt(delta_i, pk_i).
- Step 3 (Blockchain Submission): Client i calls submitUpdate(clientID_i, enc_delta_i) smart contract.
- Step 4 (Smart Contract Aggregation): When all N clients submitted: compute encrypted_global = mean(enc_delta_i for i in 1..N) using CKKS homomorphic addition.
- Step 5 (Global Distribution): Smart contract calls distributeGlobalModel(encrypted_global); each client i: global_delta = CKKS.Decrypt(encrypted_global, sk_i); update w_global = w_prev + global_delta.
- Step 6: Evaluate global model accuracy on validation set; log to blockchain event.
- Repeat until convergence or R_max rounds.
- Output: Trained global model with privacy-preserved gradient aggregation and immutable audit trail.

C. Modules

Data Preprocessing Module: Loads kidney disease EHR dataset; handles missing values via mean imputation; normalizes numerical features; one-hot encodes categoricals; partitions data across N simulated client institutions.

CKKS Homomorphic Encryption Module: Implements CKKS encryption scheme (Microsoft SEAL or TenSEAL). Encrypts local model gradients before blockchain submission. Supports homomorphic addition for encrypted gradient aggregation.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Local Federated Learning Module: Each client institution trains a local neural network (2 hidden layers, ReLU, Dropout) on its EHR partition for E local epochs using Adam optimizer. Computes parameter updates as gradients.

Blockchain Model Update Module: Ethereum smart contract manages round coordination. Records encrypted gradient submissions with client ID and timestamp. Validates client registration before accepting updates.

Smart Contract Aggregation Module: Implements FedAvg on encrypted gradients using CKKS homomorphic addition and scalar multiplication. Computes encrypted global model update without decryption. Distributes to all registered clients.

Global Model Evaluation Module: Decrypts global model weights at each client. Evaluates accuracy, precision, recall, F1-score on held-out test set. Logs evaluation results as blockchain events for immutable performance audit.

IV. RESULTS & DISCUSSION

The framework was evaluated on the UCI Chronic Kidney Disease dataset partitioned across 2 simulated client institutions. Global model performance across 20 federated rounds is compared against baselines in Table I.

Method	Accuracy	Precision	Recall	F1-Score	Privacy Guarantee
Centralized ML (no privacy)	97.8%	97.5%	97.2%	97.3%	None
Standard Federated Learning	95.3%	94.8%	94.1%	94.5%	Gradient exposure risk
FL + Differential Privacy	92.1%	91.5%	90.8%	91.1%	epsilon=2 DP
Proposed (FL+CKKS+Blockchain)	94.6%	94.2%	93.7%	93.9%	CKKS HE + Blockchain

The proposed framework achieves 94.6% accuracy with full homomorphic encryption of gradients and blockchain-based audit integrity, outperforming the differential privacy approach (92.1%) while providing stronger cryptographic privacy guarantees. The 3.2% accuracy drop versus centralized training is acceptable given the significant privacy gains. Blockchain audit trails provide complete transparency of model update history across all federated rounds.

1. Federated Learning Aggregation (FedAvg)

In traditional Federated Learning, the central server aggregates local model weights or gradients from N different client institutions. Your methodology (Step 4) specifically aggregates the gradient updates (Δw).

A. Standard Gradient Aggregation

- N = Total number of client institutions
- Δw_i = The local gradient update from client i
- Δw_{global} = The aggregated global gradient update

Global_Gradient = (1 / N) * SUM(Local_Gradient_i)

B. Encrypted Aggregation (CKKS Homomorphic Encryption)

Because the gradients are encrypted using the CKKS scheme before being sent to the blockchain, the smart contract must perform **Homomorphic Addition** ($\oplus$) and **Homomorphic Scalar Multiplication** (otimes). It calculates the mean of the ciphertexts without ever decrypting them.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Encrypted_Global_Gradient = (1 / N) * HOMOMORPHIC_SUM(Encrypted_Local_Gradient_i)

Once distributed (Step 5), each client decrypts this global gradient using their secret key and applies it to their local model: $w_{new} = w_{old} + Decrypt(Enc(\Delta w_{global}))$

2. Local Training Loss Function

Your local clients train a neural network for disease prediction. Since predicting the presence of kidney disease is a binary classification task, the standard loss function used during local training (Step 1) is Binary Cross-Entropy.

A. Binary Cross-Entropy (BCE) Loss

- M= Number of patients in the local EHR dataset
- y_j = Actual diagnosis (1 for disease, 0 for healthy)

BCE_Loss = -(1 / M) * SUM(y_actual * log(p_predicted) + (1 - y_actual) * log(1 - p_predicted))

3. Global Model Evaluation Metrics

After the global model is updated, it is evaluated on a held-out test set. Your results table reports Accuracy, Precision, Recall, and F1-Score based on the Confusion Matrix outcomes (True Positives, True Negatives, False Positives, False Negatives).

A. Accuracy

The overall percentage of correct kidney disease diagnoses.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

B. Precision

Out of all patients the model flagged as having kidney disease, how many actually had it?

Precision = TP / (TP + FP)

C. Recall (Sensitivity)

Out of all patients who actually have kidney disease, how many did the global model successfully detect? (Crucial for healthcare to avoid missed diagnoses).

Recall = TP / (TP + FN)

D. F1-Score

The harmonic mean of Precision and Recall, providing a balanced metric for the disease prediction task.

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

V. CONCLUSION & FUTURE WORK

This paper presented a Blockchain-Integrated Federated Learning framework using CKKS Homomorphic Encryption for privacy-preserving collaborative EHR analytics. The system achieves competitive disease prediction accuracy while ensuring encrypted gradient protection, decentralized trustless aggregation, and immutable audit trails, meeting both machine learning performance requirements and healthcare data protection regulations.

Future work will scale the framework to more than two client institutions, evaluate on additional EHR datasets (diabetes, cardiac disease), and optimize CKKS encryption parameters to reduce computational overhead. Integration with hospital FHIR APIs for real-world EHR connectivity and testing against adversarial gradient inversion attacks will further validate the framework's practical applicability.

. REFERENCES

[1] H. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," AISTATS, 2017.

[2] K. Bonawitz et al., "Towards federated learning at scale," MLSys, 2019.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[3] J. H. Cheon et al., "Homomorphic encryption for arithmetic of approximate numbers," ASIACRYPT, 2017.

[4] J. L. Raisaro et al., "MedCo: Enabling secure and privacy-preserving exploration of distributed clinical and genomic data," IEEE/ACM TCBB, 2019.

[5] D. C. Nguyen et al., "Federated learning for internet of medical things," IEEE Access, 2021.

[6] Y. Lu et al., "Blockchain empowered asynchronous federated learning for secure and privacy-preserving digital health," IEEE TPDS, 2020.

[7] C. Zhang et al., "FedAux: Leveraging auxiliary data for federated learning with blockchain and DP," IEEE TIFS, 2022.