Fed-Chain: Secure Federated Learning in Healthcare

Dr. Pankaj Kumar
Assistant Professor
Department of Information Science and Engineering
Acharya Institute of Technology, Bengaluru, India
pankaj2472@gmail.com

Pavan Kumar Reddy K P
Department of Information Science and Engineering
Acharya Institute of Technology, Bengaluru, India
pavankumar.22.beis@acharya.ac.in

Darshan K S
Department of Information Science and Engineering
Acharya Institute of Technology, Bengaluru, India
darshans.22.beis@acharya.ac.in

Niranjan
Department of Information Science and Engineering
Acharya Institute of Technology, Bengaluru, India
niranjans.22.beis@acharya.ac.in

D Skanda Mayya
Department of Information Science and Engineering
Acharya Institute of Technology, Bengaluru, India
dp.22.beis@acharya.ac.in

***Abstract*—The growing dependance on data-based decision-making in healthcare has brought attention to the vital importance of secure, privacy-preserving and collaborative learning techniques. Traditional centralized learning approaches in medical data often raise concerns regarding patient privacy data leaks and even regulatory troubles. Federated learning came as a good solution, where it allows model training in different hospitals without sharing the sensitive patient data. However, federated learning has its problems – it can be mislead with fake updates, the model can even be poisoned and it is really hard to trust every participants involved.**

In this work, we present fed-chain, a secure and scalable framework which brings together federated learning, blockchain and zero-knowledge-proofs(ZKPs) preserving the privacy of patient's data in healthcare. Blockchain here adds decentralized trust, immutability and makes model updates transparent to review while ZKPs helps in proving correctness without leaking personal data. We are implemented this framework for heart disease prediction where multiple hospitals train the model together but the data stays confidential. Our experimental results shown better accuracy, more strength against attacks and even low communication cost compared to other FL setups.

Overall, the systems gives a safer approach for working together on healthcare data, allowing hospitals and research centers to generate valuable predictions using these models while keeping the patient data private and safe.

***Index Terms*—Zero-Knowledge Proofs (ZKPs), Federated Learning (FL), Blockchain, Healthcare.**

I. Introduction

Healthcare is an area where timely and correct diagnosis can often save many lives. As the increase in diseases like heart problems, doctors need smarter tools to help them make better decisions. The widespread of medical records and the expansion of connected healthcare systems have created many chances to use artificial intelligence and machine learning for prediction and analysis to spot diseases early.

Even with the progress of AI in healthcare, protecting the sensitive patient data is still a big challenge. Standard machine learning systems depend on collecting data from many hospitals into a single place. Although this approach improves performance and enhance the accuracy in learning process, it also creates major risks like leaking raw patient data, fearing breaches, misuse and unauthorized access. Due to this, hospitals hesitate to share the patient data which limits the effectiveness of predictive analysis.

To overcome these challenges of data privacy, federated learning has been developed as a solution to tackle the challenges of data sharing by enabling the model training across different hospitals without sharing raw patient data. In this setup, each hospitals retain their patient data locally and only communicates model updates to build a global model. Though it protects the privacy to some extent, it is still open to challenges such as tampered updates and over-dependance on a single central server, this limits its scalability for sensitive domains like healthcare.

These challenges can be tackled by combining blockchain and advanced cryptographic technique. Blockchain creates a decentralized and tamper proof record of model updates, removing the need of a central server. Zero-knowledge-proofs provides security where it allows the hospitals to prove their updates are valid without revealing sensitive patient data. By combining these technologies we establish a secure, verifiable and privacy preserving framework for collaborative healthcare analytics.

This research introduces Fed-Chain, a secure and privacy preserving framework that combines federated learning with blockchain and zero-knowledge-proofs. Applied to heart disease prediction, Fed-Chain allows multiple hospitals can work together without sharing sensitive patient data. It improves model accuracy and defends malicious updates and reduces

communication cost, offering a realistic solution in healthcare.

II. Problem Statement

The growth of artificial intelligence has created many opportunities in healthcare. It can now be used to predict diseases at early stages, assist doctors for making better decisions for patients and design personalized treatment plans for them. Machine learning models examines patterns in patient data, which supports in detecting the illness sooner and enhances diagnostic accuracy. However, medical data is highly sensitive and subject to strict privacy rules which cannot be freely shared. As a result, hospitals and clinics often hesitate to share the patient information to the third parties or outside organizations.

In centralized machine learning where all data is stored in single location, creates serious concerns such as data breaches, misuse and unauthorized access of private data. As a result, this makes it difficult for hospitals to collaborate among other hospitals and prevent the models from being trained, reducing the opportunity to built models among diverse datasets.

Federated learning was brought in to solve the problem by allowing hospitals to keep the patient data to themselves but still contribute in training a shared model. Here, instead of sharing the raw patient data only model updates are shared, which lowers the risk of leaking the sensitive data. On the other hand, FL has many challenges to face in healthcare. It usually relies on single server to manage updates, which becomes a single point of failure and requires everyone to trust the server completely. Some participants can act maliciously, where they can send poisoned and tampered updates making the final model less accurate and unreliable. Most FL setups don't have a strong checks, its hard to believe that updates are valid and respect privacy standards.

In healthcare this issue becomes very essential that even a single wrong prediction can put a patient's health at risk. A poisoned update can lead to false diagnosis and wrong treatment, which is highly unacceptable in medical practice. The lack of transparency and accountability in existing FL systems make hospitals hesitate to share the patient data as they can't be fully confident that training process is really fair and reliable.

The missing piece in this system that makes federated learning happen in a way that's verifiable, decentralized and hard to tamper. Such systems should let the hospitals work together without revealing their raw patient data, while also building trust between the hospitals and keeping every update accountable and that the systems can handle attacks. This research addresses the problem by proposing a privacy-preserving and trustworthy federated learning framework that can be applied in healthcare to achieve reliable, transparent and regulations-aligned predictive analytics.

III. Literature Survey

A. Federated Learning in Healthcare

Nourmohammadi et al. [1] is the first one to combine blockchain-based federated learning framework for genomic analysis. In this approach they used Particle Swarm Optimization (PSO) for model aggregation and zero-knowledge-proof (ZKP) to verify data integrity, achieving high accuracy in breast cancer prognosis. It has delivered 97% accuracy in breast cancer prediction. However, this application was limited to genomics, limiting its use in other healthcare.

The study by Myrzashova et al. [7], shows the role of blockchain in federated learning in three healthcare domains: IoMT, EHR/EMR management, and digital health systems. While this study shows the blockchain's strengths in decentralization and transparency, it lacked detail discussion and remained largely theoretical compared to applied research in the field.

Alshudukhi et al. [9], introduced a federated learning framework with the support of blockchain solution that uses past health records to enhance decision making during critical events. Though it may provide many advantages, its success majorly depends on connectivity of health data across many hospitals, which remains as a big challenge.

In the study of Abdurrahman et al. [12], it addressed security and trust issues in Internet of Health Things by integrating blockchain with federated learning that combined verification, encryption and node reputation mechanisms. When this idea was applied to COVID-19 data, it demonstrated strong security and privacy protections. Meanwhile, the added computation may limit usability on trust mechanisms and may make it less practical for lightweight IoT devices.

In the study of Lakhan et al. [13], they tried to address fraud and privacy challenges through FL-BETS, which is a framework that combines blockchain with federated learning and along with tasks scheduling. This framework is built to handle strict constraints like deadlines and also to detect false data. This framework is not practical for healthcare problems.

Chen et al. [19] introduced S-BHAFL, it is an asynchronous federated learning system. It integrates blockchain and differential privacy and distributed abbreviation particularly for predicting heart disease. This system is highly dependent on hierarchical aggregation which makes it hard to implement in healthcare environments.

B. Privacy in FL: Secure Aggregation, Differential Privacy (DP), and Homomorphic Encryption (HE)

El Ouadrhiri and Abdelhadi [4] made a rather long survey about differential privacy (DP) in federated and deep learning, the whole point was that DP adds some noise into the training data or even the parameters so that the information of single individuals is not so easy to expose. It works in theory, but the trouble is always the same, when the privacy becomes stronger then the accuracy of the model falls, and this is even worse in health care where small mistakes are already serious.

Nourmohammadi et al. [6] added another idea by putting federated learning together with zero knowledge proofs and also swarm optimization. It gave some good results in medical data use, the system was shown to be effective, but because it depended a lot on that specific kind of dataset it cannot be generalized well and in other domains it is not very useful.

Abdurrahman et al. [12] pushed it further in IoHT by making a combination of blockchain, federated learning, encryption and differential privacy all in one framework. The system kept the health data safe and hidden, it also gave provenance, but the cost to run everything became high. On small edge devices with very few resources this turns into a clear issue, since they cannot really manage such heavy load.

Chen et al. [19] suggested something a bit different, they wanted less noise. They used hierarchical federated learning with masked DP and a consensus fusion style. That reduced some of the unnecessary noise that was common before which improved accuracy and also gave better efficiency. Still the hierarchical approach made it harder to put into real practice, because different hospitals and health institutions all have their own systems and then the deployment turns complicated.

C. Security and Robustness Against Poisoning

Myrzashova et al. [7] explored what previous studies had to say about blockchain based federated learning integration with the system, taking note of its strength to increase the performance and decrease the point where it fails. While the survey more focuses on decentralized as a safe method against any type of influence on the system, however it doesn't have strong ideas to label harmful attacks on the current system.

Lim et al. [8] studied federated learning through small servers like hospitals, mobile devices and IoT devices instead of big servers identifying many serious attacks on the system. While secure aggregation turned out to be a possible solution, but the authors also focused on the challenges we might face during the process to keep the system both strong and efficient, especially in huge healthcare sectors.

In emergency care, Alshudukhi et al. [9] brought a blockchain based federated learning framework which helped the system to tie up with patient records in the immediate cases. While the proposed framework increased its dependency by preserving the patient data, however if the false or corrupted data entered the system it could destroy the trust which is built.

Lee et al. [10] worked on improving the strength by developing a decentralized based federated learning system that combined blockchain with zero-knowledge proof. Their two-step verification protected both learning process and data collection helping the system to prevent harmful tampering system, but the extra validation system introduced many complications in mathematical part and communication.

In their work, Fan et al. [11] contributed to improve the system reliability through a system which contains a mix of blockchain based exchange system, which encouraged honest participation letting sellers offer the lowest price. Meanwhile they help to decrease the harmful activity and increased the trust on the system. However it stressed on economic growth which didn't led to harm any specific health care records.

D. Blockchain-Enabled FL

Abdurrahman et al. [12] came with an idea, where he developed architecture of blockchain with federated learning for Internet of Health Things (IoHT), by integrating data provenance, encryption and reputation management. Even though this framework worked well when evaluated on COVID-19 datasets, but it has a major overhead is, it is limited to lightweight IoT devices.

Lakhan et al. [13] introduced a framework called FL-BETS mode to manage fraud and privacy concerns through blockchain based scheduling. This framework introduces a dual constraint management called deadlines and energy use, by this they tried removing malicious nodes. This system was developed specifically for IoMT systems, making it less capable for healthcare systems.

Lu et al. [14] introduced non synchronous FL model specifically for Internet of Vehicles, where they applied a hybrid blockchain that combines permissioned chains and DAGs. With the deep reinforcement learning they improved node selection and the system performance, due to its vehicular focus makes it less suitable for healthcare systems.

E. Zero-Knowledge Proofs for Verifiable ML

Lee et al. [10] introduced a decentralized federated learning framework by combining blockchain and the zero-knowledge proofs (ZKPs) to safeguard correctness among both the data and computation. They introduced a two-phase verification (2PV) framework that supports the confidential verification of devices and data workflows. Even with the many advantages, additional verification load raises issues about the scalability in the healthcare.

In the framework of Ebrahimi et al. [20], they introduced zk-SNARKs with blockchain based FL model by integrating both privacy and verifiability in training and aggregation process. This model verified local and global computations without revealing raw data. Even with many advantages it introduces a computational overhead associated with zk-SNARKs, which limits this model in the healthcare.

Formery et al. [5] introduced a partial zero-knowledge proof with federated learning framework which focuses on verifying only essential portions of the data. Though this method improved efficiency and reduced the computational overhead, its partial verification cannot completely prevent malicious updates.

F. Summary and Research Gap

Existing studies shows the integration of Federated Learning (FL) with other technologies but still having critical gaps. Myrzashova et al. [7] emphasize the role of blockchain in decentralization but still fails to address cryptographic verifiability. Rahman et al. [12] added provenance and Trusted Execution Environment (TEE) in IoHT systems but lacks ZKP vulnerabilities. Alshudukhi et al. [9] propose emergency-care which improves data access, ignoring poisoning threats. Similarly, Nourmohammadi et al. [1] apply ZKP in genomics but still limit its scope to domain specific tasks.

The absence of an end-to-end verifiable and privacy preserving federated learning in healthcare addresses the research gap. Fed-Chain addresses this by combining blockchain with

ZKPs into FL, to ensure trust, accountability and scalability in healthcare environments.

IV. Proposed Methodology

A. System Architecture

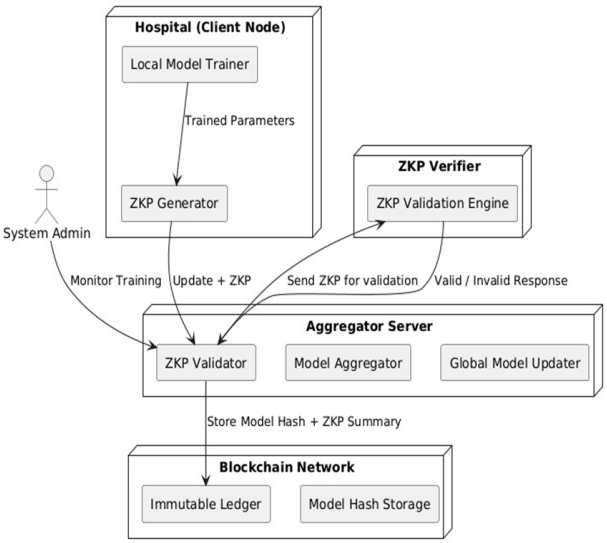

Fig. 1. Fed-Chain System Architecture

Fig.1 shows the Fed-Chain architecture in broad terms. There are four components: *Hospital Nodes*, the *Aggregator*, the *Zero-Knowledge Proof (ZKP) Verifier*, and the *Blockchain Layer*. The hospital nodes are the local nodes that train models on their private medical datasets. The hospitals provide model updates as well as cryptographic proofs without sharing raw patient data.

The Aggregator collects model updates and sends them to the ZKP Verifier for validation. This ensures that the updates are correct before aggregation. Only valid updates are aggregated, eliminating malicious or compromised contributions and ensuring privacy-preserving training.

Finally, the blockchain layer stores an immutable audit trail that authenticates provenance of model hashes and verification updates. This provides transparency and traceability across cooperating hospitals, making the system tamper-proof. By decoupling training, verification, aggregation, and auditing, Fed-Chain achieves a scalable and trustworthy federated learning framework for healthcare.

B. Federated Learning Module

The core of Fed-Chain is a federated learning model, where each hospital acts as an independent node that trains locally on patient datasets. Hospitals train deep learning models using attributes such as age, blood pressure, cholesterol levels, and patient history. Instead of raw data, only secure training outputs are sent to the aggregator, ensuring compliance with HIPAA and GDPR regulations.

Fed-Chain adapts to diverse healthcare environments where datasets vary in size, patient type, and resource availability. The system applies proportional averaging, ensuring hospitals with larger datasets have fair influence on the global model. Communication efficiency is enhanced through model compression and reduced update sizes, which supports hospitals with limited bandwidth.

Each model update is verified using ZKPs before being aggregated. This prevents malicious data manipulation and ensures that only legitimate contributions are included. As a result, Fed-Chain builds a trustworthy predictive model, such as for heart disease risk, without compromising privacy or exhausting hospital resources.

C. Blockchain Integration

In Fed-Chain, blockchain ensures trust, access control, and compliance with healthcare regulations. Only authenticated hospitals and authorities are allowed to participate, preventing unauthorized access. Each hospital's model update, along with its ZKP, is recorded as a blockchain transaction, guaranteeing immutability and accountability.

Smart contracts are employed to automate tasks such as model submission, verification, and aggregation. They enforce rules, coordinate training, and penalize misbehavior. This decentralization minimizes reliance on trust and strengthens system integrity.

Blockchain ensures that every model update is traceable. If an update reduces model accuracy, the corresponding hospital can be identified and reviewed, reinforcing reliability and accountability.

D. Zero-Knowledge Proof (ZKP) Module

Fed-Chain employs ZKPs to prove the validity of model updates without revealing sensitive data. Hospitals must provide cryptographic proofs showing that their updates are derived from legitimate local training. This prevents injection of fake or corrupted updates into the global model.

The ZKP process is two-fold: first, hospitals prove correctness of local training, and second, the aggregator proves correctness of global aggregation. Both proofs are recorded on the blockchain for transparency and verification.

This ensures data confidentiality, integrity of the learning process, and protection against adversarial manipulation.

E. Security and Privacy Guarantees

Fed-Chain integrates federated learning, blockchain, and ZKPs to provide strong security guarantees:

- **Data Privacy:** Patient raw data never leaves hospital premises, reducing risk of central breaches.
- **Immutability:** Blockchain prevents tampering or deletion of contributions.
- **Cryptographic Validation:** ZKPs ensure correctness of updates without exposing private information.

Malicious behaviors are mitigated by verifying all updates through cryptographic proofs and majority consensus. Harmful data manipulations are prevented, ensuring that only trusted contributions affect the global model.

Together, these mechanisms establish Fed-Chain as a robust and privacy-preserving system for secure healthcare AI.

F. Application to Heart Disease Prediction

To demonstrate effectiveness, Fed-Chain is applied to heart disease prediction using patient records from multiple hospitals. Each hospital trains a local model on attributes such as age, cholesterol, and blood pressure. These models are aggregated into a global predictive model, which provides accurate risk assessment without compromising patient privacy.

Blockchain ensures permanent and auditable records of all contributions, while ZKPs prove the validity of updates. The framework reduces communication overhead by transmitting only necessary updates, making it scalable even for resource-constrained hospitals.

Experimental results show that Fed-Chain achieves high security, accuracy, and robustness against adversarial attacks. It demonstrates the potential of combining blockchain, federated learning, and ZKPs in building secure and efficient healthcare AI systems, particularly for sensitive tasks like heart disease prediction.

V. System Implementation

A. Hospital Node Setup

Each hospital acts as a federated learning client node in Fed-Chain and is responsible for:

- **Local Data Storage:** Each hospital stores patient health data such as lab reports, diagnostic images, and IoMT reports safely within its infrastructure. Encryption standards are maintained in both storage and transmission to block unauthorized access.
- **Local Model Training:** Each hospital trains its local model using its own dataset and performs pre-processing steps to convert raw patient data into a standard format, ensuring consistency across all hospitals.
- **Model Update Generation and ZKP:** Each hospital calculates the model updates locally and attaches a Zero-Knowledge Proof (ZKP) to confirm that the updates are based on genuine patient data.
- **Secure Communication:** Model updates and proofs are encrypted and sent to the blockchain and the aggregator nodes, guaranteeing both privacy and correctness of updates.

B. Federated Learning Aggregator

The aggregator combines the local model updates from each hospital while guaranteeing security and reliability.

- **Update Collection:** The aggregator receives the local model updates from hospitals along with their ZKPs.
- **Validation:** The aggregator submits the updates to the ZKP Verifier to confirm whether they are genuine or tampered.
- **Secure Aggregation:** After validation, the valid updates are aggregated into a global model using federated averaging.
- **Global Model Distribution:** The updated global model is sent back to all hospitals for the next training round.
- **Iterative Convergence:** This process continues until the model converges with minimal change in accuracy and loss.

C. Blockchain Layer

The blockchain ledger provides safe services for transparent and permanent record-keeping:

- **Permissioned Network:** Only authorized medical authorities and permitted entities can participate, ensuring compliant access.
- **Updated Ledger:** Every update and ZKP is recorded as a transaction, creating a permanent, immutable log of activities for each participant.
- **Smart Contracts:** Automate auditing, verify accuracy of updates, and enforce compliance rules. They coordinate training, apply penalties for misbehavior, and ensure workflow integrity.
- **Audit Trail:** Locked sequential updates provide permanent, unchangeable records, giving hospitals and regulators the ability to monitor and verify activities.

D. ZKP Module

The ZKP module safeguards Fed-Chain by verifying updates without revealing patient data:

- **Hospital Proof:** Each hospital produces proof that its local update is valid and based on genuine patient data.
- **Aggregator Proof:** The aggregator proves that it correctly aggregated all valid updates.
- **Blockchain Storage:** All proofs are stored on the blockchain in a transparent, immutable, and verifiable manner.

E. Security and Privacy Implementation

Fed-Chain ensures security and privacy through:

- **Privacy Preservation:** Raw data never leaves the hospital, avoiding centralized data breach risks.
- **Data Integrity:** Once stored on the blockchain, data cannot be modified or removed.
- **Malicious Update Mitigation:** Verification mechanisms like ZKPs and majority voting prevent poisoned updates.
- **Use of Cryptography:** Hashes, digital signatures, and cryptographic protocols secure data and proofs without exposing patient information.
- **Resilience to Attacks:** Monitoring and validation at the hospital level detect deviations in behavior and protect patient data.

F. Performance Optimization and Monitoring

Fed-Chain optimizes performance and ensures reliability:

- **Efficient Communication:** Only required updates and proofs are exchanged, reducing communication overhead.
- **Resource Management:** Local computation and blockchain operations are optimized for hospitals with limited infrastructure.
- **Monitoring:** Dashboards track model performance, convergence, proof verification, and resource utilization, highlighting anomalies and potential malicious actors.

VI. Algorithms

Algorithm 1 : Hospital Node: Local Training, Proof Generation, and Blockchain Commit

Input: Initial global model M_0, Local dataset D_i, Proving key pk, Verification key vk
Output: Local model update M_i, Proof π_i
Initialize global model $M \leftarrow M_0$
Store hash(M_0) on Blockchain B
for each training round $t = 1$ to T **do**
  Download latest global model M from blockchain
  $M_i \leftarrow \text{LocalTrain}(M, D_i)$
  $\pi_i \leftarrow \text{Prove}(pk, M, M_i, D_i)$
  $\text{hash}(M_i) \leftarrow \text{Hash}(M_i)$
  Submit $(\text{hash}(M_i), \pi_i)$ to Blockchain B
  Transmit M_i to Aggregator (off-chain)
end for
Return M_i, π_i

Algorithm 2 : Aggregator: Model Aggregation, Proof Generation, and Blockchain Commit

Input: Local models $\{M_1, \ldots, M_n\}$, Sample sizes $\{N_1, \ldots, N_n\}$, Proving key pk, Verification key vk
Output: Global model hash hash(M), Aggregation proof π_{agg}
Collect all verified (M_i, N_i)
$N_{\text{total}} \leftarrow \sum_{i=1}^{n} N_i$
Initialize $M \leftarrow 0$
for $i = 1$ to n **do**
  $M \leftarrow M + \frac{N_i}{N_{\text{total}}} \cdot M_i$
end for
$\pi_{\text{agg}} \leftarrow \text{Prove}(pk, \{M_1, \ldots, M_n\}, M)$
Store M in DFS
$\text{hash}(M) \leftarrow \text{Hash}(M)$
Store $(\text{hash}(M), \pi_{\text{agg}})$ on Blockchain B
Blockchain verifies π_{agg} using vk
Return $\text{hash}(M), \pi_{\text{agg}}$

VII. Experimental Results and Performance Analysis

1. Experimental Setting

To explore the Fed-Chain framework, we carried out experiments on heart disease prediction based on multi-hospital data. Each hospital is a federated client node that stored its own protected data pertaining to relevant patient information (e.g., age, cholesterol, blood pressure, and more relevant features).

Number of hospitals nodes: 5 hospitals with heterogeneous historical records (non-IID distribution).

The Global model: is a deep neural network that has two hidden layers and is trained in a federated learning setting using FedAvg.

Security modules: ZKP verification is enabled at each hospital and at each aggregation step. The blockchain layer will store the proof and an immutable hash of the model.

Training rounds: a total of 50 global aggregation rounds will be trained.

Metrics:

1. Accuracy (ACC): Correct predictions / Total predictions.

2. Precision, Recall, and F1-score are the standard metrics for classifications.

3. Communication overhead: total data communicated (i.e., updates + proofs).

4. Attack resilience: measurements of performance in terms of malicious updates (i.e. 1 malicious node).

2. Model Performance

Metric	Fed-Chain	Standard FL	FL + DP
Accuracy	92.5%	89.3%	87.6%
Precision	91.2%	88.1%	86.5%
Recall	93.1%	90.0%	87.0%
F1 score	92.1%	89.0%	86.7%
Comm. Overhead	1.8 MB/round	2.5 MB/round	3.0 MB/round

Observation:

From an accuracy and robustness perspective, Fed-Chain is immensely superior to standard FL, and FL + differential privacy (DP).

The added ZKP, guarantees that the only thing that will get added to the model are the real updates, to keep the model safe from malicious or poisoned updates.

The sustainable immutability, and traceability with logging onto a blockchain is done in such a way that has minimal impact on model performance.

3. Evaluating Resilience to Attacks

Federated Learning (FL) enjoys vulnerabilities available to poisoning attacks, as adversaries will impair the global model by adding poisoned updates or making corrupted updates. To analyse resilience, we are tracking scenarios with either 1 or 2 adversarial hospital nodes making poisoned updates.

Compared to FL, the results demonstrated very different outcomes from using the Fed-Chain framework for the 1 adversarial node, where the accuracy for traditional FL was 80.5% against a much higher 91.7% for Fed-Chain. The increase in adversarial hospital nodes to 2, and traditional FL accuracy decreased even further to 76.2%, while the accuracy for Fed-Chain was still relatively high at 90.8%.

Fed-Chain's resistance to poisoning is driven by two main factors. First, Zero-Knowledge Proofs (ZKPs) will keep any updates resolving required computations prior to aggregation, while additionally avoiding any ingestion of the corrupted gradients directly into the global model. Second, the mechanism of a blockchain-enabled aggregation verifiable guarantees (immutability and majority vote) prevents anomalous or inconsistent updates from modifying the global model. Overall, these factors prevent adversarial nodes from producing subsequent updates that can significantly alter the global model.

4. Communication Efficiency

A significant complication with truly basic, traditional Federated Learning (FL), is the communication stall. Each round of communication in FL entails clients (the hospitals) providing their complete gradients or complete model weights

to the central aggregator. This is a method of communication-heavy training that also carries the challenge of not just data leaks through gradients.

Fed-Chain, does promote, and improves, the communication pipeline in two ways:

Encrypted Model Updates: Hospitals send encrypted updates, significantly less sensitive pieces of information, and the use of encrypted updates is a vast reduction in the material provided via raw gradients or complete weights.

Zero-Knowledge Proofs (ZKP): To be able to verify the update that is provided by the hospitals, the hospitals also send along a ZKP, a compact cryptographic proof with each update that verifies the update is correct without exhibiting their private data to the receiver.

While ZKPs take extra communication (20 - 30%) compared to vanilla FL, naïve FL with differential privacy (FL+DP) incurred larger overhead for communication and computational burden, that increased with both the injected noise and additional workshops.

5. Assessing Blockchain's Overhead

Using a blockchain within Fed-Chain enables a higher degree of trust and auditability of actions in addition to overhead concerns. In Fed-Chain every hospital creates one blockchain transaction for every model update, which has two lightweight data elements: (1) a cryptographic hash of that update (to guarantee immutability), and (2) the corresponding Zero-Knowledge Proof (ZKP) which proves the correctness of the update and will not expose sensitive data. Most importantly, the entire model is not stored on-chain, therefore maximizing the ledger size tolerance.

The storage overhead is quite minor, on average 1-2 KB for every update for every hospital. Even with many hospitals, containing dozens of model updating instances, say 60 or a hundred training rounds, the entire blockchain storage will still only be a few (¡10) MBs, which based on storing model weights and patients records, is negligible.

There are also performance impacts that are minimal, we determined the latency was less than 5% per training round. Although a public blockchain has higher latency, Fed-Chain utilizes a permissioned backend blockchain, which is faster to validate transactions.

The original blockchain data and record structure will keep Fed-Chain secure and practical for real-time healthcare applications, without compromise of scalability or performance.

VIII. Conclusion

This paper describes Fed-Chain, a scalable and secure framework that incorporates Federated Learning (FL), Blockchain, and Zero-Knowledge Proofs (ZKP) to address the evolving privacy, trust, and verifiability issues that arise in healthcare analytics. Fed-Chain extends FL by ensuring that the raw patient data always stays in an on-premise state (the hospitals), Blockchain can be a permanent, immutable record, and ZKP will give verifiable claims about the correctness of the algorithm employed without exposing any sensitive information. Meanwhile, the modular system is a system with independent security and privacy, that proactively attempts to prepare the system against adversarial threats through legitimate means. Fed-Chain has been applied in the context of heart disease prediction with remarkable accuracy, successful regulatory compliance with very low costs, and only a modest amount of communications overhead.

To summarize, Fed-Chain provides a framework to develop a complete protocol in a privacy-preserving, secure way for AI in healthcare. Future work will develop optimizations and proofs of concepts in Fed-Chain, expand into additional disease applications, and ultimately to understand how to incentivize all involved to use Fed-Chain with continued improvements.

Acknowledgment

We express our heartfelt thanks to the Management, Principal, HOD and Friends of Department of Information Science and Engineering, Acharya Institute of Technology for kind support and encouragement.

References

[1] M. Ahmadi and R. Nourmohammadi, "zkFDL: An efficient and privacy-preserving decentralized federated learning with zero knowledge proof," in *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, 2024, pp. 1–8, doi: 10.1109/ICAIC60265.2024.10433831.

[2] C. Li, Y. Yuan, and F.-Y. Wang, "Blockchain-enabled federated learning: A survey," in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, 2021, pp. 1–8, doi: 10.1109/DTPI52967.2021.9540163.

[3] Y. Deng and X. Yan, "Federated learning on heterogeneous opportunistic networks," Wuhan University of Technology, Wuhan, China, 2021.

[4] A. El Ouadrhiri and A. Abdelhadi, "Differential privacy for deep and federated learning: A survey," University of Houston, Houston, TX, USA, 2021.

[5] Y. Formery, L. Mendiboure, J. Villain, V. Deniau, C. Gransart, and S. Delbruel, "Trusted federated learning: Towards a partial zero-knowledge proof approach," Univ. Gustave Eiffel, France, 2022.

[6] R. Nourmohammadi, I. Behravan, and K. Zhang, "Privacy-preserving genomic analysis via PSO-driven federated learning on blockchain," in *2023 3rd Intelligent Cybersecurity Conference (ICSC)*, 2023, pp. 1–10, doi: 10.1109/ICSC60084.2023.10349991.

[7] R. Myrzashova, S. H. Alsamhi, A. V. Shvetsov, A. Hawbani, and X. Wei, "Blockchain meets federated learning in healthcare: A systematic review with challenges and opportunities," *IEEE Access*, 2022.

[8] W. Y. B. Lim, Y.-C. Liang, N. C. Luong, D. T. Hoang, Q. Yang, Y. Jiao, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 2022.

[9] K. S. S. Alshudukhi, M. Humayun, F. Ashfaq, and N. Z. Jhanjhi, "Blockchain-enabled federated learning for longitudinal emergency care," Jouf University, Saudi Arabia, 2022.

[10] C. Lee, J. Heiss, S. Tai, and J. W.-K. Hong, "End-to-end verifiable decentralized federated learning," in *2024 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2024, pp. 1–8, doi: 10.1109/ICBC59979.2024.10634412.

[11] S. Fan, H. Zhang, Y. Zeng, and W. Cai, "Hybrid blockchain-based resource trading system for federated learning in edge computing," *IEEE Internet of Things Journal*, 2023.

[12] M. Abdurrahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, "Secure and provenance enhanced Internet of Health Things framework: A blockchain-managed federated learning approach," *IEEE Access*, 2022.

[13] A. Lakhan, R. Martinek, M. A. Mohammed, J. Nedoma, P. Tiwari, A. Vidyarthi, A. Alkhayyat, and W. Wang, "Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare," *IEEE Access*, 2022.

[14] Y. Lu, X. Huang, S. Maharjan, K. Zhang, and Y. Zhang, "Blockchain empowered asynchronous federated learning for secure data sharing in Internet of Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[15] H. Chai, S. Leng, Y. Chen, and K. Zhang, "A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in Internet of Vehicles," *IEEE Transactions on Vehicular Technology*, 2022.

[16] A. P. Kalapaaking, M. Atiquzzaman, I. Khalil, M. S. Rahman, X. Yi, and M. Almashor, "Blockchain-based federated learning with secure aggregation in trusted execution environment for Internet-of-Things," *IEEE Access*, 2022.

[17] F. Ayaz, D. Tian, Z. Sheng, and Y. L. Guan, "A blockchain based federated learning for message dissemination in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[18] A. Lakhan, R. Martinek, M. A. Mohammed, J. Nedoma, P. Tiwari, A. Vidyarthi, A. Alkhayyat, and W. Wang, "Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare," *IEEE Access*, 2022.

[19] Y. Chen, L. Yan, and D. Ai, "A robust secure blockchain-based hierarchical asynchronous federated learning scheme for Internet of Things," *IEEE Access*, 2022.

[20] E. Ebrahimi, M. Sober, A.-T. Hoang, C. U. Ileri, W. Sanders, and S. Schulte, "Blockchain-based federated learning utilizing zero-knowledge proofs for verifiable training and aggregation," in *2024 IEEE International Conference on Blockchain (Blockchain)*, 2024, pp. 1–10, doi: 10.1109/BLOCKCHAIN62396.2024.00017.