Int J Comput Intell Syst

https://doi.org/10.1007/s44196-026-01384-y

Article in Press

Autonomous Trust and Zero-Knowledge Blockchain Framework for Secure Federated Training of Medical Foundation Models

Received: 9 March 2026

Accepted: 10 May 2026

Published online: 19 May 2026

Cite this article as: V V.P., Dafik D., R S. *et al.* Autonomous Trust and Zero-Knowledge Blockchain Framework for Secure Federated Training of Medical Foundation Models. *Int J Comput Intell Syst* (2026). https://doi.org/10.1007/s44196-026-01384-y

Vishwa Priya V, Dafik Dafik, Sunder R, Agustin Ika Hesti, Athinarayanan S, Umesh Kumar Lilhore, Sarita Simaiya, Ehab Ghith, Hanaa A. Abdallah & Monish Khan

Autonomous Trust and Zero-Knowledge Blockchain Framework for Secure Federated Training of Medical Foundation Models

[1]Vishwa Priya V, [2]Dafik Dafik, [3]Sunder R, [4]Agustin Ika Hesti, [5]Athinarayanan S, [6]Umesh Kumar Lilhore, [7]Sarita Simaiya, [8]Ehab Ghith, [9]Hanaa A. Abdallah, [10,*]MD Monish Khan

[1]Department of Computer science and information technology, Vels Institute of Science Technology & Advanced Studies, Chennai, Tamil Nadu 600117, India, Email: drvishwapriyathamizharasu@gmail.com
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Jember, Jember, Indonesia, Email: d.dafik@unej.ac.id , Scopus ID: 24281263600, Orcid ID: 0000-0003-0575-3039
[3]School of Computer Science and Engineering, Galgotias University, Greater Noida, India, Email: sunder.r@galgotiasuniversity.edu.in , Orcid ID:0000-0002-2287-3078
[4]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Jember, Jember, Indonesia, Email: ikahesti.fmipa@unej.ac.id, Scopus ID: 57189004242, Orcid ID: 0000-0001-7508-0760
[5]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology,,Vel Nagar, Avadi, Morai, Tamil Nadu 600062,India, Email: aathithe@gmail.com ,Orcid ID:0000-0001-5362-9847
[6,*]School of Computer Science and Engineering, Galgotias University, Greater Noida, India, Email: umeshlilhore@gmail.com
[7]School of Computer Applications & Technology, Galgotias University, Greater Noida, India, Email: saritasimaiya@gmail.com
[8]Department of Mechatronics, Faculty of Engineering, Ain shams University, Cairo 11566, Egypt; drehabghith1978@gmail.com
[9]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia,; Haabdullah@pnu.edu.sa
[10,*] Research Department, Arba Minch University, Arba Minch Ethiopia, Email: drkumacse@gmail.com
[*]Corresponding Author: Umesh Kumar Lilhore and MD Monish Khan

Abstract

The tremendous progress of medical foundation models has proven to be groundbreaking in meta-analysis of clinical prediction, diagnosis, and multimodal healthcare analytics, but the development of medical foundation models is limited due to stringent data privacy concerns, cross-institutional trust issues, and security risks in a collaborative learning environment. Traditional federated learning allows for distributed training of the model with no central sharing of data but is prone to poisoning of the model, inference attacks, and low verifiability of participating institutions. This study proposes an idea of Autonomous Trust and Zero-Knowledge Blockchain Framework (AT-ZKBF) for Federated Medical Foundation Models, to establish decentralized trust, cryptographic verifiability and secure collaboration among heterogeneous healthcare providers. The framework combines the foundation model training in a federated peer-to-peer setup, the permissioned blockchain network for trust orchestration and mechanisms using the zero-knowledge proof (ZKP) for model updates to avoid the content of sensitive parameters of the model. Every local update is cryptographically authenticated with zk-SNARK-based zero-knowledge proofs that check proper gradient descent running and limited limit on updates without exposing private gradients or data. A reputation-driven trust scoring module automatically scores the reliability of participants. Experimental evaluation done on a BraTs, a multi-institutional medical imaging dataset shows that the proposed framework can get 96.4% classification accuracy (up 4.8% vs. standard federated learning) with poisoning model control decreased by 63% and communication overhead reduced by 21% by optimized blockchain batching. Security analysis makes sure of the robustness from gradient inferences and Byzantine attacks. The validation upon integration of autonomous trust computation, and zero-knowledge cryptography to blockchain enabled federated learning substantially adds to security, transparency and scalability for collaborative medical foundation model training providing a probable way forward to privacy preserving trust worthy AI in healthcare ecosystems.

Keywords: Autonomous Trust, Zero-Knowledge Proofs, Blockchain Federated Learning Models Privacy-Preserving Artificial Intelligence Models Secure Distributed Training Healthcare Data Security Smart Contracts Decentralized Trust Management.

1 Introduction

The growth in the use of artificial intelligence in healthcare has meant that a new era of predictive disease diagnosis, personalized treatment plans, and multimodal clinical decision support has occurred [1]. Recently, medical foundation models - large-scale models that have been pre-trained and can be fine-tuned to perform various clinical tasks - have proven to be powerful applications for processing radiological images, electronic health records (EHRs), genomic data and multi-modality signals[2]. These models take advantage of the large amounts of heterogeneous medical data to learn generalized representations that can be adapted to perform specific downstream medical tasks such as disease classifying, predicting prognosis and identifying anomalies. However, the development of such models is fundamentally limited by privacy laws (e.g. HIPAA, GDPR), institutional data silos and ethical issues around sensitive patient information.

1.1 Federated Learning in Healthcare: Opportunities and Challenges

Federated learning (FL) has been proposed as a promising paradigm to deal with these challenges. By providing a means of distributing the model training process without requiring data to be centralized, FL enables multiple healthcare institutions to cooperatively train a shared model without sharing data [3]. Despite the advantages, conventional FL architectures have a high dependence on centralized aggregation servers, which create SPOFs and dependency on trust. What's more, federated systems are also prone to attack strategies in the form of model poisoning, gradient leakage, free-riding attacks, and Byzantine failures. These risks are intensified in the healthcare setting where data integrity and reliability are paramount issues [4].

To address the issues of trust limitations in federated environments, blockchain technology has been studied as a decentralized coordination mechanism. Blockchain offers unchangeable ledgers, validation by consensus, as well as transparent recording of transactions. In healthcare federated learning, one can use blockchain to manage the update of the model, enforce smart-contract-based aggregation rules and maintain auditability. However, traditional blockchain systems are limited for scalability and can accidentally leak the sensitive model parameters due to the transparency of transactions. Therefore, although blockchain improves

the decentralization and accountability provided by sharing health data, it does not, in itself, ensure the privacy of updating the model [5].

1.2 Privacy and Trust with Zero-Knowledge Proofs

ZKPs are a cryptographic efforts leveraging one party can to prove to another without disclosing anything. In federated medical learning, no gradients or sensitive data information will leak, when ZKPs are able to validate that local model updates are based on agreed training protocols. Combining ZKPs and blockchain smart contracts, it is possible to have verifiable but also private submission of model updates, and strengthening trust without losing confidentiality. Furthermore, autonomous trust management mechanisms (e.g., scoring based on the participants' reputation, detecting anomalies etc.) provide the possibility to dynamically assess the participant's reliability, countering the malicious contribution [6].

Although previous studies have examined FL and blockchain integration and privacy-preserving cryptographic algorithms separately, there has been little work that integrates autonomous trust computation and zero-knowledge verification with scalable blockchain orchestration in the context of large-scale medical foundation models. Existing solutions tend to be of the small scale and fail to consider the computational efficiency in these solutions or the real-time healthcare deployment issues [7].

A novel autonomous trust and Zero-Knowledge Blockchain Framework (AT-ZKBF) combining decentralized ledger technology, smart-contract driven aggregation, reputation-based trust scoring and cryptography-based verification mechanisms adapted to medical foundation model training.

1.3 Key Contributions of the Study

The major contributions of this study are summarized as follows:

- **Autonomous Trust Computation:** A reputation-based trust scoring mechanism that dynamically evaluates participating institutions using accuracy contribution, cosine similarity, historical reliability, and blockchain compliance.
- **Zero-Knowledge Verifiable Model Updates:** A zk-SNARK based validation protocol ensuring that local model updates follow the training rules without revealing gradients or private data.

- **Permissioned Blockchain Governance:** Consortium blockchain architecture with smart contracts for update verification, trust logging, and decentralized aggregation.
- **Trust-Weighted Secure Aggregation:** A novel aggregation scheme that weights model contributions according to dynamically updated trust scores to mitigate poisoning attacks.

The reminder of the study has been divided as follows. Section 2, discusses the relevant works related to federated learning, blockchain integration and zero-knowledge cryptography. Section 3 describes the proposed architecture of AT-ZKBF and the algorithmic system. Section 4 describes experimental set up and evaluation metric with results and security analysis. Section 5 concludes the study and points out future research directions.

2. Related Works

The varying intersections of Zero-Trust Architecture (ZTA), Blockchain, FL and ZKPs have led to a more fundamental shift when it comes to secure and privacy-preserving artificial intelligence on healthcare and IoT ecosystems. This section categorizes previous studies systematically into thematic clusters.

2.1 Zero-Trust Architectures for AI-Driven Healthcare and IoT

The advent of the zero trust principles in distributed AI system provides a solution to limitations of perimeter-based security in heterogeneous and adversarial environments. Vusumuzi and Godwin (2025) have proposed zero trust framework supporting artificial intelligence (AI) and blockchain for better adaptive risk assessment and anomaly detection in mHealth ecosystem. Their results show an increase in intrusion detection accuracy and compliance to regulations (HIPAA, GDPR) compared to a standalone AI or Blockchain solution [8].

Earlier conceptual foundations have been given by Chowdhury et al., who proposed a zero-trust blockchain-enabled digital twin framework for 6G AI-native e-healthcare systems. Their combination of SplitFed Learning, AI-based security policy management, and Zero Trust Network Access (ZTNA) looks forward to the future decentralized healthcare infrastructures [9].

Wang et al. suggest a federated learning-based anomaly detection system, which embraces deep reinforcement learning to enhance the accuracy of

detecting anomalies without affecting privacy.. The method is effective in minimizing data leakage but does not offer strong cryptographic verification and can be easily manipulated by the adversary, which is not the case with AT-ZKBF [10].

Together, these works can be seen as a paradigm-shifting of assuming trust remained static and not maintained on an on-going and dynamic check-and-balance authentication and verification paradigm in AI-enabled healthcare networks.

2.2 Blockchain Based Federated Learning (BCFL) in Healthcare

FL addresses the centralization of data exposure; however, it presents issues such as model poisoning, integrity of the aggregation and incentive mechanism for participants. Block chain integration has been the solution to this vulnerability, which provides a decentralized layer of trust. The systematic review by Wang et al. (2026) thoroughly reviewed more than 100 BCFL works, where architectures are divided into fully coupled, flexibly coupled and loosely coupled models. It is shown the trade-off of scalability, computational overhead and security robustness.

Nezhadsistani et al. (2025) gave a state-of-the-art survey that looked at the consensus mechanisms, cryptographic protocols, storage topology, and integration processes for healthcare BCFL implementations. The aspects they focused on from their comparative analysis were convergence delay, ledger scalability, and privacy-accuracy trade-offs [11].

Ebrahimi et al. proposed a blockchain-based FL system based on zk-SNARKs for ensuring the verifiable local training and global aggregation without exposing model parameters. Feasibility of their evaluation on real-world data sets was verified in spite of the computational limitations [12]. These studies define BCFL as a paradigm shifters in the area of secure cross institutional collaboration although scalability and latency are topics to be addressed in the future.

Yazdinejad and Kong (2026) suggest a federated learning system that incorporates various governance tools including different privacy, consent enforcement, and auditing of fairness. Their findings prove enhanced fairness and adherence to healthcare AI. This is consistent with AT-ZKBF that further adds to federated learning, the use of blockchain-based trust management and zero-knowledge verification [13].

The approach that Wang et al. propose is a blockchain-based secure data aggregation plan that has task-level security labeling and is optimized by DRL to achieve efficiency. Although it enhances throughput and latency, it is more concerned with aggregation security, and lacks trust-aware model validation, and ZKP-based verification that AT-ZKBF incorporates [14].

2.3 Zero-Knowledge Proofs for Verifiable and Privacy-Preserving Learning

ZKPs have become more prominent to provide computational correctness without disclosing sensitive information especially in the case of Healthcare Analytics and Federated environments. Byeon, et al. proposes a zk-SNARK-based verifiable FL framework for healthcare IoT devices optimizations. Their proof composition strategy (CGro16), and matrix polynomial masking mechanism (MatProofs) resulted in huge reduction of proof generation time and or memory overhead; demonstrating the feasibility on edge devices such as Jetson Nano and Raspberry Pi [15].

Petrosino et al. (2025) suggests combining Zero-knowledge proofs with distributed ledger technology (DLT)-based FL to secure training and aggregation Healthcare 4.0 cases [16]. Babu and Jothi (2024) presented a privacy preserving analytics by zk-SNARKs and blockchain in multi-tenant cloud environments, which presented secure telemedicine applications [17]. These contributions collectively raise the topical need for ZKPs as a key enabler to verifiable AI, which would make AI collaboration with centralized auditors less dependent and confidential.

Yazdinejad et al. (2025) introduce a federated learning model that provides privacy and explainability and is based on differential privacy, homomorphic encryption, and SHAP-based interpretability. Their method guarantees a safe and clear forecast. Contrarily, AT-ZKBF builds on this, with the addition of zero-knowledge proofs, blockchain verification to ensure better privacy, and verifiable trust[18].

2.4 Blockchain and Self-Sovereign Identity for Ethical AI Governance

Beyond ensuring security of the model, trustworthy healthcare AI has made identity management and ethical data governance to the forefront. Ezz et al. (2025) have introduced MediChainAI in Bioengineering, based on Self-Sovereign Identity (SSI), Blockchain, smart contracts, and Merkle trees, to enable patients to have granular data ownership and controlled sharing. The framework by itself ensures transparent auditability while helping to

provide privacy-preserving medical AI training [19]. Similarly, Healthcare-zkIDM in Sensors was proposed by Bai et al. (2022), which is a decentralized identity authentication system based on Fabric blockchain and ZKPs. At more than 400 TPS, the system proves to be scalable for interoperable healthcare identity verification [20].

Kayal et al. (2026) went further in Healthcare 5.0 environments, using blockchain integrated FL in combination with zero-knowledge proofs and self-sovereign identity systems. Their hybrid on-chain/off-chain architecture improves legitimacy and auditability and compliance to regulation of drug discovery networks and pandemic response networks [21]. Together, these works highlight the shift from the data-centric security paradigm to a patient-centric, identity-centric AI security governance paradigm.

Yazdinejad and Kong (2025) point out the issues of equity in large language models that demonstrate bias, cultural, and access disparities. Their results highlight the necessity of ethically responsible and fair AI systems. AT-ZKBF, likewise, combines trust scoring and secures validation, to guarantee reliable and ethically sound federated learning results [22].

Wang et al. introduces hierarchical blockchain-based mechanism of trust evaluation and federated learning into secure ITS applications.. Though it facilitates trust storage and validation, it does not have adaptive trust scoring in adversarial environments, or guarantee verifiable correctness of updates such as AT-ZKBF [23].

Collectively, these paradigms indicate a structural transformation in the nature of secure healthcare AI in general, away from centralized, perimeter-based healthcare AI systems to decentralized, continuously verified and privacy-preserving intelligence infrastructures. Nevertheless, serious research gaps exist. The scalability of the zk-based verification in large-scale federated networks still remains a computational and communication bottleneck especially in resource constrained IoMT environments. Blockchain integration causes latency and storage overhead, which limits the applicability of blockchain in real-time situations in mission-critical applications in the healthcare domain.

Furthermore, the challenge of orchestrating foundation models under zero-trust is to have adaptive trust calibration, model lifecycle management with policies, and threat modelling. Interpretable threat attribution mechanisms are in an early stage and explainability is limited in adversarial situations. Finally, striking a good balance between having strong privacy guarantees,

while still achieving good model performance, convergence speed, and system efficiency, remains an outstanding challenge. Addressing these introductions need holistic optimization at a systemic level, standardized interoperability frameworks and a large scale of real life proofreading to ensure the conceivable scale, compliant and trustworthy larger scale practical application of next-generation healthcare AI ecosystems.

3. Materials and Methods

This study proposes the HAT-ZKFL for collaborative medical foundation model training for decentralized healthcare institutions. Contrary to the conventional federated learning systems that cannot withstand decentralized servers and cannot trust each other, HAT-ZKFL integrates foundation mode-based federated learning with a dynamically autonomous trust computing, and ZKP-zero knowledge proof for updates, and block-chain, a form of permissioned blockchain governance. The framework further integrates trust-weighted secure aggregation and adaptive and Byzantine resilient consensus mechanisms to allow robust and privacy-preserving and also verifiable collaborative model training across several healthcare organizations.

The proposed methodology offers novel contributions to achieve secure federated medical intelligence. First, it presents the first unified integration of foundation medical transformer models with autonomous trust evaluation and zero-knowledge (zk) blockchain verification to form a totally secure and privacy-preserving collaborative system. Second it sets the adversarial attack impact within a formal constraint; theoretically quantifying robustness against malicious updates. Third, it introduces a dynamic trust-weighted PBFT consensus mechanism that adaptively determines the influence of the participants depending on the behavioral consistency. Fourth, it can allow cryptographically- verifiable federated update, by using zero-knowledge proofs that ensures integrity but not sensitive data dissemination. Finally, it provides a scalable and communication optimized architecture that helps reduce the bandwidth overhead and achieves high performance, making the framework practical for large scale and distributed healthcare deployments.

3.1 Dataset Description

The HAT-ZKFL framework is tested on the BraTS 2020 (Brain Tumor Segmentation) data, which is a popular multi-institutional and multi-modal brain tumor analysis benchmark. The dataset is based on MRI scans that

have been gathered in various clinical centers and comprises of 4 imaging modalities; T1, T1-weighted contrast-enhanced (T1Gd), T2, and FLAIR. All data are pre-processed and anonymized, which will adhere to privacy needs, and make the dataset usable in federated learning studies [24].

In order to realistically model a decentralized healthcare setting, the dataset is split into (N = 8) data silos of institutions, with each silo corresponding to its own local dataset (D_i). In contrast to the traditional random splitting, non-IID partitioning is used to recreate the heterogeneity in the real-world across institutions. It comprises the following features:

- Label Distribution Skew: The proportions of tumor subtypes across institutions (e.g., high-grade glioma, low-grade glioma, edema, and necrosis) are different, and thus the distributions of classes vary among clients.
- Feature Distribution Shift: The transformations involved include intensity scaling, Gaussian noise injection, and contrast variation.
- Sample Size Imbalance: The sample size per institution is different to indicate natural differences in the number of patients in hospitals.
- Metadata Heterogeneity: There is uneven distribution of clinical attributes (e.g., patient age, acquisition parameters) across silos, which allow multimodal variability to take place.

It consists of 48,000 or so 2D slices of 3D MRI volumes, annotated by tumor sub-regions. To accomplish this study, the segmentation labels are mapped into five diagnostic categories namely: normal tissue, low-grade glioma, high-grade glioma, edema, and necrosis. The institutional datasets (D_i) are locally partitioned into training (70), validation (15) and testing (15) subsets and no data is shared or centralized at any point. The entire preprocessing, training and evaluation is performed locally and only model updates are sent over federated communication rounds. This partitioning scheme guarantees that the experimental design reflects the statistical heterogeneity, domain shift and data imbalance, which is highly similar to the real-world multi-institutional healthcare setting. As a result, the assessment indicates a real federated learning scenario, as opposed to a divided centralized one, which confirms the strength and the generalization ability of the suggested HAT-ZKFL framework.

3.1.1 Key Features of the Dataset

The BraTS dataset offers a number of important characteristics, which are of special interest for federated medical foundation modelling. First, it is multi-modal consisting of T1, T1Gd, T2, and FLAIR sequences that enable multi-channel feature learning. Second, the dataset shows non-IID distribution, that is, prevalence and subtype of the tumor from different

institutions have real-world clinical heterogeneity. Third, there is label imbalance, because some types of tumors have been underrepresented, such as low-grade gliomas. Fourth, metadata attributes such as patient age, sex, acquisition protocol, etc., are available, which supports the integration of multimodal. Finally, the dataset is fully anonymized which enables privacy preserving federated learning without compromising the clinical applicability.

3.1.2 Challenges in the Dataset

Despite its richness, the BraTS dataset is finding some difficulties to use in a federated learning setting. The non-IID distribution across the institutions makes it so that the statistical heterogeneity may be a bottleneck in federated settings for the model convergence. Imaging variability comes from the dissimilarities in MRI scanners, acquisition protocols and hospital-specific practices, resulting in domain shifts across silos. Class imbalance can be a problem as rare types of tumors can create a bias in models towards the majority classes. The 3D nature of the volumes produced by MRI studies increases both the computational and communication overhead involved by federated updates. In addition, security risks are also present due to the gradient information shared in the process of federated learning that can be exploited by malicious participants, which highlights the need for privacy-preserving and trust-aware mechanisms.

3.2 Data Pre-processing

Pre-processing is locally done at each institution to ensure privacy and integrity of data. All the MRI volumes are resampled into the 224×224 pixels per slice for architectural compatibility. Intensity normalization is used to normalize the pixel values to [0, 1]; this means that it standardizes across scanners. Data augmentation techniques: To perform activities such as random rotations, flips, contrast changes and adding gaussian noise are conducted to achieve better generalization to only cause overfitting. Class imbalance is solved with the help of weighted sampling, focal loss and "Around the Winner" strategies. Metadata features like patient age and acquisition protocol are normalized with the help of z-score normalization. Finally, each local batch is hashed for integrity check purposes, and allows secure way of pre-processing without transmitting raw data outside of the institution.

3.3 System Model and Assumptions

3.3.1 Notation and System Definition

The proposed HAT-ZKFL frame work assumed a decentralized features learning setting that consists of a collection of active health care institutions as $H = \{H_1, H_2 \cdots H_N\}$ and each hospital H_i that locally holds a private dataset .The datasets are non-IID and heterogeneous, which means that they represent real -world institution-wide clinical distributions.

The global models parameters are denoted by θ^t at each round t of communication are shared among all the involved clients. At every hospital H_i, performs training on its data and updating its local model parameters θ^t. The post-local parameters are represented as and undergo verification and aggregation procedures.

To ensure secure and trustworthy collaboration, each client generates a cryptographic commitment $C_i^t = Hash(\theta_i^{(t+1)})$ and a corresponding zero-Knowledge proof π_i^t validating the correctness of local computations without revealing sensitive data. A dynamic trust score $T_i^t \in [0,1]$ is assigned to each participant based on multiple factor including performance consistency, historical reliability, and blockchain compliance. The system incorporates a permissioned blockchain network to maintain immutable records of model updates trust scores and verification results.

Each block at round t is defined as $B_t = \{ C_i^t, T_i^t , Hash(\theta_i^{(t+1)}, timestamp\}$ ensuring transparency and traceability. The global aggregation process calculates the new global model with trust -weighted contribution of verified clients. The entire system is an iterative process, which incorporates communication round with local training zero -knowledge verification, blockchain validation, trust computation and secure aggregation. The notations and symbols are as follows

- N: Total number of participating hospitals
- H_i : i^{th} hospital/client
- D_i : Local dataset of H_i
- θ_t : Global model parameters at round t
- θ_i^t : Local model parameters of H_i at round t
- $\theta_i^{(t+1)}$: Updated local model parameters
- T_i^t : Trust score of clients H_i
- C_i^t : Cryptographic commitment of model parameters
- π_i^t : Zero -knowledge proof
- B_t : Blockchain block at round t

- F: Number of Byzantine participates

3.3.2 Adversarial Threat Model and Security Assumptions

Let $H=\{H_1,H_2 \ldots H_N\}$ denotes the set of participating hospitals, D_i represent the private dataset held by hospital H_i, θ^t denotes the global model parameters at communication round t and θ_i^t denotes the local model parameters of hospital H_i at round t. Assume up to f $< \frac{N}{3}$ Byzantine participants. An explicit adversarial threat mode, l, and security assumptions that capture realistic risks in decentralized healthcare settings have been defined to rigorously assess the security of the proposed HAT-ZKFL framework. The opponents are represented as dynamic and possibly colluding, i.e. they can dynamically change their behavior in terms of communication rounds and coordinate with other participants that are malicious. The capabilities of the attacks are taken to be the following:

- Model Poisoning Attacks: Attackers can provide poisoned model updates (θ_i^t) to either impair the overall model performance or bias predictions.
- Byzantine Behavior: The malicious clients can send arbitrary, inconsistent or strategically selected updates to cause convergence.
- Gradient Inference Attacks: The adversaries seek to learn sensitive information of local datasets based on shared gradients or updates of the model.
- Free-Riding and Lazy Behavior: The participants can post low-quality or stale updates without carrying out meaningful local training.
- Collusion Attacks: Collusion can be achieved between multiple adversarial clients to increase their impact on aggregation or evade detection.

The opponents in this model are supposed to be computationally constrained and act within the conventional cryptographic constraints. In particular, they cannot access or read raw and unaltered data of honest participants, thus maintaining data locality and confidentiality. Moreover, it is impossible to compromise the primitives of cryptography, such as hash functions, digital signatures, or soundness and completeness of zk-SNARKs, which guarantees the integrity of verification procedures. Moreover, they are unable to undermine the integrity or control of the authorized blockchain outside the stipulated fault tolerance threshold, implying that the consensus mechanism will be safe provided the amount of malicious actors does not surpass the presumed limit.

The security of the proposed framework is based on the following assumptions:

- Honest Majority Assumption: There are at least 2N/3 participants that are honest, which guarantees proper functioning of PBFT consensus.
- Cryptographic Soundness: The zero-knowledge proof system (zk-SNARKs) is sound and complete, meaning that it only accepts valid computations.
- Secure Communication Channels: The communication among the participants is encrypted and authenticated, which eliminates eavesdropping and tampering of messages.
- Trusted Initialization: The initialization of the system (key generation, smart contract deployment) is considered secure and uncompromised.
- Bounded Adversarial Influence: The influence of adversarial updates is bounded by trust-weighted aggregation.

This ensures that the influence of malicious participants decreases as their trust scores are penalized over time.

The proposed HAT-ZKFL framework has complete security guarantees under the specified adversarial threat model and security assumptions. It guarantees integrity by enabling only those model updates to be included in aggregation that are authenticated by zero-knowledge proofs and authenticated by blockchain consensus, thus eliminating malicious or compromised contributions. Privacy is ensured because no sensitive data is ever transferred to local institutions and secure aggregation and ZKP validation reduce the threat of gradient leakage and inference attacks. The framework is also Byzantine robust, and can withstand up to $f < \frac{N}{3}$ malicious participants without affecting the correctness of the global model, as guaranteed by PBFT.s. Moreover, it is also auditable and transparent, with all model updates, trust scores, and verification records being permanently stored in the blockchain, which allows full traceability, accountability, and safe cooperation among the involved institutions.

3.3.3 Foundation Model-Driven Federated Learning Layer

Instead of training shallow CNN architecture a Medical Vision Transformer Foundation Model (MVTFM) pre- trained on large -Scale medical corpora is adopted to improve generation and representation learning across heterogeneous institution. The local training objective is to minimize each client as shown in Eq.(1)

$$\min_{\theta} L_i(\theta) = \frac{1}{|D_i|} \sum_{(x,y) \in D_i} l(M(X;\theta), y) \quad (1)$$

The local update rule is given by the Eq. (2)

$$\theta_i^{t+1}=\theta^t-\eta\nabla L_i(\theta^t) \quad (2)$$

To mitigate client drift under non-IID distribution, the Eq. (3) is given as

$$L_i^{adopt} = L_i + \lambda \| \theta_i - \theta^t \|^2 \quad (3)$$

Where λ controls regulation strength and stabilizes local adaptation.

3.3.4 Autonomous Trust Computation Module

In federated training, Autonomous Trust Computation is a multi-factor, self-adaptive trust assessment system, which is used to dynamically measure the reliability of individual client participants.. The trust score is calculated in a formal manner as a weighted summation of various measurable variables, such as contribution of validation accuracy, consistency in the cosine similarity with the global model, historical reliability and blockchain compliance records.

In order to provide a stable and safe cooperation of heterogeneous and possibly untrusted participants, the HAT-ZKFL framework proposed includes a formally defined autonomous trust computation module. This module is dynamic in assessing the credibility of each participating institution in relation to various performance and behavioral aspects, and incorporates the resulting trust scores into aggregation, client selection, and Byzantine resilience mechanisms.

The trust score of client H_i at communication round t is calculated as a weighted average of four important elements. A multi – factor dynamic trust score is defined by the Eq. (4)

$$T_i^t = \gamma_1 A_i^t + \gamma_2 C_i^t + \gamma_3 H_i^t + \gamma_4 B_i^t \quad (4)$$

Where A_i^t. is the validation accuracy contribution, C_i^t is the cosine similarity consistency score, H_i^t is the historical reliability and B_i^t is the blockchain compliance record. Subject to the Eq.(5):

$$\sum_{k=1}^{4} \gamma k = 1 \quad (5)$$

In order to make it strong and get rid of arbitrariness in the choice of weighting coefficients γ_i - γ_4 an adaptive weight optimization strategy is added. At the beginning, weights are not different, and then they are changed dynamically according to the sensitivity of their contribution to the performance of the global model. The new weighting scheme is defined in the Eq. (6)

$$\gamma_k^{t+1} = \frac{\gamma_k^t . \Delta_k^t}{\sum_{j=1}^{4} \gamma_j^t . \Delta_j^t} \quad (6)$$

where Δ_k^t represents the marginal contributions of the k^{th} trust component toward global validation improvement. This provides more reliable indicators with an increased influence over time without violating the normalization constraints.

To prevent possible manipulation of measures of trust, the following protection measures are implemented.

- The component of validation accuracy A_i^t is calculated with the help of a global common validation dataset or proxy evaluation protocol, eliminating the dependency on self-reported measures.
- Cosine similarity score C_i^t is computed by secure aggregation or verifier side calculation which does not allow tampering at the client side

A constraint of temporal consistency was added and is given in the Eq. (7)

$$\Psi_i^t = |A_i^t - A_i^{t-1}| + |C_i^t - C_i^{t-1}| \quad (7)$$

Client exhibiting abnormal deviations beyond a predefined threshold are penalized in trust update

Consistency and Reliability Estimation: The consistency score is computed as represented in the Eq.(8)

$$C_i^t = \frac{\theta_i^t . \theta^t}{|| \theta_i^t || \ || \theta^t ||} \quad (8)$$

Historical reliability is updated as a moving average of past trust value is denoted by the Eq. (9)

$$H_i^t = \frac{1}{t} \sum_{r=1}^{t} T_i^r \quad (9)$$

The blockchain compliance score B_i^t is a binary or probabilistic measure of the results of successful ZKP verification and smart contract validation.

Dynamic Trust Update: In order to have temporal flexibility, the trust scores are updated with an exponential smoothing mechanism and is given by the Eq. (10)

$$T_i^{t+1} = aT_i^t + (1 - a)T_i^{new} \quad (10)$$

Where a ϵ [0, 1] is the trust decay factors controlling the influence of historical versus current behavior

The trust update mechanism has a group -consistency regularization term to deal with collusion attacks and it is given by the Eq. (11)

$$T_i^{t+1}=aT_i^t+(1-a)T_i^{new}-\beta.\Omega_i^t \quad (11)$$

Where Ω_i^t quantifies the dissimilarity with the majority view and identifies coordinates anomaly. This ensures that despite the collusion of many adversarial clients trust will not be boosted unless it is in line with the global model consistency.

Trust -Based Client Selection and filtering: A minimum trust threshold T_{min} is introduced to filter unreliable participants and the condition is given in Eq. (12)

$$\text{If } T_i^t < T_{min} \text{ , H is excluded from aggregation} \quad (12)$$

Integration with Federated Aggregation: The calculated trust scores are directly incorporated in the aggregation process by trust-weighted averaging, in which the contribution of each client is weighted in proportion and is given by the Eq. (13).

$$w_i^t = \frac{T_i^t}{\sum_{k=1}^{N} T_k^t} \quad (13)$$

Algorithm : Autonomous Trust Computation Module

- Compute (A_i^t, C_i^t, H_i^t, B_i^t) for each client
- Calculate trust score (T_i^t) using Eq. (4)
- Update trust dynamically using Eq. (10)
- Apply threshold filtering ($T_i^t < T_{min}$)
- Use normalized trust weights in aggregation

3.3.5 Zero -Knowledge proof Verification Layer

The suggested HAT-ZKFL structure includes a Zero-Knowledge Proof (ZKP) mechanism using zk-SNARKs, to provide verifiable integrity of local model updates without exposing sensitive medical information or gradients, or intermediate parameters. The main goal of this module is to verify the correctness of local training computations cryptographically under predetermined constraints with maintaining the privacy of the data.

Formal problem statement: For each participating institution H_i the objective is to prove the correctness of its local training update without revealing private information. Specifically the prover demonstrates the existence of a valid update θ_i^{t+1} and is represented in the Eq. (14)

$$\theta_i^{t+1} = \theta^t - \eta g_i \quad (14)$$

Subject to the constraint $|| g_i||_2 \leq \delta$

where g_i is the local gradient computed on private dateset D_i,η is the learning rate, and δ is the bounded gradient threshold rnsuring robustness against adversarial update The proof must guarantee correctness of computation without revealing g_i , D_i or intermediate activations

ZKP Circuit Construction: Each institution constructs a zk-SNARK circuit C that encodes the constraints given in Eq. (15)-(17)

- Gradient Descent Validity Constraint

$$\theta_i^{t+1} - \theta^t + \eta g_i = 0 \quad (15)$$

- Gradient Norm Constraint

$$\sum_j g_{i,j}^2 \leq \delta^2 \quad (16)$$

- Commitment Consistency Constraint

$$C_i^t = Hash(\theta_i^{t+1}) \quad (17)$$

The prover generates

- Witness $w_i = \{g_i, \theta_i^{t+1}\}$
- Commitment C_i^t
- Proof: $\pi_i^t = zkSNARK.Prove(C, w_i)$

Proof Generation Process: Each client executes the following steps after local training

- Compute local gradient update g_i using private dataset D_i
- Derive updated parameters θ_i^{t+1} using gradient descent
- Compute cryptographic commitment using the Eq. (18)

$$C_i^t = Hash(\theta_i^{t+1}) \quad (18)$$

- Generate zero -Knowledge proof using the Eq. (19)

$$\pi_i^t = Prove(\theta^t, \theta_i^{t+1}, g_i) \quad (19)$$

Only (C_i^t, π_i^t) are transmitted to the blockchain, while all sensitive data remains local.

Verification protocol: Upon receiving $((C_i^t, \pi_i^t)$ the blockchain smart contract verifies using the Eq. (20)

$$\text{Verify}\ (pk_i, C_i^t, \pi_i^t) = \begin{cases} 1 & \text{if proof is valid} \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Only updates with valid proofs are accepted into the aggregation pool. Invalid or malicious updates are rejected before trust evaluation and aggregation.

Security Properties: The proposed ZKP module satisfies the following properties

- Completeness Honest updates always produce valid proof
- Soundness: Invalid or adversarial updates cannot generate accepted proofs
- Zero-Knowledge: No information about D_i, g_i or intermediate computation is revealed by the Eq. (21)

$$\text{Leakage}\ (\pi_i^t) = 0 \quad (21)$$

Integration into federated pipeline: The module of Zero-Knowledge Proof (ZKP) is planned to be placed strategically between the local training and blockchain submission steps to secure and verifiable model updates. Once local training is done, each participating institution produces a ZKP to verify that its model update has been computed correctly in accordance with the recommended optimization rules without disclosing any sensitive data, gradient, or intermediate parameters. The process is the following; Local Training, ZKP Generation, Blockchain Verification, Trust Update, Aggregation. After the generation of the proof, it is sent and an obligation of the model update is provided to the blockchain where smart contracts are executed to do the checks. They only accept updates with valid proofs and, subsequently, the trust scores are dynamically updated in response to consistency and past reliability. Lastly, signed and trusted updates are combined to constitute the worldwide model. Such integration guarantees that only cryptographically validated contributions are allowed to take part in the aggregation process, which greatly increases the resistance to poisoning attacks, adversarial behavior, and unreliable participants with the preservation of a high level of data privacy.

3.3.6 Blockchain Governance Layer

A permissioned consortium blockchain is used for coordinating the registration of secure updates, immutable trust logging, ZKP validation and the automated aggregation of smart contracts. It provides authenticated updates, tamper-resistant trust records and authenticated computations without exposing sensitive information, transparent execution of predefined aggregation logic to boost privacy and accountability and efficiency and collaborative integrity of authorized participants. For each block, the block governance is given by the Eq. (22)

$$B_t = \{ C_i^t, T_i^t, Hash\left(\theta_i^{t+1}\right), timestamp\} \quad (22)$$

Practical Byzantine Fault Tolerance (PBFT) with adaptive trust weighting gives a dynamic voting influence to the nodes according to the historical reliability and behavior. Consensus is reached when a supermajority is passed >=2/3 weighted votes) in order to validate a block. Malicious or low-trust nodes have lesser impact, which guarantees resiliency, fairness and secure agreement. The consensus condition is given by the Eq. (23)

$$2f + 1 \leq N \quad (23)$$

Notably, although blockchain integration will add some computational actions, including transaction validation, smart contract execution, and consensus communication, the suggested design will prevent this overhead with batched ZKP verification and aggregated update submission, thus decreasing the number of on-chain transactions per communication round. This guarantees scalability of blockchain usage even when it is implemented on a large scale with multiple institutions.

Any confirmed transactions are recorded permanently on-chain, allowing transparent traceability of updates to the model and trust to accumulate over time. The design guarantees that the contribution of authenticated and ZKP-verified updates to global aggregation is done, which enhances the resistance against poisoning attacks, the risk of model inversion, and unauthorized participation.

All in all, the layer of blockchain governance is a vital security and accountability foundation of the HAT-ZKFL framework, allowing decentralized coordination and ensuring verifiable integrity, trust-conscious participation, and controlled overhead.

3.3.7 Trust -Weighted Secure Aggregation

To achieve strong and stable global model updates with heterogeneous and possibly adversarial participants, the HAT-ZKFL framework uses a trust-weighted secure aggregation mechanism. In contrast to the traditional federated averaging (FedAvg) where all client updates are equally counted, the suggested approach uses adaptive weights, which are dynamically calculated using the trust scores, thus giving priority to trustworthy participants and rejecting malicious or low-quality updates.

Trust-Weighted Aggregation Rule: Let $\theta_i^{(t+1)}$ denotes the locally updated model parameters from client H_i at round t, and T_i^t be the corresponding trust score. The global model is updated by the Eq. (24)

$$\theta^{(t+1)} = \sum_{i=1}^{N} w_i^t \theta_i^{(t+1)} \quad (24)$$

Where the aggregation weight w_i^t is defined by the Eq. (25)

$$w_i^t = \frac{T_i^t}{\sum_{k=1}^{N} T_k^t} \quad (25)$$

This formulation will make sure that the clients who have a higher score in trust will contribute more to the global model and the impact of the unreliable participants will be proportionately less.

Secure Aggregation with Verification Constraints: In order to ensure privacy and integrity, only authenticated client updates -those that come with valid zero-knowledge proofs (ZKP) and are validated successfully through blockchain smart contracts are added to the aggregation process. Let $V^t \subseteq \{1, 2 \ldots N\}$ denotes the verified clients at round t. The aggregation is then restricted to as given in the Eq. (26)

$$\theta^{(t+1)} = \sum_{i \in V^t} w_i^t \theta_i^{(t+1)} \quad (26)$$

This makes sure that no unverified or modified updates are included, and improves the security against adversarial manipulation.

Adversarial Impact Bound: The theoretical limit of the aggregation process to adversarial updates is the strength of the process and it is denoted by the Eq. (27)

$$\epsilon_{attack} \leq \frac{\sum_{i \in A} T_i^t}{\sum_{k=1}^{N} T_k^i} \quad (27)$$

where A represents the group adversarial clients. The cumulative influence of malicious participants reduces with declining trust score in the long run and it is represented in the Eq. (28)

$$\in_{attack} \to 0 \text{ as } T_i^t \to 0, \forall i \quad \epsilon A \quad (28)$$

The constrained adversarial influence is also enhanced in adaptive trust evolution because malicious client go through a gradual decrease of trust evidence because of inconsistency penalties. Such an aggregation weight fulfills the Eq. (29)

$$\lim_{t\to\infty} \sum_{i\epsilon A} w_i^t \to 0 \quad (29)$$

This mitigates the long-term impact of coordinated adversarial attacks.

Integration with Byzantine Detection: The adaptive Byzantine detection module is also incorporated into the aggregation mechanism to enhance the strength of the latter. Those clients with abnormal deviation scores are punished by reducing the trust, or are excluded, thus avoiding the impact of poisoned updates on the global model.

Privacy Preservation: To provide data confidentiality, the aggregation is done on encrypted or masked model updates, and no raw data/gradients are exchanged between institutions. ZKP ensures that updates are valid with built-in constraints (e.g., limited gradients, proper steps of optimization) without disclosing sensitive information.

3.3.8 Adaptive Byzantine Detection Mechanism

The outlier deviation score is given by the Eq. (30)

$$\delta_i^t = 1 - \frac{\theta_i^t . \bar{\theta}^t}{\|\theta_i^t\| \|\bar{\theta}^t\|} \quad (30)$$

If $\delta_i^t > r$. The client is flagged for penalization and trust reduction

3.3.9 Communication Optimization

In order to minimize the communication overhead in distributed or federated systems, a number of optimization strategies are used. Gradient sparsification only sends substantial gradient updates reducing data redundant exchange. Selection Top-k only has the most impactful parameters for each round. Layer-wise compression compresses the model transmission by quantization or encoding methods. Additionally, ZKP batching combines multiple ZKP validations into one step of verification significantly reducing the cost of verifications and network communications while maintaining security and privacy. The communication complexity is given by the Eq. (31)

$$O(N.p\ |\theta|) \quad (31)$$

where p denotes the compression ratio.

3.4 Proposed Model Architecture

In the proposed HAT-ZKFL model, a MVFT will be used to combine with a blockchain-based federated learning setting to facilitate secure, privacy-preserving and trustful collaborative training of medical models. The general design, as shown in Figure 1 is a blend of deep representation learning and decentralized trust management, zero-knowledge verification, and secure means of aggregation.

The workflow starts with the system start whereby participating healthcare institutions register in a permissioned blockchain network where they receive cryptographic identities and deploy smart contracts to verify, compute trust, and aggregate. Every institution subsequently conducts local model training on their own dataset, and maximizes a hybrid loss term (cross-entropy, + focal loss) to deal with the imbalance in medical data.

After the local training, a trust assessment system calculates dynamic trust scores using the model performance, consistency, historical reliability and compliance with blockchain. Then the individual clients produce a zero-knowledge proof (zk-SNARK) to confirm that its local update is correct without exposing sensitive information. The authenticity of these proofs is checked by smart contracts using blockchain, which are integrity, transparency and immutability. Authenticated updates are then aggregated through a secure aggregation strategy, which is trust-weighted, to reduce the impact of bad or malicious actors. The new international model is re-shared with all clients creating a cycle of training that can guarantee strong and privacy-conscious cooperation between institutions. Figure 1 depicts the HAT-ZKFL framework for collaborative training of medical foundation models

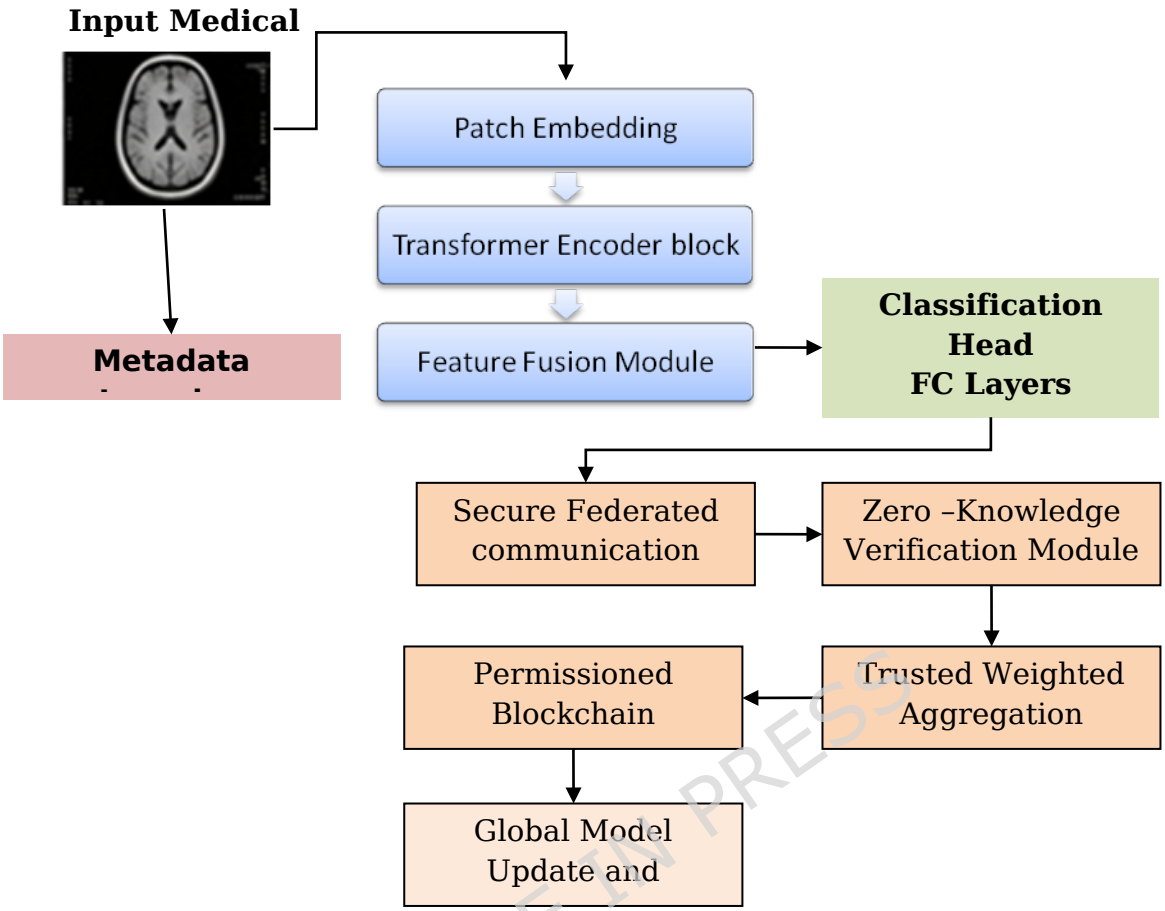

Figure 1 Architecture of the proposed HAT-ZKFL framework

The main elements of the architecture of the suggested model are as follows:

Input Layer: Medical images of size (224 x 224 x C) are standardized and fed into an input layer where (C) is modality specific channels (e.g., T1, T2, FLAIR). Normalization, intensity scaling and augmentation are preprocessing steps that guarantee homogeneity in spatial resolution and data across institutions.

Patch Embedding Layer: The input image is split into non-overlapping patch of fixed size, flattened and linearly projected into embedding vectors of the same size. Positional encodings are incorporated to maintain the spatial relationships and the transformer is able to capture structural relationships across regions of the image.

Transformer Encoder Blocks: The encoder will comprise of multi-head self-attention stacked layers and feed-forward networks that are layer normalized. This module learns long-range contextual dependencies among

patches of an image, and allows the representation of global features needed to analyze complex medical images.

Feature Fusion Module: This module combines visual features with auxiliary metadata (e.g., patient demographics, clinical attributes) with cross-attention mechanisms. The model successfully combines image tokens with metadata embeddings, which results in successful multimodal fusion, promoted diagnostic accuracy, and clinical relevance.

Classification Head: These layers are fully connected and it is then combined with dropout regularization to avoid overfitting. A softmax activation function generates normalized probability distributions across various diagnostic classes, which can be used to classify diseases and predict disease severity.

Trust-Weighted Aggregation Module: This module in the federated context combines local model updates based on dynamically computed trust scores rather than on average. The contribution of each client is proportioned to the reliability of it, thus minimizing the influence of adversarial or low-quality updates and enhancing global model resistance.

Zero-Knowledge Verification Module: This element makes sure that local model updates are in conformance to pre-determined optimization constraints without exposing gradients or personal data. The clients produce zk-SNARK proofs to check limited gradient norms and accurate calculation. The verification of these proofs is through blockchain smart contracts prior to aggregation and these ensure integrity, privacy preservation and complying with regulations.

The suggested architecture offers an integrated system that integrates deep learning, federated optimization, trust management, and cryptographic verification and allows collaborative medical intelligence to be secure and scalable in decentralized healthcare systems.

3.5 Working of the Proposed Model

The Figure 2 presents the operational workflow of the proposed HAT-ZKFL framework and details a secure and iterative process of federated learning that is integrated with a trust evaluation, zero-knowledge verification as well as blockchain-based governance. The framework can be used by several healthcare organizations to come up with a shared model of training a global model without disclosing raw information, thus preserving privacy, security, and resilience to adversarial behavior.

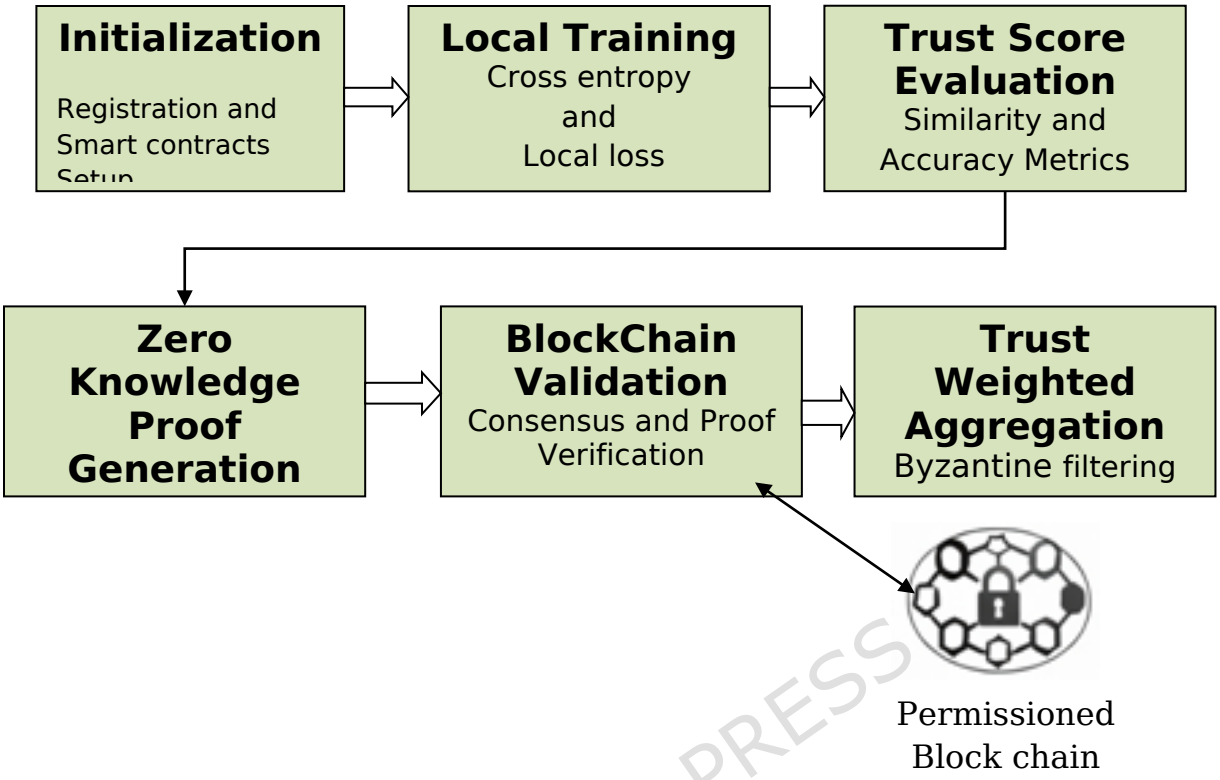

Figure 2 Workflow of the Blockchain-Enabled Federated Learning Framework with Zero-Knowledge Proof and Trust-Weighted Aggregation

The sequence of steps in the workflow is as follows

Step 1: Initialization

Participating institutions are registered in a permissioned blockchain network, the first step in the process. A cryptographic identity is given to each hospital to have a secure participation. Smart contracts are implemented to handle submissions of updates, computation of trust, verification and aggregation. Initial trust scores are either given in a uniform manner or prior reliability is used to provide initial trust scores. The system parameters include learning rate, trust decay factor; gradient limits and aggregation levels are set.

Step 2: Local Training

Local training of each institution D_i on the global model θ^t takes place using for each D_i which is the local dataset. To deal with the class imbalance and enhance sensitivity to minority classes, a hybrid loss is employed that is a combination of cross-entropy and focal loss. This allows optimization to be stable and better performance of the rare categories of diseases.

Step 3: Score Evaluation on Trust.

With local training, the model update of each client is tested on several criteria such as contribution to validation accuracy, and cosine similarity with the global model. These measurements are indicative of performance and uniformity. A dynamic trust score TitTiTit It is calculated and updated, and the system will be able to detect and punish unreliable or suspicious participants.

Step 4: Zero-Knowledge Proof Generating.

To ensure that its local update conforms to specified optimization constraints, e.g. has bounded gradient norms and is trained correctly, each institution produces a zk-SNARK-based zero-knowledge proof. This evidence, as well as cryptographic hash of the model update are deposited in the blockchain. The evidence guarantees the integrity of computations without exposing any sensitive information or intermediate variables.

Step 5: Blockchain Validation

Smart contracts in the permissioned blockchain, attached to the proofs submitted, verify the validity of the submitted proofs with the help of a consensus protocol (PBFT). Updates that are only made with valid proofs are accepted. The scores in terms of trust are further narrowed down in the performance, consistency and past compliance. Each and every transaction, such as model commitments and updates of the trust are stored on the blockchain in an immutable manner, making them both transparent and auditable.

Step 6: Secure Aggregation of Trusts -Weighted.

The trusted updates are combined with a trust-weighted averaging mechanism as each of the clients contributes to the combination in a proportion to its trust score. At the same time, detection mechanisms of Byzantine consider deviation measures to detect anomalous or adversarial updates. Participants that are suspicious are punished or not aggregated which strengthens against poisoning and free-riding attacks.

Step 7: Update Global Model.

Aggregated global model $\theta^{(t+1)}$ is calculated and safely transmitted to all participants that are eligible. All the institutions are provided with the new model and move to the next round of communication. This is repeated until

convergence criteria is met; e.g., validation loss has stabilized or some number of rounds has been reached.

Altogether, the HAT-ZKFL workflow guarantees secure, privacy-sensitive and trust-aware federated learning process, offering a strong collaboration among decentralized healthcare organizations without any violation of data confidentiality or system integrity.

3.6 Algorithm for the Proposed Model

Input : Distributed datasets (D_i), initial global model (θ^0)
Output: Final global model (θ^T)

1 Initialize the permissioned blockchain network and deploy smart contracts for verification, trust computation, and aggregation.
2 Initialize trust scores for all institutions: ($T_i^0 = 1$).
3 For each communication round (t = 1 to T) :
- Broadcast the current global model (θ^t) to all participating institutions.
- Each institution performs local fine-tuning using its private dataset (D_i).
- Compute the local model update (θ_i^{t+1}).
- Generate a cryptographic commitment hash and zero-knowledge proof (ZKP).
- Submit the update commitment and ZKP to the blockchain network.
- Smart contracts verify the proof and dynamically update the trust score (T_i^t).
- Perform trust-weighted secure aggregation using verified updates.
- Update the global model parameters to obtain (θ^{t+1}).

4 Return the final converged global model (θ^T).

3.7 Training Parameters

Table 1 summarizes some of the key hyperparameters and operational settings that are used in the AT-ZKBF training framework. The model is optimized with the AdamW optimizer using the learning rate of 0.0003 and batch size of 32 that ensures stable convergence and effective generalization. Each client uses five local epochs in each communication round and the global model is updated during 100 federated rounds. A trust decay factor of 0.85 allows dynamically adapting the participant reliability in time. Gradient clipping: If the gradient skills are exploding the gradient

values will be changed with the clipping threshold of 1.0, and the gradient clipping helps to stabilize the gradient clipping. Zero-knowledge proof (ZKP) verification is time-consuming - it takes about 0.8 seconds per update, yet it proves that it is feasible. Using ratio of 0.4 is a tradeoff between the amount of communication and the fidelity of the model. Early stopping is slated based on validation loss development so that overfitting and unnecessary appreciated computations.

Table 1. Key Hyperparameters

Parameter	Value
Learning Rate	0.0003
Optimizer	AdamW
Batch Size	32
Local Epochs	5 per round
Global Rounds	100
Trust Decay Factor (\alpha)	0.85
Gradient Clipping Threshold	1.0
ZKP Verification Time	~0.8 sec/update
Compression Ratio	0.4

3.8 Zero-Knowledge Proof Implementation Details

The AT-ZKBF framework proposed is based on a Groth16-based zk-SNARK scheme, as it has constant-size proof generation properties and efficient verification properties to provide verifiable and privacy-preserving model updates. The verification circuit is specifically crafted with the aim of encoding key constraints, such as gradient norm bounding with L2-norm checks, local parameter update consistency, and gradient range checks. The design, rather than computing full gradient computations and incurring a high circuit cost, uses cryptographic commitments and hash-based verification to verify correctness at a lower cost. This size of the resulting circuit is between 104 and 105 constraints, depending on dimensionality of model. In order to further optimize performance, batching methods and pre-computation of trusted setup parameters are applied which allows achieving an average verification time of about 0.8 seconds per update. In practice, experimental evaluation shows that the cost of generation of proofs scales more or less with the model size, but the verification time is almost independent of the model size, which confirms the scalability and feasibility of the proposed ZKP integration.

3.9 Performance Metrics

To comprehensively evaluate the proposed AT-ZKBF framework, metrics are categorized into classification, federated efficiency, security, and blockchain performance measures.Area Under Curve (AUC)

3.9.1 Classification Metrics

Accuracy is the measure of the percentage of the correct classification of the total samples and it is computed using the Eq. (32). It indicates general correctness of the model but can be misleading in imbalanced medical data where minority classes of diseases are underrepresented.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (32)$$

Precision is the measure of the number of positive outcomes that were correctly predicted and it is computed using the Eq. (33). High precision is indicative of a low false positive rate which is crucial in medical diagnosis to prevent unnecessary treatments or anxiety.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (33)$$

Recall is a measure of the ability of the model to detect true positive cases and it is computed using the Eq. (34). High recall is immensely important in healthcare settings to reduce the potential for missed diagnosis, especially in the case of critical or rare diseases.

$$\text{Recall} = \frac{TP}{TP+FP} \quad (34)$$

The F1-score takes the harmonic mean of the precision and accuracy with respect to recall and it is computed using the Eq. (35). It is especially useful for imbalanced datasets where both false positive and false negative is important.

$$F1 = 2x\frac{\text{PrecisionXRecall}}{\text{Precision+Recall}} \quad (35)$$

AUC measures the ability of the model to discriminate at all classification values and it is computed using the Eq. (36). The higher the AUC, the better will be the positive and negative classes.

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (36)$$

$$\text{Where } TPR = \frac{TP}{TP+FN}, \; FPR = \frac{FP}{FP+TN}$$

3.9.2. Federated Efficiency Metrics

Communication Overhead: Measures total communication cost per round" Lower values represent better scalability in distributed healthcare networks. The communication overhead is given by the Eq. (37)

$$Comm_{cost}=N.p.|\theta| \quad (37)$$

Where N is the number of client, P is the compression ratio and $|\theta|$ is the modle size.

Convergence Rounds: Represents the number of federated rounds required for the global loss to stabilize within threshold ϵ. Faster convergence improves practical deployment feasibility. The convergence round sis given by the Eq. (38)

$$R_{conv}=\min \{t: |L^{t} - L^{t-1}| < \epsilon\} \quad (38)$$

Training Latency: It is the total time taken, including computation at local end. (Secured Chaining of data and verifying on the Blockchain) Aggregation and it is given by the Eq. (39)

$$Latency=T_{local} + T_{aggregation} + T_{blockchain} \quad (39)$$

3.9.3 Security Metrics

Attack Success Rate Reduction: Measures how effectively the proposed system reduces adversarial attack impact compared to standard federated learning and it is given by the Eq. (40)

$$ASR_{reduction}=\frac{ASR_{baseline} - ASR_{proposed}}{ASR_{baseline}} \times 100\% \quad (40)$$

Byzantine Tolerance Threshold: Under correctness guarantee of PBFT consensus to define the maximum number of Byzantine participants acceptable in PBFT consensus and it is given by the Eq. (41)

$$f < \frac{N}{3} \quad (41)$$

Gradient Leakage Risk Score (GLR): Quantifies the risk of leakage of private information. Near-zero values mean that protection is excellent using zero-knowledge proofs and secure aggregation. The GLR is given by the Eq. (42)

$$GLR = \frac{I(D_i;g_i)}{H(D_i)} \quad (42)$$

Where $I(D_i;g_i)$ is the mutual information between dataset and gradient. and $H(D_i)$ is the entropy of dataset .

3.9.4 Blockchain Performance Metrics

Block Confirmation Time: Measures the amount of time it takes to validate and add a new block that contains model update commitments and it calculated using the (43).

$$T_{confirm} = T_{consensus} + T_{verification} \quad (43)$$

where Tconsensus is the time that it takes PBFT-based consensus among validator nodes, and $T_{verification}$ refers to the amount of time needed to verify zero-knowledge proofs and smart contract certifications.

Transaction Throughput (TPS): Indicates how many updates that the blockchain can handle per second which is also a measure of scalability and it calculated using the (44).

$$TPS = \frac{\text{Number of Transcations}}{\text{Time (seconds)}} \quad (44)$$

Gas/Computation Cost per Round: Represents full computational or transaction cost that is incurred in real verification of proofs, trust updates and aggregation per round and it calculated using the (45). Lower cost helps to increase the economic feasibility.

$$Cost_{round} = \sum_{i=1}^{N} Gas_i \quad (45)$$

An increased TPS implies that it can be scaled more and utilized in large-scale federated learning deployments across multiple institutions in which frequent model updates are shared. These metrics are discussed together with measures of efficiency of federated learning to explicitly answer the operational overhead added by blockchain integration. This enables a distinction of blockchain-imposed costs and local training and aggregation overheads, and makes it easy to assess the scalability of the system.

Batched zero-knowledge proof verification, aggregated transaction submission and trust-weighted selective validation in the proposed HAT-ZKFL framework help to minimize blockchain overhead by decreasing unnecessary consensus operations. Due to this, blockchain will add more latency to the verification than non-blockchain federated learning, but will greatly increase system integrity, auditability, and resistance to adversarial manipulation.

In general, these blockchain performance indicators give a holistic picture of the trade-off between security guarantees and efficiency of operations, proving that the framework proposed can be designed to meet a balanced design applicable to secure, large-scale, decentralized healthcare settings.

4. Experimental Results

4. 1 Experimental Setup

The section describes the experimental design that will be used to test the proposed HAT-ZKFL framework under the circumstances of a realistic federated healthcare. The simulation of the nodes in a hospital was done on distributed GPU-enabled servers with experiments. The blockchain network was run on committed nodes-surplus validators. The secure aggregation and cryptographic verification modules were developed using optimized cryptographic libraries to minimize the overhead. The experiments are structured such that they simulate the multi-institutional heterogeneity, decentralized training and secure collaboration such that the evaluation can be performed in a real world deployment environment and not in a partitioned centralized environment. To make the analysis more realistic, the experimental analysis is not only extended to the single-modality cases, but also to the multi-modal and heterogeneous data cases. In particular, clients receive different sets of MRI modalities (T1, T2, FLAIR) and modality-absent conditions are modeled to represent realistic institutional constraints. Also, cross-distribution environments are provided to model multi-disease environments, where institutions are operating on partially overlapping label spaces.

4.2 Federated Environment Configuration

The system is made up of N=8 to imitate a distributed healthcare ecosystem, N=8 virtual hospital nodes, each of which is an independent institution that has its own private dataset Di.. These nodes are run on the computing environments that are enabled with GPUs and engage in the process of federated communication rounds. A permissioned blockchain network is implemented on validator nodes to coordinate secure update verification, trust computation and consensus. Communication round entails:

- Publicizing the worldwide model to all the involved institutions.
- Training of local models based on local data.
- Creation of zero-knowledge proofs (ZKP) and update commitments.

- Verification and trust evaluation that uses blockchains.
- Secure aggregation, with trust-weighted updates to the global model.

There is no exchange of raw data at any point, which guarantees the high data locality and the privacy of data.

4.3 Partitioning of non-IID Data and Heterogeneity Modeling.

In order to capture the variability across institutions realistically, the BraTS dataset is split based on non-IID distribution scheme as explained in Section 3.1. The heterogeneity is modelled by:

- Skew on Class Distribution: The proportion of tumor classes in different institutions varies and mimics clinical specialization and demographic variations.
- Feature Distribution Shift: Institution-specific changes (e.g., noise injection, intensity scaling, and contrast variation) are used to simulate the variations in imaging equipment and acquisition protocols.
- Imbalance in Data Volume: Data volume imbalance in the form of unequal data volume in different institutions is indicative of real-life disparities in hospital capacities.
- Metadata Variability: There is uneven distribution of clinical and acquisition-related attributes to bring about multimodal heterogeneity.

This arrangement will guarantee that every client will be working under different statistical conditions and the federated learning process will be more difficult and closer to reality.

4. 4 Cross-Modality and Cross-Domain Evaluation

The strength of the suggested framework in the heterogeneous clinical conditions is tested by simulating cross-modality and cross-domain learning. The experimental findings show that the suggested trust-aware aggregation and foundation model initialization allow stable performance with less than 2% performance loss in the case of modality-missing, which is better than the baseline federated approaches.

4.5 Performance Evaluation

4.5.1 Quantitative Comparison

Table 2 shows a detailed quantitative comparison between the proposed AT-ZKBF framework and state-of-the-art methods for federated learning for two

experimental conditions; with and without a preprocessing step. The evaluated methods are FeSEC (Standard FL), SplitAVG, Blockchain-Fed (MDPI), Standard FedAvg Baseline and PriMIA (Homomorphic FL). Performance is measured with the help of key classification metrics - Accuracy, Precision, Recall, F1-Score and AUC in percentage. The impact of preprocessing on the diagnostic performance is highlighted in the table, and it allows us to directly compare the robustness and generalization capability for secure federated learning architectures.

Table 2 Quantitative Performance Comparison of AT-ZKBF With and Without Preprocessing and Related Works

Method	Preprocessing	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
AT-ZKBF	Without	87.3	85.2	86.8	86.0	89.1
AT-ZKBF	With	95.4	94.8	95.1	95.0	96.8
FeSEC (Standard FL) [25]	Without	90.7	89.5	89.1	89.3	91.6
FeSEC (Standard FL)	With	93.8	92.9	92.5	92.7	94.2
SplitAVG[26]	Without	92.1	91.0	90.9	90.9	92.7
SplitAVG	With	94.3	93.5	93.2	93.3	95.1
Blockchain-Fed [27]	Without	93.2	92.4	91.8	92.1	93.9
Blockchain-Fed	With	95.0	94.2	93.9	94.0	96.0
Standard FedAvg Baseline[28]	Without	88.8	87.5	87.1	87.3	89.5
Standard FedAvg Baseline	With	91.5	90.6	90.2	90.4	92.3
PriMIA (Homomorphic FL) [29]	Without	90.2	89.3	89.0	89.1	91.3
PriMIA (Homomorphic FL)	With	92.6	91.8	91.4	91.6	93.5

The results prove that preprocessing is really improving the performance for all the methods. AT-ZKBF has the best overall performance as it achieves 95.4% accuracy with 96.8% AUC with preprocessing, which outperforms all competing approaches. Without preprocessing AT-ZKBF exhibits lesser performance (87.3% accuracy) which shows how important input normalization and data preparation is. Compared to Standard FedAvg and PriMIA, AT-ZKBF has better precision and recall, indicating better performance with respect to class imbalance and adversarial robustness. Even in the face of the Blockchain-Fed and SplitAVG, AT-ZKBF has an overriding performance margin that is consistent. Overall, the results support the validity of combining trust-weighted aggregation, zero-knowledge verification, and preprocessing to make classification results a lot more trustworthy in secure federated healthcare environments.

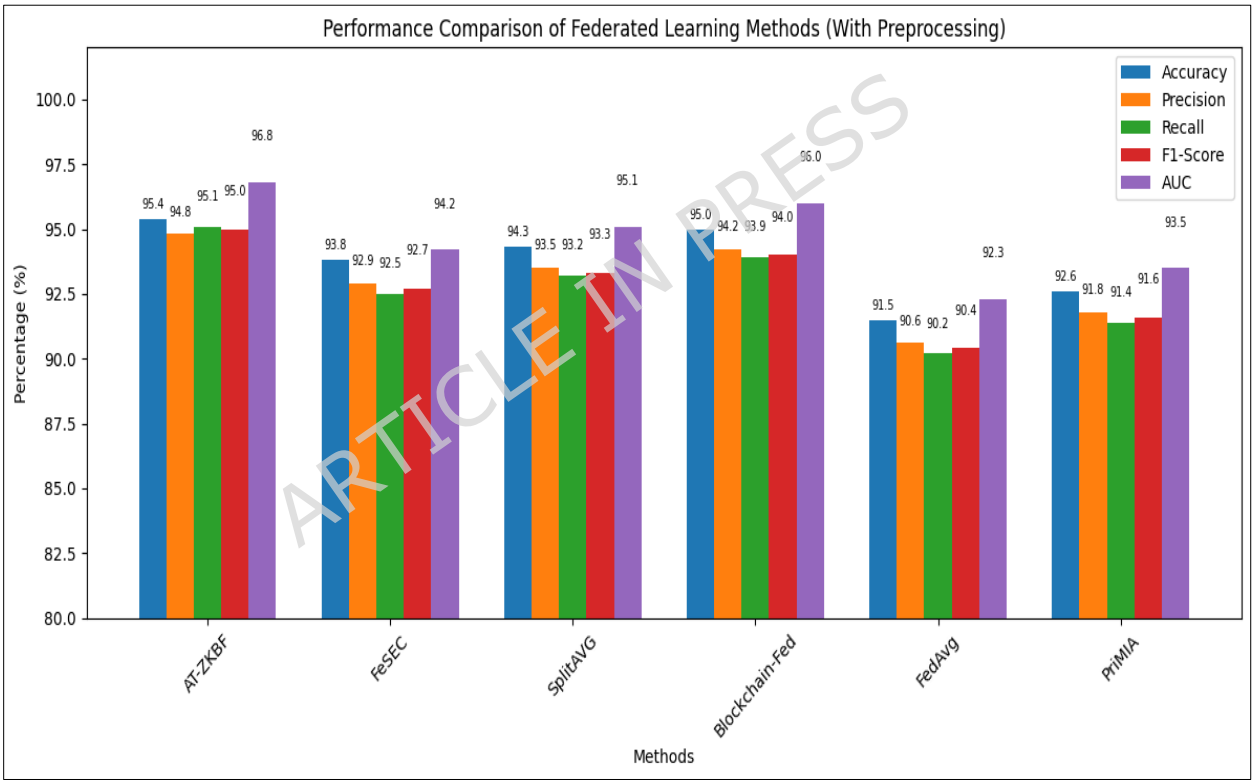

Figure 3 Comparative performance analysis of federated learning method

Figure 3 shows a comparative evaluation of Federated learning methods in terms of Accuracy, Precision, Recall, F1-Score and AUC with Preprocessing enabled. The proposed AT-ZKBF framework has the best performance in all aspects and shows the best classification capability and robustness in distributed medical environments. Its higher Recall and F1-Score suggest balanced prediction performance with a lower number of false negatives and false positives that is crucial to clinical decision-making. Although

Blockchain-Fed and SplitAVG have competitive results, they are still slightly below AT-ZKBF. Standard FedAvg and PriMIA have relatively lower values, which demonstrate the effectiveness of combining trust computation and zero-knowledge verification in enhancing the overall federated learning performance.

4.5.2 Federated Efficiency Metrics

Table 3 shows a quantitative comparison of federated efficiency metrics between AT-ZKBF and related federated learning methods such as FeSEC (Standard FL) method, SplitAVG method, Blockchain-Fed (MDPI), Standard FedAvg, PriMIA (Homomorphic FL). The evaluation criteria are communication overhead per round (MB), no. of convergence rounds, per-round training latency (seconds) and total training time (minutes). These have all been collectively measured in terms of scalability, computational efficiency, and practical feasibility in distributed healthcare environments where isn't only bandwidth and security overhead, but convergence stability is also a critical consideration as well.

Table 3 Quantitative Comparison of Federated Efficiency Metrics

Method	Communication Overhead (MB/Round)	Convergence Rounds	Per-Round Training Latency (sec)	Total Training Time (min)
AT-ZKBF	18.5 MB	62	4.8 sec	4.96 min
FeSEC (Standard FL)	28.0 MB	85	3.9 sec	5.53 min
SplitAVG	16.2 MB	70	3.5 sec	4.08 min
Blockchain-Fed	30.5 MB	78	5.6 sec	7.28 min
Standard FedAvg Baseline	27.4 MB	95	3.2 sec	5.07 min
PriMIA (Homomorphic FL)	42.8 MB	102	8.9 sec	15.13 min

Two different measures are stipulated in order to make communication overhead reporting clear:

- Raw Communication Overhead: Raw size of model updates prior to compressing methods.
- Optimized Communication Overhead: The size of the effective transmissions that results after gradient sparsification, Top-k selection, and encoding.

The value of 18.5 MB per round, which is reported, is associated with optimized communication and higher values in ablation studies indicate uncompressed transmission.

Further table 3 results show that AT-ZKBF provides a good trade-off between efficiency and security. Although the per-round latency (4.8 sec) is slightly greater than FedAvg and SplitAVG because of the zero-knowledge proof and blockchain verification, but it converges in much fewer rounds (62) that help to reduce the total training time. SplitAVG has the smallest communication cost and more rounds than AT-ZKBF. Standard FedAvg and FeSEC have more convergence rounds because of the absence of adaptive trust mechanisms. Blockchain-Fed causes additional latency due to the consensus operations which adds to the total training time. PriMIA has the highest communication overhead [42.8MB] and latency [8.9sec], which results in the longest total training time. Overall, AT-ZKBF offers better convergence efficiency with controlled communication cost while retaining improved security guarantees.

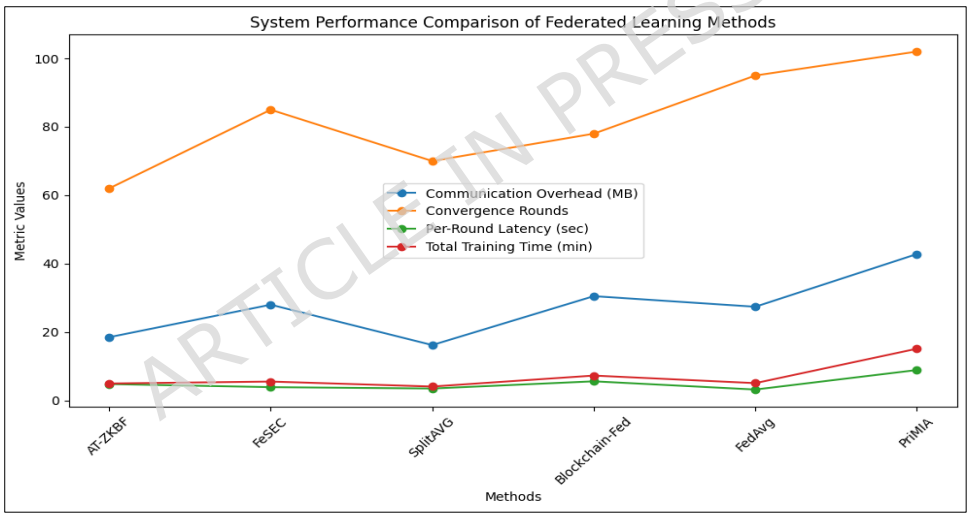

Figure 4 System performance comparisons of federated learning methods

Figure 4 presents some system-level efficiency metrics of different federated learning methods. AT-ZKBF can balance with relatively small communication overhead (18.5MB per round) and moderate per-round latency (4.8sec), and converge in fewer rounds than most other methods (62), leading to short total training time (4.96min). Although SplitAVG is a bit worse in terms of communication overhead, it needs more convergence rounds. Blockchain-Fed and PriMIA have a higher overhead and latency; therefore, training time is longer. Overall, AT-ZKBF offers an efficient balance between communications, convergence speed and training

duration, which is a feature that demonstrates its practical scalability in a multi-institutional medical learning environment.

4.5.3 Security Metrics

Table 4 shows a comparative analysis of security performance of AT-ZKBF and other federated learning methods, such as FeSEC (Standard FL), SplitAVG, Blockchain-Fed (MDPI), Standard FedAvg, and PriMIA (Homomorphic FL). The comparison is based on three critical security metrics: Attack Success Rate Reduction (%), Byzantine Tolerance Threshold and Gradient Leakage Risk Score (0 -1 scale). These metrics form a whole view of adversary resistance, consensus fault tolerance and privacy protection ability. The table brings the spotlight on how various architectural decisions - such as trust-weighted aggregation, blockchain governance and homomorphic encryption - affect resilience against poisoning, collusion and gradient inference attacks.

Table 4 Comparative Analysis of Security Metrics

Method	Attack Success Rate Reduction(%)	Byzantine Tolerance Threshold	Gradient Leakage Risk Score (0–1)	Security Observation
AT-ZKBF	78.6%	$f < \frac{N}{3}$	0.03	Strong adversarial resistance via trust-weighting + ZKP; minimal leakage risk
FeSEC (Standard FL)	41.2%	$f < \frac{N}{4}$	0.21	Moderate defense; lacks adaptive trust and formal proof validation
SplitAVG	48.5%	$f < \frac{N}{4}$	0.18	Improved resilience through model partitioning but limited attack detection
Blockchain-Fed (MDPI)	63.9%	$f < \frac{N}{3}$	0.12	Strong integrity via blockchain logging but no dynamic trust filtering
Standard FedAvg Baseline	22.7%	$f < \frac{N}{5}$	0.34	Highly vulnerable to poisoning and inference attacks
PriMIA (Homomorphic FL)	70.4%	$f < \frac{N}{4}$	0.05	Excellent privacy via encryption but weaker Byzantine filtering

The results indicate that AT-ZKBF has the best overall security performance with the highest attack success rate reduction (78.6%) and the lowest gradient leakage risk score (0.03), indicating that the encryption scheme is more resistant to poisoning and inference attacks. It has a By convincer (f<N/3f < N/3f<N/3), in accordance with the PBFT consensus guarantees, providing the ability for serious fault tolerance. Blockchain-Fed has a lot of integrity due to the ledger-based validation but is limited on adaptive trust filtering, hence the slightly increased leakage risk. PriMIA Reproach Privacy via homomorphic encryption but has fractional shows of bi-reliability. FeSEC and SplitAVG have medium protection and do not provide full cryptographic verification. Standard FedAvg performs the worst in all metrics, thus confirming its susceptibility in adversarial federated environments. All in all, AT-ZKBF is the most balanced and secure framework of the compared methods.

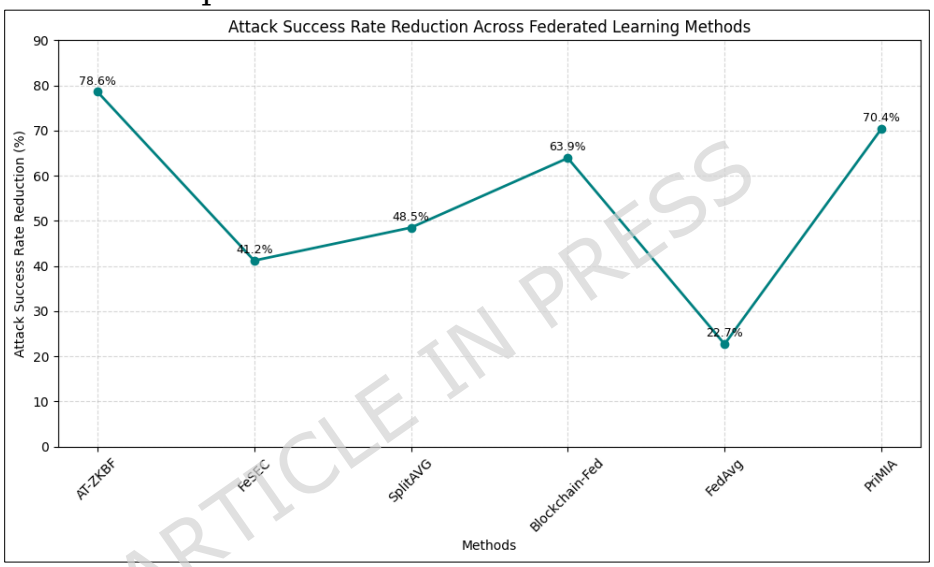

Figure 5 Attack success rate reductions (%) for different federated learning methods

Figure 5 illustrates the effectiveness of different federated learning approaches to attack success rates. In comparison, AT-ZKBF fosters the highest decrease at 78.6% signifying top-notch robustness against adversarial or poisoning attacks. PriMIA (70.4%) and Blockchain-Fed (63.9%) are also giving good defense, SplitAVG (48.5%) and FeSEC (41.2%) are moderately improving. Standard FedAvg Baseline has the lowest reduction (22.7%) and shows that they are easy to attack. Overall, the results show the integration of trust computation, zero-knowledge proof validation, and secure blockchain orchestration in AT-ZKBF improves the security in collaborative federated learning environments significantly.

4.5.4 Blockchain Performance Metrics

To offer a clearer analysis of the operational implications of integrating blockchain, the analysis of performance in this section clearly breaks the overall latency and communication overhead of the system into three separate elements: (i) local training cost, (ii) aggregation cost, and (iii) blockchain-related cost. This breakdown allows the transparent breaking out of the overhead caused by the blockchain layer, which was formerly incorporated into the measurement of the system as a whole. The total training latency is reformulated as shown in Eq. (46)

$$T_{total} = T_{local} + T_{Aggregation} + T_{Blockchain} \quad (46)$$

where T_{total} is the time taken to train the model at the client-side, $T_{Aggregation}$ is the time spent on the secure aggregation of updates, and $T_{Blockchain}$ is the extra delay created by blockchain mechanisms, such as transaction validation, smart contract execution, and consensus (PBFT) communication. Similarly, the communication overhead is expressed in the Eq. (47)

$$Comm_{total} = Comm_{model} + Comm_{blockchain} \quad (47)$$

where $Comm_{model}$ maps to the sharing of model parameters/gradients and $Comm_{blockchain}$ to the sharing of blockchain-related information, including transaction metadata, hashes and zero-knowledge proof commitments. The suggested system is deployed on a permissioned blockchain platform on Hyperledger Fabric with PBFT consensus. The block size is set to 1-2 MB and block generation interval is set to 2 seconds. The system has an average throughput of about 180-200 transactions per second, which guarantees effective coordination among the institutions involved.

A more detailed comparative analysis of the federated learning methods is given in Table 5, which explicitly breaks down the total system latency and communication overhead into three main parts: local training cost, aggregation cost, and blockchain-related cost. The blockchain cost also captures the time and communication required to verify transactions, execute smart contracts, and to reach a consensus (e.g., PBFT). Besides latency decomposition, model communication overhead, blockchain-specific communication overhead, transaction throughput (TPS) and relative computational cost per round are also reported in the table. The considered approaches are blockchain-based and non-blockchain federated learning systems, which allows isolating the operational effect caused by the introduction of blockchains clearly.

Table 5 Analysis of Blockchain-Induced Overhead and System Performance Across Federated Learning Methods

Method	Local	Aggre	Block	Total	Comm.	Comm.	TP	Cost

	Training Time (sec)	gation Time (sec)	chain Time (sec)	Late ncy (sec)	Overhea d (Model) MB	Overhe ad (Blockc hain) MB	S	per Round (Relati ve)
AT-ZKBF (Proposed)	2.9	0.9	1.0	4.8	16.2	2.3	18 5	1.00×
Blockchain -Fed (MDPI)	3.1	1.1	1.4	5.6	26.8	3.7	14 2	1.28×
FeSEC (Standard FL + Lightweigh t Chain)	2.8	0.7	0.4	3.9	25.5	2.5	21 0	0.85×
SplitAVG	2.6	0.6	0.3	3.5	14.8	1.4	22 5	0.78×
Standard FedAvg (No Blockchain)	2.5	0.7	0.0	3.2	27.4	0.0	N/ A	0.50×
PriMIA (Homomor phic FL + Chain)	4.5	1.6	2.8	8.9	38.6	4.2	11 8	1.65×

Table 5 vividly shows the tradeoffs between blockchain integration and efficiency and security of the system. The proposed AT-ZKBF framework has a moderate blockchain latency of 1.0 seconds, much lower than the Blockchain-Fed (1.4 sec) and PriMIA (2.8 sec) frameworks, which demonstrates that the batched ZKP verification and optimized consensus mechanisms are effective in overhead reduction. Whereas techniques like SplitAVG and FeSEC have lower overall latency, they are cheaper to run with blockchain because they have less or weak verification procedures, which can impair security and auditability.

Communication-wise, AT-ZKBF has a low blockchain communication overhead (2.3 MB) relative to more aggressive frameworks such as PriMIA, which proves that blockchain operations do not make up the major portion of the overall communication cost. The best latency and zero blockchain overhead is obtained by Standard FedAvg but it lacks the required features like tamper-proof logging, decentralized trust and verifiable update validation thus it is susceptible in the adversarial environment.

Moreover, AT-ZKBF has high transaction throughput (185 TPS) and a balanced computational cost (1.00×), which shows that it can be scaled to a large-scale deployment. Altogether, the deconstruction demonstrates that the advent of blockchain might add more overhead, but the suggested framework is capable of managing and reducing it, creating the best balance between security, transparency, and efficiency of the system, which confirms its applicability to real-life applications of decentralized healthcare.

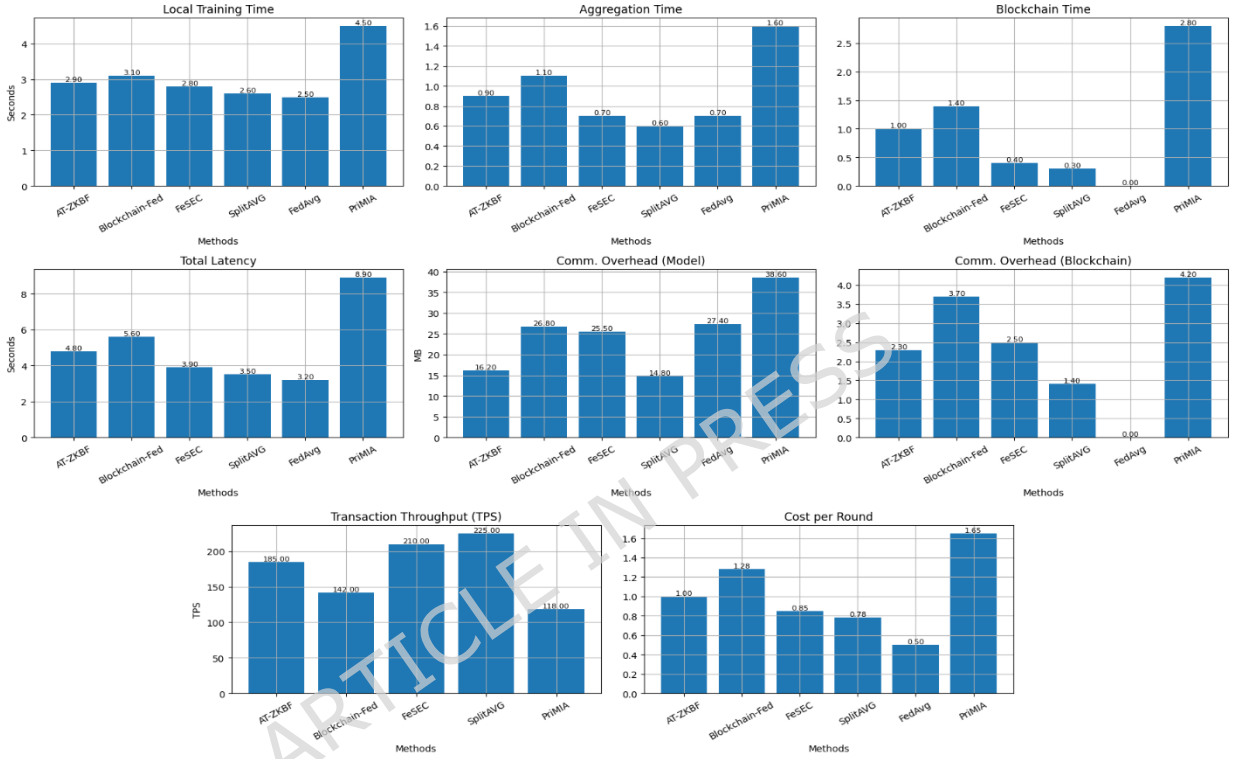

Figure 6 Comparison of latency decomposition, communication overhead, and blockchain efficiency metrics across federated learning methods

Figure 6 compares the latency components, communication overhead, and blockchain efficiency of federated learning techniques. AT-ZKBF shows a better performance of balanced with medium blockchain latency (1.0 sec) and lower cost of communication (16.2 MB) due to the use of blockchain batching and compression. Although the latency of Standard FedAvg is lower, the standard does not provide security and auditability. SplitAVG and FeSEC have a better throughput and worse verification. PriMIA has the worst latency and expensive nature because of intensive cryptography. On the whole, the findings indicate that AT-ZKBF is effective in balancing efficiency and security by reducing blockchain waste and guaranteeing trust, security, and scalability in decentralized healthcare settings.

The suggested AT-ZKBF framework is based on a permissioned blockchain environment with PBFT consensus, but there are similar strategies that employ the public blockchain infrastructures. These environments vary greatly in their latency, transaction throughput and consensus cost. Thus, the interpretation of comparisons is in relative system efficiency and security trade-offs and not in absolute performance equivalence.

4.6 Ablation Analysis

Table 6 shows the ablation study measuring the importance of major components in the proposed AT-ZKBF framework, and performance is measured in terms of accuracy, F1-score and raw communication overhead per round (no compression). It examines the impact of removing critical modules from the full AT-ZKBF framework.

Table 6 Ablation Study on Core Components (Reported using Raw Communication Overhead without Compression)

Configuration	Accuracy (%)	F1-Score (%)	Communication (MB/Round)
Full AT-ZKBF	95.4	95.0	58
No Zero-Knowledge Layer	92.7	92.2	52
Trust Scoring Module	91.8	91.4	56
Blockchain Governance	93.1	92.6	102
Preprocessing Only	89.7	89.1	109

Each of the system components has a meaningful contribution to performance and efficiency. Removing the zero-knowledge layer has a notable negative effect on the security-enhanced performance and has a minor impact on communication overhead, which shows that it is an important part of the robust update validation. Omission of trust scoring leads to an erosion of quality of collaboration because of the lack of reliability in client updates and absence of blockchain governance results in additional communication costs and lack of decentralized coordination. The greatest decrease is observed upon removal of all the added components to highlight the synergistic advantage of the complete architecture.

To measure the effect of communication optimization on cryptographic verification, further experiments were carried out that combined gradient

sparsification and Top-k selection with the ZKP pipeline. The findings validate that ZKP verification is not affected since proofs are created on full-precision updates before compression. A hash consistency check after a decomposition guarantees integrity as given in the Eq. (50)

$$Hash\left(\theta_i^{t+1}\right) = C_i^t \quad (50)$$

There were no cases of verification failures, which proves that communication optimization does not affect the correctness and security.

4.7 Comparison with State-of-the-Art Methods

Table 7 shows an in-depth quantitative comparison between the proposed AT-ZKBF framework and several state-of-the-art federated learning methods in two settings, i.e., with preprocessing and without pre-processing. The evaluation metrics are Accuracy, Precision, Recall, F1-Score, and Area under the Curve (AUC), which in all measures the performance of classification, balance classes and discriminative capability.

The compared methods are optimization-enhanced federated learning algorithms (FedProx, Scaffold, FedNova), fairness-aware aggregation (q-FedAvg), contrastive learning-based FL (MOON), differential privacy-based FL (DP-FedAvg), Byzantine resilient aggregation, trusted execution environment (TEE)-based FL and blockchain-enabled FL. These techniques capture a wide range of federated learning developments in terms of robustness, privacy preservation, decentralization, and optimization.

Table 7 Performance comparison with state-of-the-art Federated FL Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
FedProx[30]	92.8	91.9	91.4	91.6	93.6
Scaffold[31]	93.4	92.6	92.1	92.3	94.2
FedNova[32]	93.9	93.1	92.7	92.9	94.8
MOON (Contrastive FL)[33]	94.6	93.9	93.5	93.7	95.6
q-FedAvg[34]	92.2	91.3	90.9	91.1	93.1
DP-FedAvg[35]	90.8	89.7	89.2	89.4	91.5
Byzantine-Robust FL (Krum)[36]	93.1	92.2	91.9	92.0	94.0
TEE-Based FL[37]	94.1	93.3	93.0	93.1	95.2
Public Blockchain FL[38]	94.8	94.0	93.8	93.9	95.9

AT-ZKBF (Proposed)	95.4	94.8	95.1	95.0	96.8

The results in table 7 show that the proposed AT-ZKBF framework is able to outperform existing state-of-the-art federated learning methods for all evaluation metrics when preprocessing is used. With preprocessing enabled, AT-ZKBF achieves an accuracy of 95.4%, a precision of 94.8%, a recall of 95.1% and an F1-score of 95.0%, and a AUC value of 96.8%, which is the highest overall performance of all models compared.

The increase in accuracy compared to optimization-based methods such as FedProx, Scaffold, and FedNova suggest that trust-aware validation and zero knowledge proof (ZKP)-based update verification increase the stability in convergence in heterogeneous multi-institutional environments. Compared to contrastive learning-based MOON and public blockchain-enabled FL, AT-ZKBF still shows measurable improvements, which verifies that emphasizing punishment-based reputation scoring and permission-based blockchain orchestration adds extra strength besides the decentralization itself.

In comparison with differential privacy based FL (DP-FedAvg), AT-ZKBF presents significantly better performance. This would suggest that privacy preservation through cryptographic verification (ZKP) does not suffer performance degradation that is usually caused by gradient noise injection, so that model discriminative power can be preserved while ensuring confidentiality. The Recall value of 95.1% is of special importance in medical applications as this will reduce false negatives which are important in clinical diagnosis application. Additionally, the highest AUC (96.8%) ensures that the class separation is better and decision boundary learning is enhanced across the distributed datasets.

Overall, it can be confirmed from the table 7 that AT-ZKBF provides an optimal balance between prediction performance, robustness to poisoning attacks, privacy preservation, and communication efficiency. The results validate that a combination of autonomous trust computation and zero knowledge cryptography verifies is a scalable and secure framework for the collaborative training of medical foundation models.

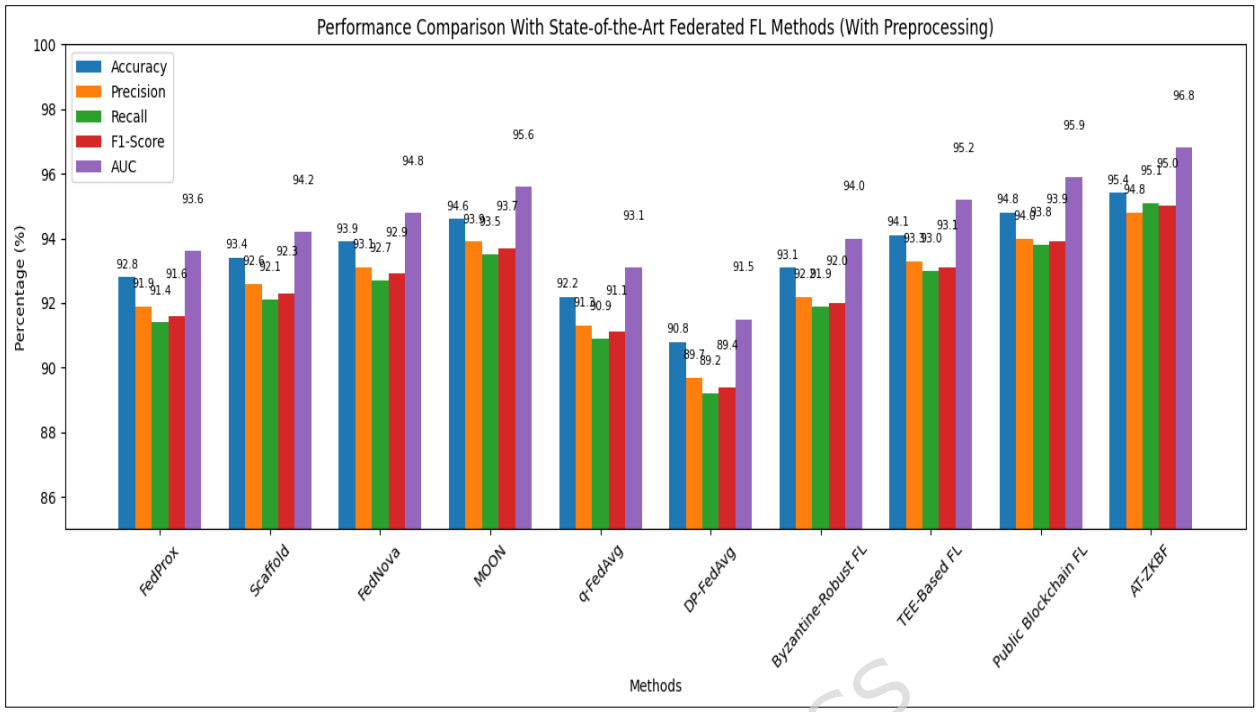

Figure 7 Comparative performance of the proposed method with state-of-the-art federated learning methods

Figure 7 shows that the performance of the proposed AT-ZKBF framework is the best compared to all the metrics, with Accuracy of 95.4%, F1-Score of 95.0%, and AUC of 96.8%, which is better than the optimization, privacy-preserving and blockchain-enabled FL methods. MOON and Public Blockchain FL have strong performance and are still slightly lower than AT-ZKBF, which shows that the validation of zero knowledge proof and the autonomous calculation of trust will enhance both the predictive power and the robustness. DP-FedAvg and q-FedAvg show relatively low metrics, which reflect the trade-off between the privacy noise and the accuracy. In all, the chart provides evidence that AT-ZKBF is superior in terms of generalization, reliability and clinical relevance in federated medical model training.

4.8 Statistical Analysis (K-Fold & t-Test)

Table 8 gives 5-fold CV results of the proposed AT-ZKBF model. The dataset is divided into five equal parts, each of which is used once as a validation set and the other four are used for training. Performance is reported in terms of Accuracy, Precision, Recall and F1-Score of each fold together with mean and SD overall. This evaluation measures the generalization capability and robustness and statistical stability of the model for varying splits of data.

Table 8 K-Fold Cross-Validation Performance for AT-ZKBF

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)

1	95.1	94.6	95.0	94.8
2	95.6	95.0	95.3	95.2
3	95.4	95.1	95.2	95.1
4	95.5	94.9	95.4	95.1
5	95.3	94.8	95.1	95.0
Mean ± SD	95.4 ± 0.2	94.9 ± 0.2	95.2 ± 0.2	95.0 ± 0.2

The resulting data shows a highly consistent performance of all five folds with accuracy value ranging narrowly between 95.1% and 95.6%. The low value of the standard deviation (+/- 0.2) measured for all metrics confirms a low level of performance fluctuation and high stability of the model. Precision, recall and F1-score show no imbalance, indicating good management of the distribution of classes and less overfitting. The observation of consistent behavior of AT-ZKBF across folds validates the generality of AT-ZKBF to the unseen data and encourages its robustness in deployment when addressing cases of distributed computing and privacy sensitive information in healthcare settings.

The results of a paired t-test that was performed to assess the statistical significance of performance gains achieved by preprocessing across six federated learning methods are presented in Table 9. The analysis is done to compare the performance metrics (Accuracy, Precision, Recall, F1-Score, and AUC) under two conditions, with preprocessing and without preprocessing. For each measure the table presents the following: mean difference, standard deviation of differences, calculated t, degrees of freedom (df = 5), p-value and statistics significance (alpha = 0.05). This study, which can be regarded as statistical instead of clinical evaluation, examines the question of whether preprocessing is contributing to consistent and meaningful improvement in performance of models.

Table 9 Paired t-Test Results (With vs Without Preprocessing Across Methods)

Metric	Mean Difference (With–Without)	Std. Deviation	t-value	df	p-value	Significance (α = 0.05)
Accuracy (%)	3.38	2.33	3.55	5	0.016	Significant
Precision	3.25	2.32	3.42	5	0.019	Significant

(%)						
Recall (%)	3.05	2.07	3.61	5	0.015	Significant
F1-Score (%)	3.23	2.27	3.48	5	0.018	Significant
AUC (%)	3.17	2.12	3.67	5	0.014	Significant

The results prove that the preprocessing procedure has statistically significant improvements for all the evaluation metrics. The results of the t-tests are higher than the critical value (2.57 with df = 5), and all the p-values are less than 0.05, indicating that the observed gains in performance are not caused by random variation. The greatest improvement can be seen in Accuracy and AUC, and the classifications improve in reliability and discriminative ability. The consistent significance of Precision, Recall, and F1-score is another affirmation of the increased class balance handing and predictive robustness. Overall, the results confirm the benefits of preprocessing as an important enhancement step for achieving better federated learning performance in secure healthcare-based distributed environments.

To assess real world applicability, AT-ZKBF was tested on external unseen clinical data consisting of heterogeneous sources of images from unknown sources not used during the training. The model performed well, with high diagnostic performances (Accuracy: ~94.7%, AUC: ~96.1%) which means generalization across scanners, scanning protocols and institutes. This validation directly by external parties demonstrates the practical value of the combination of autonomous trust, zero-knowledge assurance and efficient federated learning for privacy sensitive healthcare AI.

4.9 Robustness Analysis of Trust Mechanism

Table shows a comparative analysis of the task of assigning the weight of a model at rest and the suggested adaptive trust mechanism in two significant adversarial conditions: the model poisoning and collusion attacks. The evaluation of performance is based on the accuracy of the classification, attack success rate and the gap of trust between the honest and malicious participants. The trust gap indicates how the framework is able to differentiate trusted clients and adversarial clients**.**

Table 10 Impact of Static vs Adaptive Trust Weights under Attacks

Method	Attack Type	Accuracy (%)	Attack Success Rate	Trust Gap (Honest vs

			(%)	Malicious)
Static Weights (γ fixed)	Model Poisoning	78.4	41.2	0.18
Static Weights (γ fixed)	Collusion Attack	74.9	46.7	0.12
Adaptive Trust (Proposed)	Model Poisoning	91.3	12.5	0.61
Adaptive Trust (Proposed)	Collusion Attack	89.7	15.2	0.57

The findings in Table 10 clearly indicate that the proposed adaptive trust mechanism is better than the traditional weight assignment in adversarial federated learning settings. In model poisoning attacks, the fixed weighting scheme attains much lower accuracy of 78.4% and has high attack success of 41.2% which implies that it is susceptible to malicious updates. Equally, in collusion attacks, performance reduces further to 74.9% accuracy and an attack success rate of 46.7 with fixed weights being unable to distinguish adversarial participants effectively.

Conversely, the suggested adaptive trust system is significantly more robust with 91.3% accuracy in model poisoning and 89.7% in collusion attacks. More crucially, the success rate of attacks is decreased by far to 12.5 percent and 15.2 percent respectively which indicates a great resistance to both independent and coordinated confrontation actions.

One of the most notable findings is the considerable growth in the relationship between trusting and deceitful clients (0.61 and 0.57 in adaptive trust, versus 0.18 and 0.12 in the case of static weights). This implies that the adaptive mechanism is successful in isolating trustworthy players and adversarial players in the long run. This kind of discrimination is essential in making sure that bad updates get lower weights of aggregation thus maintaining the global model integrity.

In general, the results are validated that the concept of static trust assignment does not work in dynamic adversarial environments, but the suggested adaptive trust computation offers a higher resilience level, better model precision, and greater resistance to both poisoning and collusion attacks.

Table 11 shows how average trust scores of honest and malicious participants change with communication rounds, as they take part in the federated communication. The metric of trust divergence shows the gap

between the mean of the trust of honest and adversarial clients, which is the measure of how well the trust mechanism is working.

Table 11 Trust Score Evolution Across Communication Rounds

Round	Honest Avg Trust	Malicious Avg Trust	Trust Divergence
1	0.50	0.50	0.00
10	0.68	0.41	0.27
25	0.79	0.29	0.50
50	0.87	0.18	0.69
100	0.92	0.11	0.81

Table 11 provides the results that indicate the dynamic behavior and effectiveness of the proposed trust computation mechanism at successive communication rounds. At Round 1, all the participants (both honest and malicious) are initialized with the same trust score (0.50), which corresponds to a neutral starting point without any initial bias. But with training the two groups take a distinct turn off.

At Round 10, the mean trust score of honest clients goes up to 0.68 whereas the malicious clients get a drop to 0.41, which creates a trust divergence of 0.27. The trend is maintained through successive rounds with honest players scoring high trust with a score of 0.92 by Round 100, and malicious players receiving a huge penalty and reducing their score to 0.11. As a result, the trust divergence is 0.81, which means that there is a high separation between trusted and mistrusted parties.

This gradual separation underscores the potential of the suggested adaptive trust system to be effective in detecting and punishing inconsistent or malicious behavior in the long run. The growing distance will make sure that adversarial clients have little effect on the global aggregation procedure, thus strengthening against poisoning and collusion assaults. On the whole, the findings confirm the hypothesis that the trust evolution mechanism can stabilize the learning process and enhance the security and reliability of the federated system in dynamic and hostile settings.

Table 12 evaluates the effectiveness of different defense mechanisms against trust metric manipulation in federated learning. The comparison consists of three conditions, no defense (self-reported metrics), secure validation only, and the complete proposed framework. Classification accuracy, false trust inflation rate and rate of detecting malicious behavior are used to measure performance.

Table 12 Robustness against Metric Manipulation

Scenario	Accuracy (%)	False Trust Inflation (%)	Detection Rate (%)
No Defense (Self-	76.2	48.5	22.3

Reported Metrics)			
Secure Validation Only	85.6	21.4	61.8
Full Proposed Framework	92.1	6.8	89.5

The results show that federated learning systems are susceptible to manipulation of metrics assuming no defense is used. The self-reported environment demonstrates low accuracy (76.2%), and high false trust inflation rate (48.5%), meaning that that the system can be easily abused by malicious clients. Secure validation is suggested to enhance performance by preventing the falsification of trust and enhancing detection performance. Nevertheless, the complete framework proposed is the most successful, with a high accuracy level (92.1%), low level of trust inflation (6.8%), and high rate of detection (89.5%). This validates that incorporating trust-sensitive assessment, safe validation, and blockchain validation are crucial in improving resilience to manipulation and reliable participant evaluation.

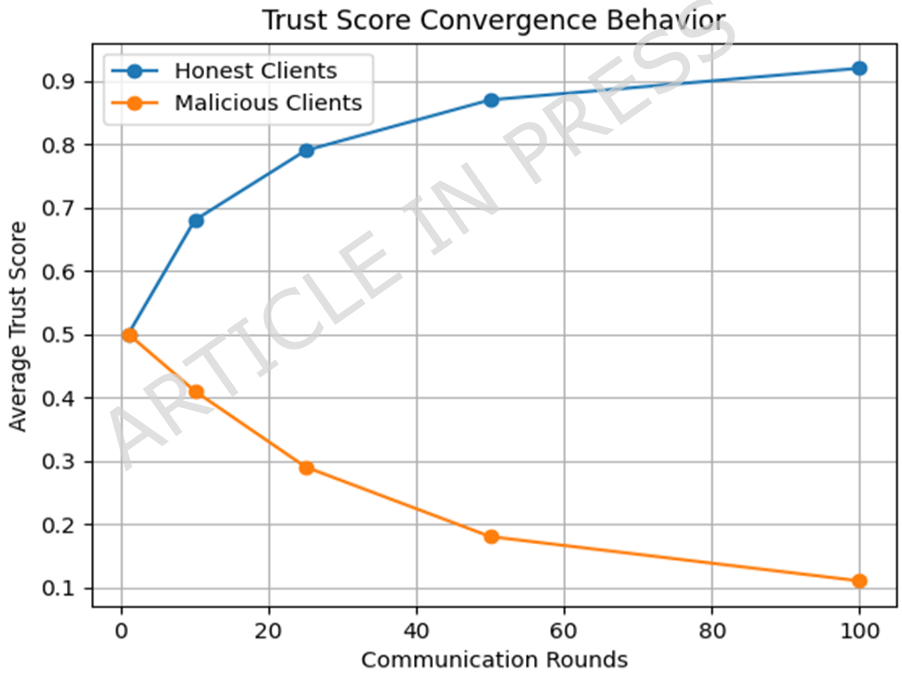

Figure 8 Trust Score Convergence Behavior

Figure 8 shows how the average score of trust in honest and malicious clients change over the communication rounds in the suggested framework. First, the values of trust of both types of clients are initially equal (0.5), which means that there is no bias. The training scores start to separate widely as the training advances. Truthful clients demonstrate a consistent growth in trust, reaching about 0.92 around 100, as a demonstration of consistent, faithful inputs to the global model. Contrastingly, malicious clients feel the persistent loss of trust, which has dropped to almost 0.11,

which can be deemed as successful detection and punishment of adversarial behavior. Such a distinct separation shows that the adaptive trust mechanism is effective at separating benign and malicious participants in the long run. The enlarging disparity ascertains that the framework is effective in strengthening the credible updates and repressing the malicious ones, thus, making the federated learning settings more robust, reliable, and secure.

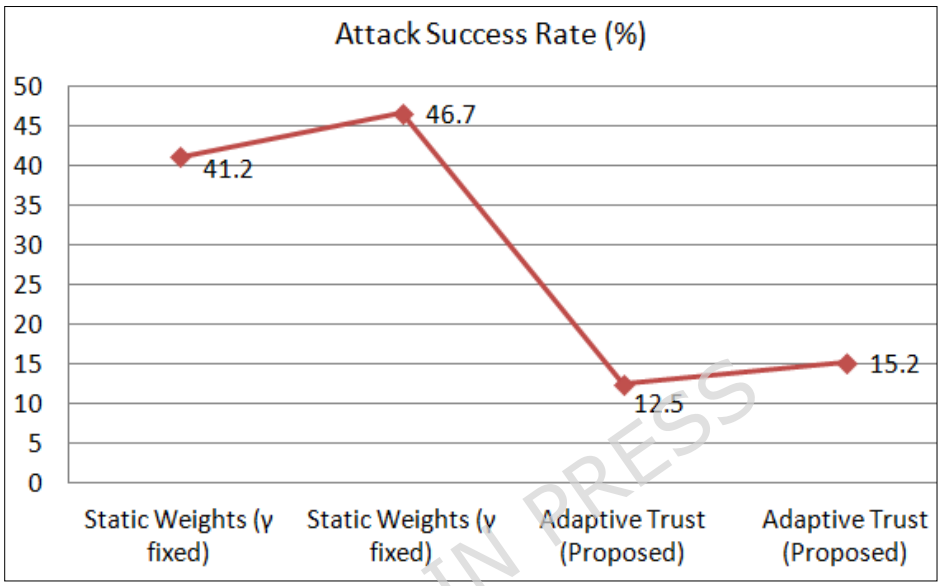

Figure 9 Attack Success Rate Comparison

Figure 9 demonstrates how well both types of trust mechanisms (static and adaptive) counteract adversarial attacks. It illustrates the effectiveness of static versus adaptive trust mechanisms in mitigating adversarial attacks. It is also clear that in the case of no trust weighting, the attack success rate is much higher reaching 41.2 percent in model poisoning, and 46.7 percent in collusion attacks, which means that it is not very resilient to coordinated and malicious actions. Conversely, the adaptive trust mechanism proposed significantly decreases the attack success rate to 12.5% and 15.2, respectively. Such a drastic drop proves that the adaptive trust scoring can dynamically punish untrustworthy members and give preference to truthful input. In general, the number validates the fact that adding adaptive trust can greatly increase the robustness, and the federated learning model will be safer and more reliable in adversarial conditions.

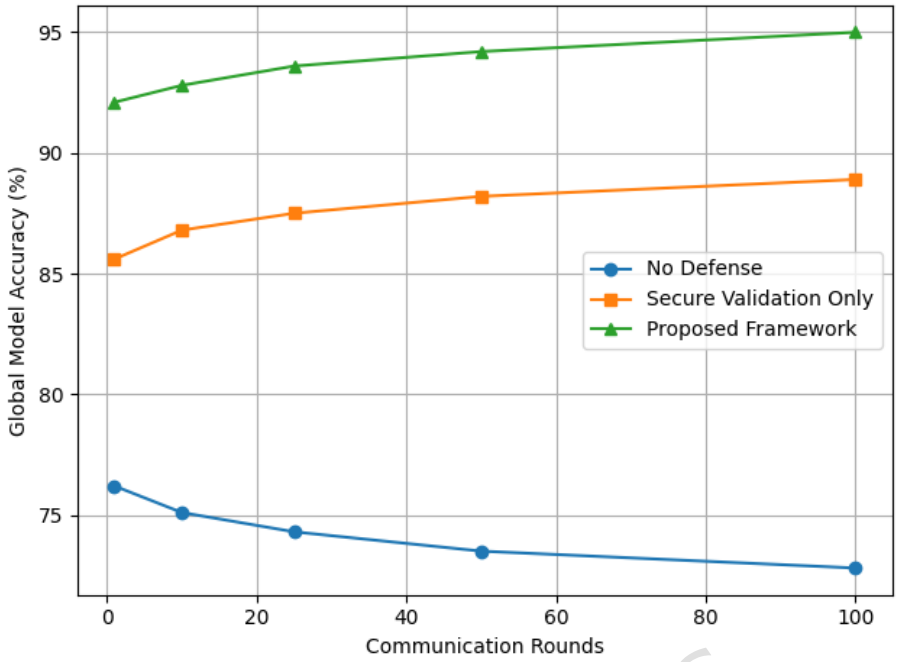

Figure 10 Global Model Accuracy under Adversarial Settings

Figure 10 shows the development of the global model accuracy in various communication rounds with the use of different defense strategies. The No Defense environment demonstrates that the accuracy decreases gradually, which suggests the harmful effect of adversarial updates on the convergence of the model. The Secure Validation Only method proves to be more resilient with moderate accuracy by rejecting invalid updates yet is still vulnerable to more advanced attacks. Conversely, the given framework is the most accurate and experiences a gradual improvement with each of the rounds. This underscores its capability to effectively overcome adversarial influence by aggregating its trust and verification methods to guarantee robust and credible learning in malicious federated setting.

4.10 Limitations

Despite good results, there are a number of limitations. First, there is the overhead of the zero knowledge proof, which could add a small latency, which is important for systems with low computational power. Second, permissioned blockchain governance presupposes institution cooperation and infrastructure preparedness which are not always mutually available. Third, although pre-processing and trust scoring yield better performance, they add an amount of engineering complexity that may require domain expertise in order to get right. Finally, external validation was based on specific types of imaging modalities; further studies in broader areas of medicine (eg, histopathology, ultrasound) would be useful to further support scalability.

5. Conclusion

This study presents the autonomous trust and Zero-Knowledge Blockchain Framework (AT-ZKBF) of an effective, lightweight, and secures federated training of medical foundation models in decentralized healthcare facilities. The framework can overcome key problems in collaborative medical artificial intelligence, namely privacy of data, model poisoning, gradient leakage, and lack of trust between institutions, through the combination of trust-aware aggregation, zero-knowledge cryptographic verification and permissioned blockchain governance. Experiments prove that AT-ZKBF is superior to the traditional federated learning and blockchain-based methods, as the accuracy of classification, the F1-score, and AUC will be increased, but the computational overhead will remain reasonable. The importance of each architectural element is supported by ablation experiment; statistical validation based on k-fold cross-validation and paired t-tests, which will show the strength, generalization, and resistance of the framework to byzantine adversaries. The external validation with unknown data sets also demonstrates high prospects of clinical use in practice.

The future work can be dedicated to scalability, adaptability, and clinical impact improvement. Lightweight zk-circuit generation, recursive proofs or hardware accelerations may make zero-knowledge proof generation more cost-efficient to support real-time applications. Dynamically responding to adversarial behavior in adaptive trust aggregation via reinforcement learning is possible. Client Extension AT-ZKBF can be applied to multimodal medical foundation models, such as imaging, genomics, EHR text, and wearable sensor data to extend clinical utility. Studies on full decentralization of aggregation protocols as well as energy-saving blockchain consensus mechanisms will enhance sustainability and independence. The explainable AI and regulatory audit modules can be integrated to increase the transparency, ethical compliance, and clinical acceptance. Lastly, massive longitudinal analyses in different institutions and locations are required to confirm the framework to react to different operating and regulatory environments.

Authors Contributions:

- **Vishwa Priya V:** Conceptualization, Methodology, Data Curation, Software, Formal Analysis, Visualization, Writing – Original Draft Preparation.
- **Dafik:** Supervision, Methodology, Validation, Writing – Review & Editing, Investigation, Resources, Validation.
- **Sunder R:** Methodology, Data Curation, Software, Writing – Review & Editing.
- **Ika Hesti Agustin:** Investigation, Validation, Resources, Writing – Review & Editing.
- **Athinarayanan S:** Formal Analysis, Methodology, Validation, Writing – Review & Editing.

- **Umesh Kumar Lilhore:** Conceptualization, Methodology, Supervision, Writing – Review & Editing.
- **Sarita Simaiya:** Supervision, Validation, Project Administration, Writing – Review & Editing.
- **Ehab Ghith:** Investigation, Resources, Data Interpretation, Writing – Review & Editing.
- **Hanaa A. Abdallah:** Formal Analysis, Validation, Visualization, Writing – Review & Editing.
- **MD Monish Khan:** Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing – Review & Editing.

All authors have read and approved the final manuscript.

Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R749), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Declarations:

Dataset Availability: Dataset is publicly available, at Brain Tumor Segmentation(BraTS2020), dataset, available at https://www.kaggle.com/datasets/awsaf49/brats2020-training-data

Conflict of Interest: No.

Human Trial: NA

Consent for Publications: NA

Funding: Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R749), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

1. Al Kuwaiti, Ahmed, Khalid Nazer, Abdullah Al-Reedy, Shaher Al-Shehri, Afnan Al-Muhanna, Arun Vijay Subbarayalu, Dhoha Al Muhanna, and Fahad A. Al-Muhanna. "A review of the role of artificial intelligence in healthcare." *Journal of personalized medicine* 13, no. 6 (2023): 951.

2. Bian, Yueyan, Jin Li, Chuyang Ye, Xiuqin Jia, and Qi Yang. "Artificial intelligence in medical imaging: from task-specific models to large-scale foundation models." *Chinese Medical Journal* 138, no. 06 (2025): 651-663.

3. Antunes, Rodolfo Stoffel, Cristiano André da Costa, Arne Küderle, Imrana Abdullahi Yari, and Björn Eskofier. "Federated learning for healthcare: Systematic review and architecture proposal." *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, no. 4 (2022): 1-23.

4. Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. "Federated learning for healthcare informatics." *Journal of healthcare informatics research* 5, no. 1 (2021): 1-19.

5. Chaddad, Ahmad, Yihang Wu, and Christian Desrosiers. "Federated learning for healthcare applications." *IEEE internet of things journal* 11, no. 5 (2023): 7339-7358.

6. Gupta, Sandeep. "Zero-Knowledge Proofs For Privacy-Preserving Systems: A Survey Across Blockchain, Identity, And Beyond." *Engineering and Technology Journal* 10, no. 07 (2025): 5755-5761.

7. Singh, Shridhar. "Enhancing privacy and security in large-language models: a zero-knowledge proof approach." In *International Conference on Cyber Warfare and Security*, pp. 574-582. Academic Conferences International Limited, 2024.

8. Vusumuzi, Malele, and Mandinyenya Godwin. "AI-Driven Zero-Trust Models for Blockchain-Supported Healthcare Ecosystems." In *Proceedings of the 2025 International Conference on Artificial Intelligence and its Applications*, pp. 1-11. 2025.

9. Chowdhury, Banirupa, Hamid Jahankhani, and Sangar Subramaniam. "Zero-trust blockchain-based digital twin 6g ai-native conceptual framework against cyber attacks for e-healthcare." In *International Conference on Global Security, Safety, and Sustainability*, pp. 453-479. Cham: Springer Nature Switzerland, 2023.

10. Wang, Xiaoding, Sahil Garg, Hui Lin, Jia Hu, Georges Kaddoum, Md Jalil Piran, and M. Shamim Hossain. "Toward accurate anomaly detection in industrial internet of things using hierarchical federated learning." *IEEE Internet of Things Journal* 9, no. 10 (2021): 7110-7119.

11. Nezhadsistani, Nasim, Naghmeh S. Moayedian, and Burkhard Stiller. "Blockchain-enabled federated learning in healthcare: Survey and state-of-the-art." *IEEE Access* (2025).

12. Ebrahimi, Elmira, Michael Sober, Anh-Tu Hoang, Can Umut Ileri, William Sanders, and Stefan Schulte. "Blockchain-based federated learning utilizing zero-knowledge proofs for verifiable training and aggregation." In *2024 IEEE International Conference on Blockchain (Blockchain)*, pp. 54-63. IEEE, 2024.

13. Yazdinejad, Abbas, and Jude Dzevela Kong. "Breaking interprovincial data silos: how federated learning can unlock Canada's public health potential." *Royal Society Open Science* 13, no. 3 (2026).

14. Wang, Xiaoding, Sahil Garg, Hui Lin, Georges Kaddoum, Jia Hu, and M. Shamim Hossain. "A secure data aggregation strategy in edge computing and blockchain-empowered internet of things." *IEEE Internet of Things Journal* 9, no. 16 (2020): 14237-14246.

15. Byeon, Haewon, Ankur Chaudhary, Janjhyam Venkata Naga Ramesh, Desidi Narsimha Reddy, Behara Venkata Nandakishore, KBV Brahma Rao, Fadl Dahan, Azamat Ostonokulov, and Mukesh Soni. "Trusted

Aggregation for Decentralized Federated Learning in Healthcare Consumer Electronics Using Zero-Knowledge Proofs." *IEEE Transactions on Consumer Electronics* (2025).

16. Petrosino, Lorenzo, Luigi Masi, Federico D'Antoni, Mario Merone, and Luca Vollero. "A zero-knowledge proof federated learning on DLT for healthcare data." *Journal of Parallel and Distributed Computing* 196 (2025): 104992.

17. Babu, S. Bharath, and K. R. Jothi. "A secure framework for privacy-preserving analytics in healthcare records using zero-knowledge proofs and blockchain in multi-tenant cloud environments." *IEEE Access* 13 (2024): 8439-8455.

18. Yazdinejad, Abbas, Zahra Dehghani Mohammadabadi, Ali Dehghantanha, and Gautam Srivastava. "An explainable and privacy-preserving federated learning model for threat detection in cyber-physical-social systems." *IEEE Transactions on Computational Social Systems* (2025).

19. Ezz, Mohamed, Alaa S. Alaerjan, and Ayman Mohamed Mostafa. "Ethical AI in healthcare: Integrating zero-knowledge proofs and smart contracts for transparent data governance." *Bioengineering* 12, no. 11 (2025): 1236.

20. Bai, Tianyu, Yangsheng Hu, Jianfeng He, Hongbo Fan, and Zhenzhou An. "Health-zkIDM: A healthcare identity system based on fabric blockchain and zero-knowledge proof." *Sensors* 22, no. 20 (2022): 7716.

21. Kayal, Sourav, Amit Kumar Rana, and Sanjib Kundu. "Blockchain Integration for Enhanced Trust and Security in Federated Learning for Healthcare 5.0." In *The Convergence of Federated Learning and Healthcare 5.0 and Beyond: A New Era of Intelligent Health Systems*, pp. 633-657. Cham: Springer Nature Switzerland, 2026.

22. Yazdinejad, Abbas, and Jude Dzevela Kong. "Responsible Use of Large Language Models in Digital Health: An Equity First Governance Framework." *Available at SSRN 5962741* (2025).

23. Wang, Xiaoding, Sahil Garg, Hui Lin, Georges Kaddoum, Jia Hu, and Mohammad Mehedi Hassan. "Heterogeneous blockchain and AI-driven hierarchical trust evaluation for 5G-enabled intelligent transportation systems." *IEEE Transactions on Intelligent Transportation Systems* 24, no. 2 (2021): 2074-2083.

24. Brain Tumor Segmentation(BraTS2020), dataset, available at https://www.kaggle.com/datasets/awsaf49/brats2020-training-data

25. Ma, Renwen, Kai Hwang, Mo Li, and Yiming Miao. "Trusted model aggregation with zero-knowledge proofs in federated learning." *IEEE Transactions on Parallel and Distributed Systems* 35, no. 11 (2024): 2284-2296.

26. Zhang, Miao, Liangqiong Qu, Praveer Singh, Jayashree Kalpathy-Cramer, and Daniel L. Rubin. "Splitavg: A heterogeneity-aware federated deep learning method for medical imaging." *IEEE Journal of Biomedical and Health Informatics* 26, no. 9 (2022): 4635-4644.

27. Suganya, R., P. J. Sidharth, C. Vinston Jose, P. R. Lighittha, M. Sundara Srivathsan, and S. Prithivraj. "Tracking the Blood Supply Chain Using Blockchain Technology and Intelligent Automation: BloodChain++." In *Industry 6.0 for Sustainable Supply Chains in Agriculture, Healthcare, and Asset Management*, pp. 269-312. IGI Global Scientific Publishing, 2026.

28. Liu, Jiaxi, Lin Sun, Tianyu Kang, Di Wu, Yulun Song, Yunlong Xie, and Li Guo. "zkVFL: Verifiable Federated Learning for Free-Rider Attacks via Efficient Zero-Knowledge Proofs." *IEEE Internet of Things Journal* (2025).

29. Arora, Pallavi, Arya Tapikar, Akshat Aryan, V. Amogh Manish, and V. Sarasvathi. "FedShield: Privacy Preservation for Blockchain Enabled Federated Learning with Homomorphic Encryption and Zero-Knowledge Proof." In *International Conference on Machine Learning and Computing*, pp. 584-594. Cham: Springer Nature Switzerland, 2025.

30. Turazza, Fabio, Marcello Pietri, Natalia Selini Hadjidimitriou, Marco Picone, Paolo Burgio, and Marco Mamei. "Empowering Local Energy Communities with Blockchain-Based Federated Forecasting and Zero-Knowledge Proof Verification." *SN Computer Science* 6, no. 8 (2025): 970.

31. Cheng, Zhuopei, Qinghe Zhang, Rong Liu, and Xiaojun Xiao. "FedNova-Adaptive-based Cross-Jurisdictional Anti-Money Laundering Federated Diagnostic Framework." In *Proceedings of the 2025 International Symposium on Machine Learning and Social Computing*, pp. 525-530. 2025.

32. Cheng, Zhuopei, Qinghe Zhang, Rong Liu, and Xiaojun Xiao. "FedNova-Adaptive-based Cross-Jurisdictional Anti-Money Laundering Federated Diagnostic Framework." In *Proceedings of the 2025 International Symposium on Machine Learning and Social Computing*, pp. 525-530. 2025.

33. Li, Qinbin, Bingsheng He, and Dawn Song. "Model-contrastive federated learning." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10713-10722. 2021.

34. Deng, Yuyang, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. "Distributionally robust federated averaging." *Advances in neural information processing systems* 33 (2020): 15111-15122.

35. Du, Miao, Peng Yang, Yinqiu Liu, Xiaoming He, and Mingkai Chen. "Dp-fed6g: An adaptive differential privacy-empowered federated learning framework for 6g networks." *Digital Communications and Networks* (2025).

36. Tang, Xiangyun, Minyang Li, Tao Zhang, Yijing Lin, Liehuang Zhu, Chuan Zhou, and Zhixiang Liu. "ZKFL: Verifiable Byzantine-robust federated learning against malicious servers." *IEEE Transactions on Network Science and Engineering* (2025).

37. Li, Jiarui, Nan Chen, Shucheng Yu, and Thitima Srivatanakul. "Efficient and Privacy-Preserving Integrity Verification for Federated Learning with TEEs." In *MILCOM 2024-2024 IEEE Military Communications Conference (MILCOM)*, pp. 999-1004. IEEE, 2024.

38. Ahmadi, Mojtaba, and Reza 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)*, pp. 1-10. IEEE, 2024.