Journal of Theoretical and Applied Information Technology
15th February 2026. Vol.104. No.3
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PPBEFL: PRIVACY-PRESERVING BLOCKCHAIN ENABLED FEDERATED LEARNING FOR HEALTHCARE DATA SECURITY

TANISHA BHARDWAJ[1], SUMANGALI K[2]

[1]Research Scholar, School of Computer Science Engineering and Information Systems,
Vellore Institute of Technology, Vellore 632014, Tamilnadu, India
[2]Associate Professor, School of Computer Science Engineering and Information Systems,
Vellore Institute of Technology, Vellore 632014, Tamilnadu, India
E-mail: [1] tanisha.2023@vitstudent.ac.in, [2] ksumangali@vit.ac.in

ABSTRACT

Blockchain (BC)-based platforms have emerged as a result of the technology's explosive expansion., each with distinct structures and consensus mechanisms. This has heightened the focus on blockchain interoperability, facilitating interactions between different platforms. In decentralized learning environments, maintaining security, transparency, and trust is particularly challenging, particularly in delicate sectors like financial services and healthcare, where data centralization is not a viable option. The Blockchain-Enabled Federated Learning (BEFL) approach, which combines blockchain technology with this work, federated learning (FL) is provided as a dependable and effective alternative. By using an aggregation technique to eliminate anomalous model parameters and integrating blockchain procedures and privacy strategies to synchronize client privacy protection, PPBEFL protects against poisoning attempts. The PPBEFL model trains the client's datasets using local models of the Visual Geometry Group 19 (VGG19) for images and a Convolutional Neural Network (CNN) for the dataset. The updates are serialized to avoid concurrency issues and recorded immutably on the blockchain. Aggregated updates refine a global model iteratively in a transparent and verifiable way. Heart Disease (HD) and Breast Cancer (BC) Image Datasets (Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM)) are used for experiments, Brain Tumor Magnetic Resonance Imaging (MRI) Dataset, and Breast Cancer Wisconsin (Diagnostic) comparison with typical FL schemes like FedAvg, Federated Learning with Multi-Party Computation (FL-MPC), Federated Learning with Robust Aggregation in Edge Computing (FL-RAEC), and the privacy-preserving and efficient FL (PEFL) all show improved defense against different types of attacks based on metrics like throughput, precision, loss, and delay.

Keywords: *Blockchain, Consensus Mechanism, Differential Privacy (DP), Federated Learning (FL), privacy security, Neural Network, Data protection, Decentralized learning, Privacy-Preserving Blockchain Enabled Federated Learning (PPBEFL).*

1. INTRODUCTION

In today's AI-driven world, data is of paramount importance. However, while AI has the potential to improve human lives, it also increases the risk of cyber threats, enabling attackers to compromise user privacy. Thus, it is essential to protect the data's privacy. FL enables multiple client entities to participate in a model-based training process, and this FL is utilized to address those problems. The raw data is extracted by local devices and used by every local Machine Learning (ML) model for independent training. Without asking users to share their original, private information, a model might facilitate collaborative learning [1]. When comparing conventional ML models with centralized data processing and this method, a global model was created and then optimized by FL through the sharing of parameter updates among several devices. Each client is given access to their local datasets in FL for training the chosen ML model. There is not much data in these local datasets. Using the locally learned model parameters, clients connect with a central coordinating server instead than disclosing unprocessed data. This approach protects the privacy of data, enabling local models to collect important data features associated with the same issue across multiple datasets owned by various clients. As FL's entities function according to a collaboration agreement, it is also known as collaborative learning. Kairouz et al. [2] provided a more comprehensive definition, characterizing FL as a machine learning framework where a central service provider or server controls the number of customers who interact to solve a learning task. Figure 1 depicts FL's general organizational structure.

However, FL faces difficulties in ensuring that the training procedure is transparent and secure, as it requires aggregating updates from multiple clients without relying on a centralized authority [3]. Its enduring and decentralized nature, to solve these problems, blockchain technology offers a workable alternative [4]. Blockchain, functioning as a distributed shared ledger and database, possesses key attributes such as decentralization, immutability, consensus mechanisms, traceability, privacy protection, fault tolerance, and the ability to execute smart contracts. Smart contracts, in particular, enable parties that lack mutual trust to interact by automatically verifying and executing predefined scripts on the blockchain [5,6].

Blockchain also enabled the creation of an unalterable ledger that maintained a verifiable history of each contribution, keeping the issue of accountability transparent [8]. As a result, medical data is exchanged securely and anonymously, facilitated by blockchain technology, which enables distributed consensus in transactions between patients and research institutions and automates the management of resources. By integrating blockchain with FL, each client model update can be securely recorded, creating a verifiable history of contributions. The result of the overall process is a more robust tamper-proof aggregation process. This will foster trust among participants [7,8].

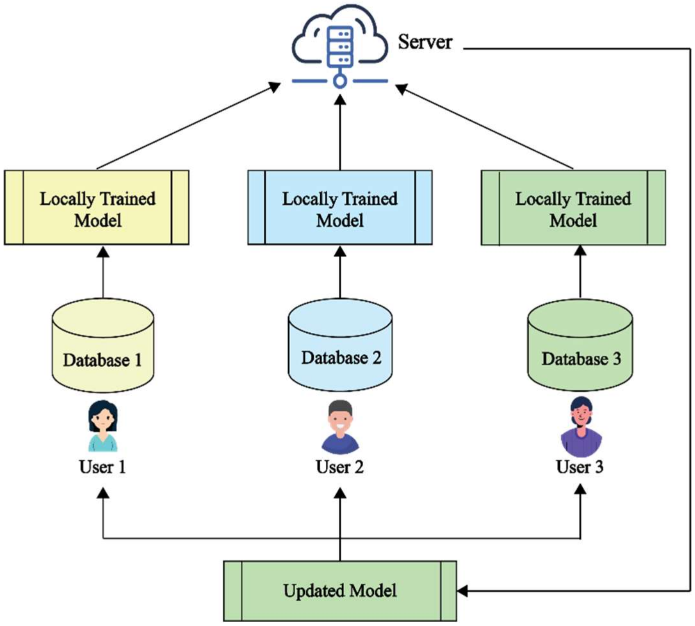


Figure 1: Basic Architecture For FL

Blockchain-enabled Federated Learning (BEFL) addresses important shortcomings in existing methods, representing a substantial advance in the field of machine learning [9]. BEFL incorporates the creativity of blockchain into FL to facilitate the decentralization of aggregation [10]. Central server dependency is eliminated with this system, and hence data integrity, along with confidentiality, remains intact across different nodes [11]. BEFL inherently possesses novelty, as it can combine secure model sharing with decentralized consensus mechanisms, ensuring strong data provenance and tamper-proof record-keeping. BEFL also scales and tolerates faults more efficiently compared to traditional models due to its characteristic of having nodes coordinate flawlessly with each other without relying on a central reference entity [12]. This paradigm shifts the balance of privacy and collaboration, making BEFL a more transformative approach to sensitive data-driven applications, such as healthcare and finance [13].

Although blockchain technology provides an effective means to mitigate trust issues in FL, concerns regarding data privacy and model security persist. In a distributed learning system, privacy preservation methods primarily focus on two key aspects: (i) maintaining the training data private and secret and (ii) protecting the local model's parameters, which are shared with other nodes or a central server and are obtained using optimization techniques like gradient descent. To achieve these objectives, Data anonymization, Homomorphic encryption (HE), secure multi-party computation (SMC), and differential privacy (DP) are popular privacy-preserving methods in machine learning. However, each approach has its limits. It needs to be encrypting and decrypting local models, with a significant increase in the overhead of the calculation. Strong consolidation algorithms generally assume that the server is honest, which is often unrealistic in practice. Additionally, the committee mechanism can also increase high computational costs. In addition, balanced efficiency with privacy protection is an important challenge for many service providers, as compromising on the performance of the system can hinder their ability to effectively manage existing services.

The goal of the study is to check how BC Technology (BCT) is included in the FL Framework. It addresses important issues related to data privacy, model integrity and transparency in BCT decentralized networks. This study introduces ppbefl, a safe and effective approach that integrates FL and BCT. Multiple customers participate in the FL process simulation, and for each client, BC keeps a local model extension. The global model is updated via federated consolidation of various local models. Benchmark datasets collected from Kagal, FL and BC can be used with the basic neural network (NN) model. PPBEL enhances the overall protection and transparency of the FL process. The accuracy of the model is significantly improved with PPBEFL. BEFL enhances the protection, transparency and effectiveness of machine learning in collaborative settings.

Motivation

Especially maintaining the integrity, authenticity and integrity of data when training collaborative model training collaborative models due to the widespread use of data sources spread in industries such as healthcare. The possibility of unauthorized access and data leaks increases as the centralized machine learning framework requires data aggregation. Problems with confidence in central aggregator, estimates attacks, and model toxicity can still affect federated learning (FL), even if it allows decentralized without revealing the raw data during training. This is not suitable for use in contexts where the results may be severe because there are no proven ways to ensure that updates are accurate and customers are responsible. Due to these restrictions, a reliable and open system is required to protect user data, identify malicious inputs and verify the validity of model modifications made by third parties.

Research Objectives

Addressing security and privacy flaws is the primary objective of this research in standard federated learning (FL) architecture, which protects confidentiality by taking advantage of blockchain to enable federated learning. To ensure that model changes are transparent and can be detected, the suggested model uses blockchain technology, which provides decentralized consensus and irreversible recording. It has a strong aggregation process that can detect and detect harmful or abnormal model parameters, preventing poisoning. The use of architecture of VGG19 for local image-based data training and CNNs are ensured by CNNs for generic data methods. The objective is to create a decentralized education environment that is resistant to tampering, guarantees the integrity of updates, and enables dispersed customers to safely cooperate.

Research hypothesis

- H_1: Blockchain-enabled federated learning does not exhibit different poisoning vulnerabilities compared to traditional federated learning.
- H_2: Privacy-preserving aggregation in blockchain-enabled federated learning does not prevent information leakage or maintain model utility.
- H_3: Integrating anomaly detection with blockchain consensus does not significantly mitigate the impact of malicious client updates.

The main contribution of study is following

- The creation of a federated learning system with blockchain capabilities that employs serial model update transactions and smart contract-based verification to ensure irreversible, tamper-resistant and conflict-free parameter aggregation.

- The implementation of a strong discrepancy filtering mechanism combines local customer-side gradient verification with global blockchain-assisted aggregation that filters to reduce the attacks of poisoning and effectively increase model integrity.
- Design of a privateness-keeping, verbal exchange-optimized machine helping heterogeneous statistics modalities by leveraging VGG19 for image records and CNN for different datasets, achieving steady and scalable deployment in useful resource-confined healthcare environments.

2. LITERATURE REVIEW

Recent research has extensively explored the integration of FL with blockchain technology and advanced privacy-preserving mechanisms to address security vulnerabilities, privacy leakage, and poisoning attacks in distributed learning environments.

Feng et al. [14] was developed to improve FL's effectiveness and security, blockchain-based asynchronous federated learning (BAFL) architecture. Blockchain technology prevents manipulation, ensuring the integrity of model data, while the asynchronous learning accelerates global aggregation. Additionally, within the BAFL structure, the ranking and contribution of the local models trained on the participating devices is evaluated using a novel entropy weight technique. By controlling local training processes, reducing delays in communication and maximizing block production rates, energy consumption and local model update efficiency are balanced. Comprehensive evaluation results suggest that the suggested BAFL structure improves other distributed machine learning approaches in both efficiency and flexibility against poison attacks. Though, this approach mainly focuses on efficiency and energy balance, posing restricted insight into robust privacy preservation in adversarial settings.

Li et al. [15] introducing an efficient privacy-preserving federated Learning (EPPFL) scheme to address the presence of incredible users. This approach is designed to reduce the adverse effects of incredible participants, which assures the use of high quality data for the model updates. The FL model maintains minimal communication and processing costs, obtaining rapid convergence through continuous application of "irrelevant components" and "weighted aggregation" processes. As a result, this method improves training effectiveness and model purity. A threshold pillier protects the user related to the user through cryptocystom-based safe structure training. In addition, the treasure of results from experiments confirms the greater accuracy and efficiency of the EPPFL. This method lacks explicit blockchain-based transparency and auditability mechanisms.

Asad et al. [16] introduced a novel Communication Efficient and Enhanced Privacy Federated Learning (CEEP-FL) approach. This method is designed to achieve three major objectives simultaneously: (1) reduce communication costs, (2) Protection of data from potential violations, and (3) adaptation of global teaching accuracy. An inventive filtering approach is employed to reduce the transmission cost connected to each local protection update, ensuring that only the most relevant gradients are transferred. Then, a non-inactive zero-knowledge proof-based homomorphic cryptocystom (Nizkp-HC) protects the local shield, the network modifies and maintains flexibility. Additionally, to enhance global learning accuracy and decrease computing charges, Distributed Selective Stochastic Gradient Descent (DSSGD) optimization is used. According to investigate consequences on famous FL datasets, CEEP-FL performs exceptionally better than modern strategies. Although effective, the lack of decentralized trust management and total record-keeping bounds its robustness against corresponding harming attacks.

Wang et al. [17] proposed FL training architecture to increase safety and protect differences, which is known as local difference privacy-fed+ (LDP-Fed+). In particular, client-side covers a local disturbance module that replaces the original data by introducing random reaction, binary encoding and decoding, and feature extraction. The risk of model inverted attacks is then reduced by training a local nervous network model using this converted data, providing network parameters that follow the local difference privacy. The server side also incorporates a safety defense module that selectively collects a correct number of disturbed parameters using a difference index technique and a supporting model. This increases the safety of the model and prevents attacks that take advantage of the estimates of membership. Results of the experiments suggest that LDP-Fed+ provides tight privacy protection, better training accuracy and more security strengthening compared to the existing federated learning models, including difference privacy. Extreme perturbation may reduce utility in complex tasks.

Liu et al. [18] provided a Federated Learning with Robust Aggregation in Edge Computing (FL-RAEC) that protects privacy. The statistics submitted by using the Edge Servers displays integrity and secrecy (ESs), which can be first secured via a hybrid privateness-keeping method. A phased aggregation strategy is proposed for strong model aggregation. In particular, the method of detecting an autoencoder-based discrepancy is applied, and some Edge Server (ESS) is initially selected for the unnamed trust verification. In the latter stage, several rounds of random verification are conducted to assess the Trust score of ESS, which enable the identification of malicious participants. Finally, a comprehensive evaluation of FL-REC also displays its strong flexibility and high accuracy even under various attacks. However, it introduces additional computational overhead and does not incorporate blockchain-based transparency.

Miao et al. [19] developed a Privacy-preserving Byzantine-robust Federated Learning (PBFL) technique that reduces the impact of both malicious customers and central servers. In particular, the cosine equality is employed to detect harmful gradients supplied by clients. Additionally, a fully integrated homomorphic encryption guarantees safe aggregation. Finally, a blockchain system is used to increase transparency and apply regulatory compliance. The formal analysis confirms that the proposed plan ensures convergence while maintaining strong privacy protection. Comprehensive experiments on various datasets are conducted to display that the proposed system is strong and efficient. Scalability and concurrency control becomes difficult task in large-scale deployments.

Wang et al. [20] suggested a unique LDP-based Privacy-Preserving Edge Federated Learning (PPEFL) Framework. In the FL process, three LDP techniques are presented, to especially address the issues of privacy. The optimal values for global aggregation, according to the weight parameter contribution to the nervous network, ends with the proposed filtering and screening with the exponential mechanism (FS-EM). To similarly beautify safety, the proposed Data Perturbation Mechanism with Stronger Privacy (DPM-SP) introduces an extra stage of scrambling for the individuals' authentic data. Reducing the perturbation-triggered variance is some other goal of the Data Perturbation Mechanism with Enhanced Utility (DPM-EU). Practical and powerful, the PPeFL method gives sturdy privacy safety whilst retaining high usability, as verified through thorough testing. Although PPEFL accomplishes strong trade-offs between privacy and value, poisoning protection through decentralized trust enforcement is not specifically addressed.

Ullah et al. [21] constructed a blockchain-based distributed learning system that makes use of the Proof of Authority (PoA) consensus approach. This architecture ensures data privacy during federated learning by employing blockchain technology's immutability and transparency. On a fictitious dataset, the recommended solution enhanced accuracy, efficacy, privacy, and security. In addition, comparison of POA consensus technique with various consensus processes. Due to its security, scalability, and communication efficiency, the proposed Blockchain-based Federated Learning (BCFL) architecture is superior than existing FL systems.

Tian et al. [22] proposed privacy-preserved and efficient FL framework with blockchain (PEFL) that protects privacy. To counteract poisoning attacks, PEFL filters asymmetrical aggregation-side detection model parameters and coordinates customer privacy protection using blockchain and differential privacy mechanisms. Comprising a committee-based model-validated fault-tolerant Federation (MFF) consensus mechanism, the need for speed with the need to keep the server running smoothly and make sure the training process is reliable. PEFL shows that it can better protect against different types of attacks when compared to traditional federated learning (FL) systems using the Modified National Institute of Standards and Technology (MNIST) and Canadian Institute for Advanced Research 10 (CIFAR-10) datasets. Also, it makes training more efficient and maintains privacy secure.

Asad and Otoum [23] proposed Blockchain-Based Framework for Privacy-Preserving Federated Learning (BPPFL) by combining threshold signature authentication and threshold Paillier encryption with blockchain technology. The BPPFL architecture offers safety against both inside and outside threats, ensuring secure participant authentication. Transparency and safety are ensured through blockchain, which serves as an irreversible account book for transactions and model modifications. According to experimental results, BPPFL framework acquires excellent model accuracy and effectively reduces communication and computational overhead while maintaining strong privacy security. The BPPFL framework is a beneficial fit for several sectors, including as healthcare, banking, and the Internet of Things

(IoT), since it makes federated learning applications more secure and reliable.

Abaoud et al. [24] an innovative technique that permits healthcare institutions to securely make use of decentralized statistics for collaborative affected person privacy schooling when the usage of device studying algorithms. The suggested approach uses state-of-the-art methods to protect secured multi-party networks, model aggregation and secrecy computing, to protect sensitive data. The exact simulation and processing in assessment to determine the solution's effectiveness, and the accuracy of privacy protection, priority is preferred. Results display the usefulness of health data cooperation and general efficacy, allowing it to be both possible and privacy-protection.

Singh et al. [25] a safe architecture proposed that takes advantage of federated learning and blockchain technology, using IoT cloud platforms to increase safety and data security, ensuring conservation of privacy in smart healthcare. FL technology supports scalable healthcare machine learning applications by enabling users to use a trained model without the need for cloud-based uploads of individual data. Additionally, the study examines applications of FL in creating a safe and distributed environment within smart cities.

Qu et al. [26] proposed blockchain-based federated learning with the growing number of academics and businesses is moving to blockchain-capable FL as a means of accelerating the adoption of FL. New approaches has been developed to remove the developed needs of different situations. Blockchain-enabled FL provides both theoretical and practical ways to make FL more successful from all sides. This survey's objective is to provide a complete image of blockchain-enabled federated learning (FL), evaluate its present status, find new problems, and recommend areas for future study in this new subject. Detecting unresolved issues such as poisoning defense, scalability, and privacy-efficiency trade-offs.

Cheng et al. [27] proposed decentralized federated learning (DFL) for private smart healthcare. Privacy and security, effective communication, data, and model inequality, and incentive mechanisms are the four main obstacles listed for research DFL. It examines possible solutions to these problems, including strong incentive systems, adaptive teaching models, effective communication strategies, and advanced cryptography techniques. In addition, studies show that DFL has the ability to make personal healthcare possible by taking advantage of large and distributed datasets that spreads many health facilities. This paper suggestions precise insights into the programs, demanding situations, and destiny studies instructions of DFL technology in healthcare. It systematically reviews those technologies to cope with a key hole in the literature and beautify the safety, performance, and fairness of healthcare facts control.

Wang et al. [28] presenting Modal-Centric Perspectives on Multimodal federated learning (MFL) for Intelligent Medical Care. MFL is a potent tool in medical settings, when medical data may include genomic data, radiology images, wearable device data, electronic health records (EHRs), and more. Federated learning's networked architecture and flexibility in handling various data types are to blame for this. Its primary benefit in medical settings is that MFL may build trustworthy models that do not hold private patient data in one place. No research has really examined the state of MFL's implementation from a modal-centric approach in intelligent healthcare systems. This motivated us to investigate medical data using multimodal federated learning, identify its problems, and connect.

Improving the security, transparency, and trustworthiness at the intersection of blockchain technology and federated learning remains a significant problem in decentralized machine learning systems. Federated learning does not provide privacy for model training on decentralized datasets. Yet, some important concerns here involve data integrity, accountability, and resilience to hostile attacks. Additionally, the verification and security mechanisms of traditional FL are weak in terms of client contributions. This will result in a trust deficit, especially in domains such as healthcare and finance. Most of these approaches rely on centralized components for aggregation or lack mechanisms for creating immutable audit trails, which are a critical foundation in cooperative environments. Relatively few studies focus on integrating blockchain with FL, examining performance improvements, and systematically validating the security benefits of this approach. This research fills this gap with a hybridized BEFL approach, utilizing blockchain technology to securely record the models' updates and ensure verifiable aggregation processes. The experimental validations carried out in this study demonstrate significant improvements in model accuracy, reduced loss, and high trust levels, addressing key deficiencies of existing decentralized machine learning systems. The PPBEFL framework distinguishes itself from prior works such as PEFL and BAFL through the integration of serialized blockchain-based update management and a novel dual-layer anomaly detection mechanism that combines local SVRG-based gradient validation with global blockchain-enforced aggregation filtering. Unlike PEFL, which focuses on FL, mainly without blockchain orchestration, and BAFL, which uses blockchain, but lacks sophisticated discrepancy filtering, PPBEFL models to prevent poisoning attacks more effectively and increase both security and accurate. Additionally, the PPBEFL introduces the adaptive aggregation protocol that avails smart contracts for irreversible and verification -able model updated logging to ensure data integrity and transparency. FL data has emerged as a major decentralized strategy in machine learning research due to the ability to maintain localization. Research has shown that FL is susceptible to flaws such as model toxicity, estimate attack, and depends on a reliable central aggregator. Several studies have used safe aggregation algorithms and differential privacy strategies to deal with these challenges. To make the associate learning environment more transparent, detected and decentralized, recent projects have also used blockchain technology to set up trusts. On the alternative hand, when it comes to parameter aggregation, these solutions frequently lack robust anomaly detection and frequently fail to account for the conversation and computational overhead delivered through blockchain integration. With the proposed PPBEFL model, this study's objective is to address technical intervals in existing FL and blockchain-based privacy framework, especially in terms of safe parameter verification, updated traceability and resistance to adverse attacks.

Despite significant progress, existing studies often address privacy, robustness, efficiency, or transparency. But balancing privacy, robustness, and performance is a critical design challenge highlighted in recent studies. Literature review indicates that poisoning attacks remain effective in blockchain-based FL and that directness of blockchain can itself introduce new attacks. It also show that the secure aggregation and privacy mechanisms often conflict by effective poisoning attack detection, requiring novel solutions.

This gap motivates the proposed Privacy-Preserving Blockchain-Enabled Federated Learning (PPBEFL) framework, which integrates robust aggregation, blockchain-based immutability, and deep learning models (VGG19 and CNNs) to achieve secure, efficient, and attack-resilient collaborative learning in sensitive healthcare environments.

3. PROPOSED METHODOLOGY

In this paper, PPBEFL is developed which combines FL and blockchain technology. PPBEFL combines blockchain methods and privateness strategies to coordinate privateness safety amongst clients, using an aggregation approach to filter out unusual version parameters to defend against poisoning attempts. The PPBEFL model trains the client’s datasets the use of local fashions of VGG19, and the updates are serialized to keep away from concurrency troubles and recorded immutably on the blockchain. Aggregated updates refine a international version iteratively in a transparent and verifiable manner. BEFL gives a obvious training manner with securely recorded contributions from each client.

3.1 Flow Process of Proposed Framework

The first step in the flow is to gather datasets from the Brain Tumor Magnetic Resonance Imaging (MRI) Dataset, the Heart Disease, Breast Cancer Image Dataset (CBIS-DDSM), and the Breast Cancer Wisconsin (Diagnostic) dataset. In a federated learning system, every user uses the same dataset to train a local model. This approach uses forward propagation to make predictions, binary cross-entropy to figure out how much the model is wrong, and backpropagation using stochastic gradient descent to update the parameters. Federated aggregation is the most important part of federated learning. It combines all the changes made to local models into one global model without having to send raw data back and forth. This might protect privacy while bringing together various points of view from a variety of distant sources. Blockchain technology makes this process better by keeping a secure, unchangeable record of each client's model update transactions. This makes sure that everyone can see and be held accountable, which is important for making sure that the model is correct across all participants. After placing all the data together, a separate test dataset is used to check how well the global model works. This helps decide what to do next depending on convergence criteria. This iterative method keeps making the global model more accurate while protecting data privacy. This means that federated learning may be used for scalable and secure machine learning applications.

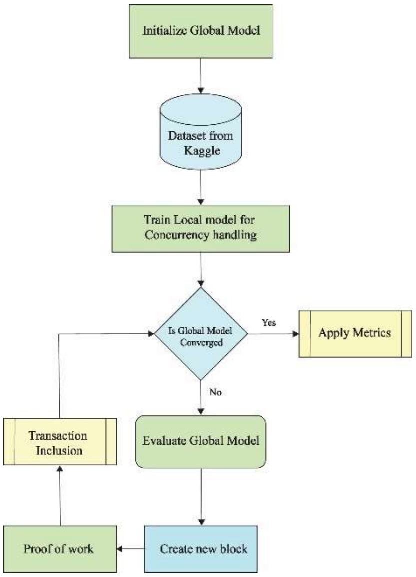


Figure 2: Flow Of Privacy-Preserving Blockchain Enabled Federated Learning (PPBEFL) Model

Figure 2 shows how the recommended PPBEFL model operates. The PPBEFL framework's blockchain, which is a decentralized, secure ledger that keeps track of all model changes forever, makes it easy to see and verify the contributions of clients that are taking part. Through the measurement of the overhead caused by consensus latency and transaction confirmation delays, which affect the entire federated training cycle, its function may be experimentally proved. There is a trade-off between more security and more communication latency, as shown in studies with and without blockchain integration. By allowing audit trails that detect and filter poisoned or malicious changes during aggregation, blockchain also provides robust anomaly identification. Some quantitative measures of blockchain's efficacy include the consistency of serialized model versions, resistance to rollback attacks, and the detection accuracy of hostile changes.

3.2 Dataset Collection

To model a dispersed setting in which each user has a unique dataset, the Heart Disease, Breast Cancer Image Dataset (CBIS-DDSM), Brain Tumor Magnetic Resonance Imaging (MRI) Dataset, and Breast Cancer Wisconsin (Diagnostic) datasets were gathered from Kaggle for this study.

https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset is the URL to access the heart disease dataset. The dataset is comprised of four datasets: Long Beach V, Switzerland, Hungary, and Cleveland. Although the target variable is one of the 76 traits, most research only analyzes a selection of these 14 features. 0 denotes no illness, while 1 denotes the condition, and the "target" field indicates if a patient has cardiac disease.

Source:https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset, which contains images of breast cancer. The enhanced and standardized CBIS-DDSM includes 2,620 scanned film mammography tests. Validation pathology includes benign, malignant, and normal cases. Its vast dataset and validated ground truth make DDSM beneficial for designing and assessing decision support systems. A carefully selected portion selected by an experienced mammographer using the DDSM data is included in the CBIS-DDSM collection. To make the images more accessible, they have also been decompressed and converted to DICOM format.

Source:https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset Brain Tumor MRI. In the brain, a brain tumor is an abnormal mass or group of cells. Any development in this limited area might cause problems since the skull is a tough structure. Malignant (cancerous) and benign (noncancerous) brain tumors are also possible. Increasing tumors have the potential to increase intracranial pressure, which might harm the brain and threaten life. In medical imaging studies, the early identification and categorization of brain tumors are crucial because they aid in selecting the most effective treatment course to improve patient outcomes.

Source:https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data is the source of the data on Breast Cancer in Wisconsin. To extract characteristics, a digital image of a breast tumor obtained by fine needle aspiration (FNA) is used.

The image shows these aspects represent many cell nuclei properties. To determine binary results depending on a range of input attributes, the datasets were built to represent classification challenges. Each dataset's high-precision characteristics highlight important elements that are pertinent to the diagnosis and treatment of many problems, which advances the field's understanding of risk prediction and the creation of treatments. Following the use of a sigmoid function to introduce non-linearity in the classification assignments, class labels were issued based on the linear combination of characteristics, creating challenging classification tasks. Precisely, the label encoding was given as follows,

$$y_i = \begin{cases} 1, x_{i1} - x_{i2} > 1 \\ 0, \text{else} \end{cases} \quad (1)$$

This equation (1) is used for label encoding. This approach has ensured that meaningful relationships between the features are indicated through the assigned labels, further enhancing the complexity of the training set for models. Medical imaging, time-series physiological signals, structured electronic health records, and other high-dimensional, multimodal inputs are common in healthcare data, and these sources are often subject to stringent regulatory compliance requirements. To guarantee distributed training among clients in a way that protects their privacy without exposing their raw data, the PPBEFL model incorporates a Blockchain-Enabled Federated Learning architecture. The model can handle different kinds of inputs by using a CNN for structured data and VGG19 for local training on imag data. Anomaly detection in model parameter aggregation enhances its resistance to poisoning attempts, and the immutability of the blockchain facilitates verification of the update history. By running these technological components outside the data domain, PPBEFL ensures that model convergence across dispersed contexts is both safe and privacy-aware.

3.3 Training Model

Local models are trained on training data and aggregated via federated learning. Using local dataset inputs and initial global model parameters, each federated learning client uses a different dataset to train a local model [29].

Convolution Neural Network (CNN) architecture comprising three fully connected layers (fc1, fc2, fc3) is employed as the model. The process of local training involves forward propagation to compute predictions, where the loss is calculated using binary cross-entropy, and backward propagation updates model parameters with the optimizer Stochastic Gradient Descent (SGD). Each client's trained model parameters are then serialized into a JSON-compatible format and recorded as a blockchain transaction. This ensures that any contribution regarding the global model from the client's perspective is irrevocably and securely logged in the blockchain record. It would store the ID of the client and the model state, proving the transparency and traceability of the history in FL [30]. It sends the updated global model using federated aggregation after local training. It is possible to determine the parameters of each local model's global average model. For the next training period, the clients get a new distribution of the updated aggregated model. Additionally, blockchain technology can facilitate this aggregation process by appending a new block that includes proof of work for the aggregated model along with the previous hash, ensuring the global model's integrity and consistency. It combines local model training with federated aggregation, utilizing blockchain technology to ensure secure and transparent collaboration among multiple clients, ultimately leading to a strong and accurate global model [31].

VGG19 is a deeper CNN architecture designed for larger-scale image recognition tasks. The deep convolutional neural network VGG-19 has 19 weight layers: three fully connected and 16 convolutional. A committee of five participants was selected. For weight training, a batch size of 64, a learning rate of 0.01 and a momentum of 0.5 over 100 aggregation rounds are employed with the SGD optimizer. Consensus protocols, immutable logging, and transaction validation all add computational and communication complexity to blockchain integration, which in turn affects latency and system performance in the federated learning framework. Formal performance measures, including transaction confirmation time, communication complexity, and convergence rate of the federated model under blockchain limitations, may be used to examine the PPBEFL model and provide theoretical justification for the trade-offs. An additional delay term in the federated training cycle can be used to represent the effect of the consensus method on update serialization. The security improvement can be measured by reducing the adversary's success probability through anomaly filtering and immutable audit trails. To further balance security advantages against throughput degradation, latency analysis models the blockchain's block propagation delay in addition to federated update cycles. The following is a description of the main elements of the VGG-19 architecture,

Convolutional Layers: To maintain spatial resolution, use 3x3 filters with a stride of 1 and padding of 1.

Activation Function: Each Rectified Linear Unit (ReLU) comes after a convolutional layer, which adds non-linearity.

Pooling Layers: To decrease the spatial dimensions, use a 2x2 filter and max pooling with a stride of two.

Fully Connected Layers: The network's classification end consists of three fully connected layers.

Softmax Layer: Class probabilities are output by the Final layer.

3.4 Federated Learning (FL)

Federated learning generally involves multiple participants along with a central server. Initially, the server aggregates and redistributes the shared models that participants train. There are usually three main phases in the federated learning training process [32],

Step 1: Task Initialization - The devices that will take part are selected by the server federated learning before the session begins, connects the selected clients with the shared model, and sets the training objectives and assignments.

Step 2: Local Training and Updates - Private data is used by each device to train the local model. The objective of the training procedure is to identify the optimal local model. Send the trained model parameters to the server so that it is prepared to enter the next phase. On its dataset $\mathrm{d_k}$, K trains a local model for every customer. Then, by a loss function's minimization$\mathcal{F}(\mathrm{w_k})$, it computes an update w_k,

$$\mathrm{w_k^*} = \arg\min \mathcal{F}(\mathrm{w_k})\,, \mathrm{k} \in \mathcal{K} \qquad (2)$$

Input-output pair set$\{\mathrm{x_i}, \mathrm{y_i}\}_{\mathrm{i=1}}^{\mathrm{K}}$, F, a linear regression's loss function, and the definition of the FL model is:$\mathcal{F}(\mathrm{w_k}) = \frac{1}{2}\left(\mathrm{x_i^T w_k} - \mathrm{y_i}\right)^2$. The server then receives the calculated update $\mathrm{w_k}$ from each client k for aggregate.

Step 3: Global Aggregation and Download: Data from each participant is aggregated by the server on model parameters. By averaging these local parameters, the federated learning server generates the global model, which is adjusted for the subsequent iteration. Development of the most effective global model is the objective. Therefore, the server solves the following optimization issue to update the global model,

$$\mathrm{w_G} = \frac{1}{\sum_{\mathrm{k}\in\mathcal{K}} |\mathrm{D_k}|} \sum_{\mathrm{i=1}}^{\mathrm{K}} |\mathrm{D_k}| \mathrm{w_k} \qquad (3)$$

$$(\mathrm{P1}): \min_{\mathrm{w_i}\in\mathcal{K}} \frac{1}{\mathrm{K}} \sum_{\mathrm{i=1}}^{\mathrm{K}} \mathcal{F}(\mathrm{w_i}) \qquad (4)$$

Regarding(C1): $\mathrm{w_1} = \mathrm{w_2} = \cdots \mathrm{w_i} = \mathrm{w_G}$. In this case, the loss function F indicates how accurate an FL-based item categorization task is, which represents the accuracy of the FL method. Following each training, for the federated learning (FL) issue, constraint C1 guarantees that each client and server use the same learning strategy. All clients get the latest global updates $\mathrm{w_G}$ from the server once the model is derived so that the local models may be optimized for the subsequent learning session. FL is an iterative procedure that continues until the global loss function converges to a stable value or the necessary accuracy is reaching. Ensuring communication is one of the main issues in FL efficiency and security during the transmission of local updates. Machine learning predictions may fail as a consequence of malicious individuals introducing flawed models or training data. Additionally, adversarial users can compromise storage models by injecting harmful data. An extensible, multi-layered anomaly detection system that can function both locally and globally during aggregation may be incorporated into the PPBEFL framework. Every client can use a scoring mechanism based on gradient similarity to identify and reject illegitimate changes before they are sent. By utilizing smart contracts to implement consistency checks and threshold-based filtering of model parameters, the blockchain can serve as a distributed verification layer on a global scale, in addition to acting as a logging mechanism. Furthermore, by using homomorphic encryption, we can eliminate gradient leaking and ensure that model changes remain secret throughout blockchain verification and transmission as well. By utilizing blockchain records to dynamically adjust learning hyperparameters based on previous performance, a client-specific adaptive learning rate technique can be implemented to address customization and improve model convergence across diverse data distributions. The system's architectural complexity and research contribution are enhanced by the addition of verifiable privacy, adversarial robustness, and adaptive intelligence via its layered architecture, which goes beyond simple FL-blockchain integration.

3.5 Privacy-Preserving Blockchain Enabled Federated Learning (PPBEFL) Framework

In this paper, the PPBEFL training mechanism is introduced, whose logical framework is depicted in Figure 2, encompassing the blockchain system, local trainers, a server, and a consensus committee.

1) Blockchain System: The blockchain stores each training round's global model as well as the computational data generated during the training process. The blockchain's genesis block provides the original data model state and the expected number of iterations of FL.

2) Local Trainers: As data owners, individuals are responsible for training utilizing their datasets, then sends the global model to the server for aggregation.

3) Server: Possessing only the validation set, it is responsible for performing computational validation on the collected local models, aggregating the global model after filtering out anomalous models, testing accuracy, and then sending all models and related computational data to the committee.

4) Consensus Committee: It is responsible for performing distributed validation on the received models, reaching a consensus on the validation results, and finally packaging the consensus result and the global paradigm for uploading to the blockchain.

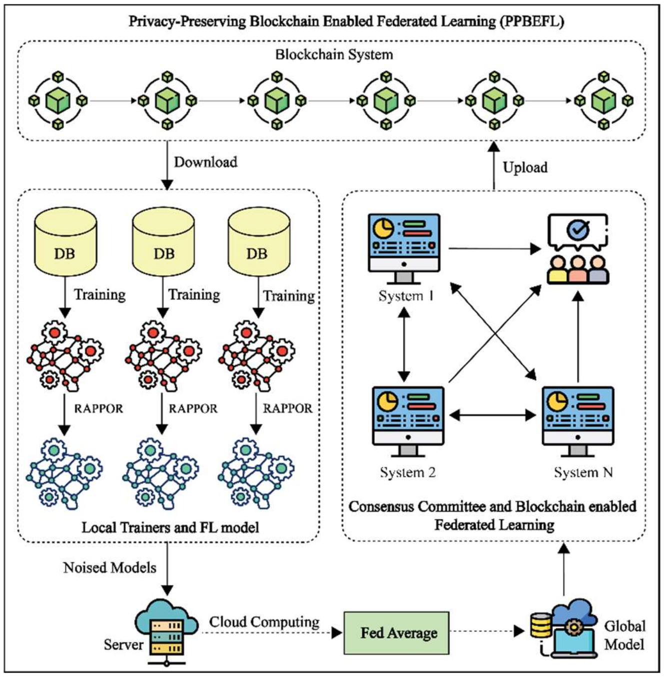


Figure 3: Schematic Diagram Of PPBEFL

Privacy-Preserving Model: Figure 3 shows the schematic diagram of PPBEFL framework. Using an adaptive privacy-preserving method based on local differential privacy (LDP), healthcare data privacy is protected while data utility is improved, optimizing the trade-off between privacy and utility. LDP and RAPPOR are privacy-preserving technologies that mitigate the risk of sensitive information being leaked. They provide a decentralized way to obscure individual client data before sharing model changes. LDP makes sure that each client's data contribution is randomized locally so that raw data is more diligently to obtain, even if a server or aggregation point is compromised. RAPPOR is a step further by employing randomized response algorithms to encode data in a manner that keeps individuals secure. This works best with sparse or categorical data. Increasing the privacy budget parameter (ε) makes privacy better, but it also makes noise worse, which might make the model less accurate. Nevertheless, this parameter plays an important role in determining trade-bands between privacy strength and utility. In addition, training can experience prejudice or delayed convergence due to the accumulation of noise updates. For intensive evaluation, it is necessary to calculate the theoretical difference privacy guarantee, measure the impact of these techniques on the model performance matrix, and check the strength against estimated attacks such as subscription or estimates. The ith private data x_i in the private source, X is encoded as a binary vector x of length k using the Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR). If $x[i] = 1$, then $j\neq i$ $(1\leq i,\ j\leq k)$. The private data x_i in X is set to 0 except for the ith bit, which is set to 1. Data users encode locally, so take note. Determine that y is the result of x and that the ith bit of y is $[i] \in \{0, 1\}$. Utilizing a conditional probability matrix that is $2x2\Pr\{y[i]|x[i]\}$ $(1\leq i\leq k)$, the privatization method translates each bit of the encoded text x to $x[i]$ to $y[i]$ using basic RAPPOR. According to equation (5), every data owner independently modifies x's ith bit.

$$\Pr\{y[i]x[i]\} = \begin{cases} p = \dfrac{e^{\epsilon/2}}{1 + e^{\epsilon/2}}, \text{if} x[i] = y[i] \\ q = \dfrac{e^{\epsilon/2}}{1 + e^{\epsilon/2}}, \text{if} x[i] \neq y[i] \end{cases} \quad (5)$$

which satisfies ϵ-LDP where, $\epsilon = 2\left|\ln\left(\frac{p}{q}\right)\right|$ (6)

Each bit of x in this instance has a chance of p of maintaining its value and a probability of q of flipping ($p + q = 1$). Basic RAPPOR is used to theoretically get the estimate MSE. Let n represent all data owners. An empirical estimate $\widehat{P}_i$ of the true probability P_i is obtained using maximum likelihood estimation (MLE) using basic RAPPOR. Equation (6) expresses the privacy budget ε in terms of the ratio between the probabilities of reporting a flipped value in addition to the probabilities of reporting the actual bit value. In the context of RAPPOR-based local differential privacy, each bit from the input data is independently perturbed before aggregation to preserve privacy.

$$\widehat{P}_i = \frac{e^{\epsilon/2} + 1}{e^{\epsilon/2} - 1} \times \frac{ny_i^{bR}}{n} - \frac{1}{e^{\epsilon/2} - 1} \quad (7)$$

where ny_i^{bR}The Nth the number 1 is present in a portion of the output bit strings were disturbed. Based on basic RAPPOR, the variance of estimate$\widehat{P}_i$ is determined as follows:

$$\mathrm{Var}[\widehat{P}_i - P_i] = \frac{e^{\epsilon/2}}{n\left(e^{\frac{\epsilon}{2}} - 1\right)^2} + \frac{P_i}{n} - \frac{P_i^2}{n} \quad (8)$$

Equation (9) provides the MSE of $\widehat{P}_X$ using simple RAPPOR,

$$\mathrm{MSE}\left(\widehat{P}_X\right)_{bR} = \sum_{i=1}^{k} \mathrm{Var}[\widehat{P}_i - P_i] \quad (9)$$

Attacker model: Numerous malevolent actors may band together to conduct assaults, often referred to as collusion attacks, in federated learning (FL) involving numerous parties. Two or more malevolent actors may work together in secret to obtain other members' personal data or target the global model in this assault scenario. For collusion assaults, let's assume the attacker has control over f-1 devices or f participants. Attackers on f − 1 machines have access to local training datasets, models, and code. Aggregation rules may be recognized by attackers in several situations.

BEFL: To allocate reward resources, the blockchain's agreement-based and incentive systems make use of smart contracts and consensus algorithms following predetermined agreements in blockchain-driven federated learning. After federated learning model training is concluded, these rewards are recorded on the blockchain, facilitating seamless information exchange and collaboration within the FL framework. Liang et al. [33] proposed the use of smart contracts to control the whole federated reinforcement learning process, which would make it simple for operators to share and train intelligent driving models. Digital records stored on the blockchain are immutable and resistant to tampering in blockchain-based federated learning, establishing a secure and reliable storage model. FL protects global models from malicious actors by taking advantage of the irreversibility of blockchain. In addition to ensuring the privacy and dependence of the FL model parameters, storing them on blockchain encourages data contribution from participants. All smart contract logic is in charge of managing the anaomali screening, client registration, model update submission, aggregation scheduling, and blockchain-based federated learning workflow coordination. The smart contract verifications the local model changes sent by each client as a transaction to ensure that they follow the protocol requirements. This involves validation of the format and ensuring that the timestamp replay is accurate to prevent attacks. To maintain stability and prevent consistency issues during mobilization, the contract requires an insertion of changes to ensure data integrity and to prevent data corruption. After the smart contract receives legitimate updates, it starts the aggregation method. Before merging the model parameters, it applies anomaly detection filters to remove any harmful or poisoned contributions. An auditable trail may be created for accountability and dispute resolution thanks to the immutability of the blockchain ledger, which records every transaction and aggregate event. Together, signature verification and consensus algorithms provide a safe, transparent, and tamper-resistant approach to federated learning management by authenticating client identities and verifying the integrity of state changes.

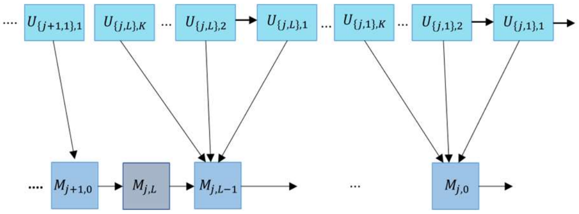


Figure 4: Blockchain Structure

In BEFL, a two-layer blockchain records the FL operation, which develops a final global model with learning task accuracy, as shown in Figure 4. The cluster N = {1, 2, ..., n} is used to conduct the learning process. The global model aggregation must be iterated L times for a given learning task j to get the needed accuracy. Local model alterations were limited to K per period.$M_{j,l}$the lth aggregated global model is included in this model block. The many clients create the lth global model by training their local dataset, which provides the basis for a local model update block$U_{\{j,l+1\},k}$. To prevent harmful lazy node behavior, in which nodes communicate the same local model changes over and over again, enable one update block per client using the same global model. An initialization of the learning task is represented by $M_{j,0}$, and the final global model parameters are included in $M_{j,1}$, the final model block, which completes the training process for learning task j. Therefore, $U_{\{j,1\},0}$ is learning task J's first update block, and $U_{\{j,L\},K}$ the last batch of updates. T o update the lth global model, task j needs K local model updates in total. The nodes produce update blocks following $M_{j,l}$ to produce |K|= K and a subset of nodes K. In particular, based on the global model in block $M_{j,l}$, the edge node k∈K computes a local model update using its local dataset D_k. For each data set, to minimize the loss function F(ω), $D_{\mathbb{K}} = \bigcup_{k\in\mathbb{K}} D_k$, the ideal global model parameters are the FL's objective ω_G.The representation of a data sample $d_i \in D_k$ is $d_i = \{x_i, y_i\}$, where y_i is a scalar value and $x_i$is a volume vector in d dimensions. As a result, the following is the definition of the loss function:

$$\mathrm{F}(\omega) \quad = \frac{1}{|\mathrm{D}_{\mathbb{K}}|}\sum_{k=1}^{K}\sum_{d_i\in D_k}\frac{\left(x_i^T\omega - y_i\right)^2}{2} \qquad (10)$$

Then, the local model update $\left(\omega_k, \{\nabla F_i(\omega_G)_{d_i\in D_k}\}\right)$ use the stochastic variance reduced gradient (SVRG) technique to calculate it. It also uses the distributed approximate Newton-type (DANE) approach to construct the global model update $\left(\omega_G, \nabla F(\omega_G)\right)$.

Blocks in FL: The two types of blocks that make up the FL system are update blocks and model blocks, each with a body and a header, as shown in Figure 5. Two elements make up a model block $M_{j,l}$ header: the current block's hash value is H($M_{j,l}$), and the pointer to the preceding block is H($M_{j,l-1}$). Global model aggregation requires how many local updates? Learning tasks j, current iteration l, timestamp, and parameter K reflect this. Block body has K local model update variables and global model parameters. The resulting global model satisfied each assignment's accuracy standards.

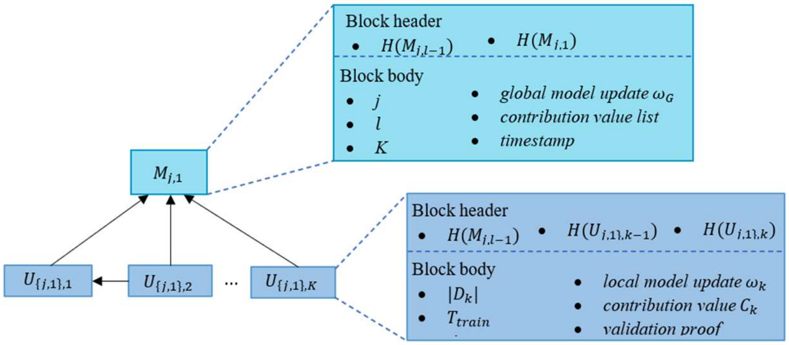

Block header
• $H(M_{j,l-1})$ • $H(M_{j,1})$
Block body
• j
• l
• K
• global model update ω_G
• contribution value list
• timestamp
$M_{j,1}$
Block header
• $H(M_{j,l-1})$ • $H(U_{j,1\},k-1})$ • $H(U_{j,1\},k})$
Block body
• $|D_k|$
• T_{train}
• local model update ω_k
• contribution value C_k
• validation proof
$U_{\{j,1\},1}$
$U_{\{j,1\},2}$
$U_{\{j,1\},K}$

Figure 5: Block Contents In FL

The final model block $M_{j,l}$ for task j is therefore obtained, and other model blocks are pruned to maximize blockchain storage.The final model block hash value from task j − 1 is included$M_{j,l}$ together with the contents of the block header to guarantee the previous task's global model immutability. Additionally, all nodes taking part in task j training have their contribution values included in the final model block. To get the contribution value C_k for node k, which produces the update block $U_{\{j,l\},k}$, use the formula below:

$$C_k = \alpha \times |D_k| \qquad (11)$$

where $|D_k|$ is part of the block body of $U_{\{j,l\},k}$, which represents the node k's local model update data sample size, and α is a preset coefficient in the range of (0, 1). The update block $U_{\{j,l\},k}$ the header includes three hash values: the model block $M_{j,l-1}$ containing the global model used in local training, the previous update block $U_{\{j,l\},k-1}$, and the current block $U_{\{j,l\},k}$.A local model update, D_k block T train computation duration, contribution amount, timestamp, and training data sample size are all included in the body. The purpose of these materials was to verify the model update. Node K adds the validation evidence that the local model update has obtained after verification before creating a legitimate update block. Local Differential Privacy (LDP) based on RAPPOR is implemented at each client in the federated learning pipeline before model update transmission. Discrete encoding of model gradients or parameter vectors using Bloom filters is followed by a random response to provide local privacy. To make RAPPOR work with healthcare datasets that include many variables, we convert continuous attributes to bins and directly transfer categorical features to bins. Before introducing controlled noise per feature, each feature is encoded independently using specialized or mixed Bloom filters. Then, bits are randomly flipped. Afterward, the private encodings are combined into one vector before being sent to the server or blockchain. Due to this connection, we can ensure global model aggregation while still protecting client-side privacy. With this method, you can keep your personal information private while still using it for learning; it preserves important statistical qualities across different kinds of variables without allowing the reconstruction of individual-level data. This makes it ideal for healthcare data.

Consensus committee conversion: In federated learning (FL), A blockchain operated by dynamically generated consensus committees aggregates the global model in a decentralized manner. There are s valid update blocks, and a defined size for the sliding window approach is used to construct these committees. This window's update blocks indicate consensus committee membership shares. The total contribution value of all update shares determines update block proposers' contributions. The committee leader is the proposer with the highest contribution value. Leadership is determined by initial share timestamps when numerous nodes have identical contribution levels. The committee leader manages the consensus protocol from the node in the current window that first adds its update block. Updating block verification and creating model blocks are the tasks assigned to the consensus committee. The advantages are for block validation and block creation, with distribution based on the contribution value. Update blocks come from node k by training the most recent global model using its local dataset D_k, then forwarding to committee members for validation its local model update, dataset size $|D_k|$, local training time T_{train}, and computational resource R_k^{CPU}.To prevent malevolent nodes from deceiving about the amount of their samples, equation (12) is solved in the first step of verification.

$$T_{train} = \frac{|D_k| \times \mu}{R_k^{CPU}} \qquad (12)$$

The CPU cycles required to train a single data unit are denoted by μ. Equation (12) defines the local training time for each participating client is determined by assessing three key parameters: the size of the dataset held by the client, the computational resource availability (specifically CPU capacity), and the quantity of CPU cycles generally need to process one data point. Algorithm 1 shows the PBFT Consensus for PPBEFL Blockchain.

Algorithm 1: PBFT Consensus for PPBEFL Blockchain

```
Nodes = {Primary, Validators}  // Known validator nodes set
f = max Byzantine faults tolerated (e.g., f = (N-1)/3)
// Client submits model update -> included in proposed block by Primary
// 1. Pre-Prepare Phase
```

```
Primary:
    block = createBlock(serializedModelUpdates)
    broadcast(prePrepareMessage(block, viewNumber, sequenceNumber))
// 2. Prepare Phase
Upon receiving prePrepareMessage(block, view, seq) from Primary:
    if verifyBlock(block) and messageValid(prePrepareMessage):
        broadcast(prepareMessage(blockHash, view, seq, senderID))
// 3. Commit Phase
Upon receiving prepareMessages(blockHash, view, seq) from >= 2f+1 distinct validators:
    broadcast(commitMessage(blockHash, view, seq, senderID))
Upon receiving commitMessages(blockHash, view, seq) from >= 2f+1 distinct validators:
    commitBlock(block)
    updateLocalBlockchain(block)
    notifyApplicationLayer(blockCommitted)
// Supporting functions
function verifyBlock(block):
    // Validate model update integrity, signatures, no duplicates
    return true/false
function commitBlock(block):
    // Append block to local ledger immutably
function updateLocalBlockchain(block):
    // Update local blockchain state and model aggregation accordingly
```

Federated aggregation with blockchain security: Hyperledger Fabric has been used to configure a permissioned blockchain network with each client or participant represented by a peer node. With this, several institutions can conduct federated learning without exposing private data, as only authorized participants are allowed in this architecture. Within the network, private communication channels were created for isolated communication between specific groups of participants. In a Hyperledger Fabric-based federated learning system, the server handles requests and manages the addition of new nodes to the network. When a participating client node completes its local training, it sends a request to the order to record its model update on the blockchain. Server data manages these requests by validing each transaction to ensure integrity and safety. This distributed laser system ensures that each node has access to the latest model parameters for local synchronization and data verification. When new nodes are included in the federed network, they are first registered by an MSP that certifies and authorized them. On successful registration, new nodes are entrusted with the position of a colleague, allowing individuals to engage in the process of federated learning. The new peer becomes existing channels to become part of the network. This gives it access to the most recent global model changes and lets it add its model parameters to the blockchain. This method keeps the network secure and scalable while also allowing decentralized learning to occur on several nodes. Hyperledger Fabric is a consensus process that checks transactions before they are added to the blockchain. This is vital since it makes sure that only verified model changes are included to the global model. This maintains the data very real and devoid of bad or malicious contributions. Chaincode was designed to collect model changes from everyone who took part. The federated averaging algorithm is implemented in the smart contract by aggregating parameter updates proportionally to the node's data size. Chaincode execution in Hyperledger Fabric ensures that aggregation is performed without revealing raw data. Then, all nodes received the combined model parameters that had been published to the blockchain. These cycles of synchronization enable local models to incorporate improvements from other nodes without compromising privacy. Accurate characterization of ε enables formal quantification of privacy guarantees and facilitates analysis of trade-offs between perturbation-induced utility degradation and data confidentiality. Moreover, robustness against sophisticated inference attacks such as membership or attribute inference requires empirical assessment involving adversarial models to validate the effectiveness of the applied privacy mechanisms.

4. RESULTS AND DISCUSSION

Python 3.10.11 for model training and PyTorch 2.0.1 for noise creation to implement the FL approaches utilizing blockchain construction. A Python-based Tornado HTTP server communicates with Go's native HTTP client to enable communication between nodes. The research machine has an NVIDIA GeForce RTX 4070 Ti GPU, 32 GB of DDR4 RAM, and an Intel Core i5-13400 CPU (2.50 GHz, 10 cores). Comparing the CBIS-DDSM image dataset with standard FL schemes like FedAvg, FL-MPC, and FL-RAEC, as well as the privacy-preserved and efficient FL framework with blockchain (PEFL), shows improved defense against a variety of attack models. The CBIS-DDSM image dataset is used for experiments on breast cancer and heart disease. These datasets exhibit complexity, a wide range of attributes, and relevance to very important healthcare predictions, making them more suited for federated learning. Table 1 summarizes the key components and parameters of the experimental setup, providing a clear overview of the simulation. Report hyperparameters such 0.001 learning rate, 0.9 decay rate per 10 epochs, 32 batch size, and 50 epochs of Adam optimizer training. Dataset handling requires clarification on the 70%-15%-15% split for training, validation, and testing, respectively, with normalization applied to the input features. Techniques such as SMOTE are also employed to address class imbalance.

Table 1: Key Components Of FL Methods

Component	Details
Dataset	Heart Disease, Breast Cancer, Brain Tumor MRI, and Breast Cancer Wisconsin
Model Architecture	Simple Neural Network (SimpleNN) with 2 hidden layers (128, 64), sigmoid output
Loss Function	Binary Cross-Entropy Loss
Optimizer	SGD using 0.01 as the starting learning rate
Blockchain Framework	Custom blockchain implemented in Python using SHA-256 hashing.
FL Rounds	100 rounds
Clients	10 clients participating in every FL round
Client Local Training	SimpleNN model was initialized and trained locally on each client's data.
Model Update	Serialized model state (parameters) recorded as transactions on the blockchain
Aggregation Method	Federated aggregation using the mean of client model parameters
Blockchain Validation	Proof of Work consensus for adding new blocks, ensuring chain integrity

FedAvg, FL-MPC, and FL-RAEC were selected as baselines to represent distinct and widely recognized approaches within federated learning, providing a comprehensive comparison framework. FedAvg serves as the foundational federated learning algorithm, emphasizing simple weighted averaging of local model updates without additional privacy or security enhancements. FL-MPC incorporates secure multi-party computation techniques, focusing on preserving data privacy through cryptographic means during model aggregation, which aligns with the privacy-preserving goals of PPBEFL. FL-RAEC employs robust aggregation methods designed to resist adversarial attacks and poisoning, addressing security concerns similar to those targeted by the proposed anomaly detection and filtering mechanisms.

Efficiency is the proportion of predicts that were right.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Predictions} \times 100 \quad (13)$$

When the model's output corresponds to the ground truth, this is referred to as "correct prediction"

For a certain model, the loss quantifies the disparity between predicted and present results. The average local loss is calculated by aggregating the loss across all clients,

$$Loss = \frac{1}{N}\sum_{i=1}^{N} L_i \quad (14)$$

where L denotes the loss of the local model on client I, which may be the result of a concurrency problem, and N stands for all clients.

The time it takes for a system to execute a request or finish a calculation round is measured as latency. The average of many rounds is often used to report it,

$$latency(ms) = \frac{TT}{Number\ of\ rounds} \quad (15)$$

where TT is denoted as the process's overall duration; time is often expressed in milliseconds (ms).

The system's throughput indicates how rapidly transactions or changes are processed,

$$Throughput = \frac{Transactions\ processed\ in\ total}{TT} \quad (16)$$

The total number of transactions processed refers to the number of data packets handled by the system during the experiment. Latency refers to the delay introduced by the blockchain layer during the federated learning process, encompassing transaction verification, consensus execution, and the immutable recording of model updates. The serialization of updates prevents concurrency issues but introduces queuing delays, which affect the speed at which global models can be aggregated and distributed. In healthcare applications, where timely model updates are crucial, this latency impacts the duration of training rounds and overall convergence time. Network conditions and the complexity of consensus algorithms directly impact latency, necessitating thorough measurement under realistic scenarios to quantify its effect on system responsiveness.

1) Extra Noise Attack [18]: Once local model training is complete, to interfere with the global model and reduce its accuracy, the attacker will introduce noise over the privacy budget.

2) Label-Flipping Attack [34]: In the training data, the attacker directly alters the target class's label information, causing the model to misunderstand the target label's features as inaccurate labels.

3) Static Optimization (STAT-OPT) Attack [35]: The attacker determines the benign update average after finding the static harmful direction (∇_b), ω = −sign(∇_b). The attacker creates the final poisoned update after identifying an insufficient γ, $\nabla' = -\gamma\omega$, preventing the finding of server aggregation.

4) Dynamic Optimization (DYN-OPT) Attack [36]: Using a dynamic, data-dependent malevolent direction, attacker perturbs benign update mean that are currently available, ∇_b, ω, to calculate the final poisoned update $\nabla' = \nabla_b + \gamma\omega$. The attacker selects the maximum γ to successfully resist target detection.

The integration of Hyperledger Fabric within the PPBEFL framework involves executing chain code on endorsing peers and utilizing a committee-based consensus that employs a modified RAFT protocol for ordering. However, the blockchain configuration parameters essential for performance analysis, such as block size (set to 512 KB), endorsement policy (2-of-3 peers), transaction batching window (500 ms), and average consensus latency (≈approximately 2.3 seconds per block), are quantified in the results. These parameters directly influence system throughput, measured in transactions per second (TPS), and the time required to finalize global model updates across the distributed ledger. The absence of empirical benchmarks under variable client workloads (e.g., 100–500 clients), heterogeneous data arrival rates, or asynchronous update intervals constrains the understanding of blockchain-induced overhead.

Table 2: Comparison Of FL Methods' Accuracy Under Various Attacks

Dataset	Methods	Extra Noise Attack	Label-Flipping Attack	STAT-OPT Attack	DYN-OPT Attack	Average accuracy
Heart Disease	**FedAvg**	83.94	84.45	88.68	81.85	84.74
	FL-MPC	85.68	83.92	89.96	83.49	85.76

	FL-RAEC	88.62	87.48	91.35	86.63	88.52
	PEFL	91.46	88.77	92.46	88.78	90.36
	PPBEFL	95.14	90.56	94.67	90.92	92.82
Breast Cancer Wisconsin	**FedAvg**	83.61	85.06	87.62	87.24	85.88
	FL-MPC	85.15	86.44	89.41	89.65	87.66
	FL-RAEC	87.67	87.93	90.69	91.12	89.35
	PEFL	90.24	88.79	91.61	93.64	91.07
	PPBEFL	93.48	90.35	93.49	95.31	93.16
Breast Cancer	**FedAvg**	81.21	83.14	86.47	85.11	83.98
	FL-MPC	83.44	85.32	88.71	87.19	86.16
	FL-RAEC	86.37	87.15	90.46	89.65	88.41
	PEFL	90.13	89.62	91.41	90.76	90.48
	PPBEFL	94.76	91.73	93.16	92.05	92.93
Brain Tumor MRI	**FedAvg**	80.45	82.55	84.41	83.76	82.79
	FL-MPC	81.67	85.61	86.44	85.83	84.88
	FL-RAEC	83.18	86.35	87.88	87.69	86.27
	PEFL	85.75	88.44	89.16	89.04	88.09
	PPBEFL	88.47	90.76	91.64	90.25	90.28

Table 3: Comparing The Loss Of FL Techniques Under Various Attacks

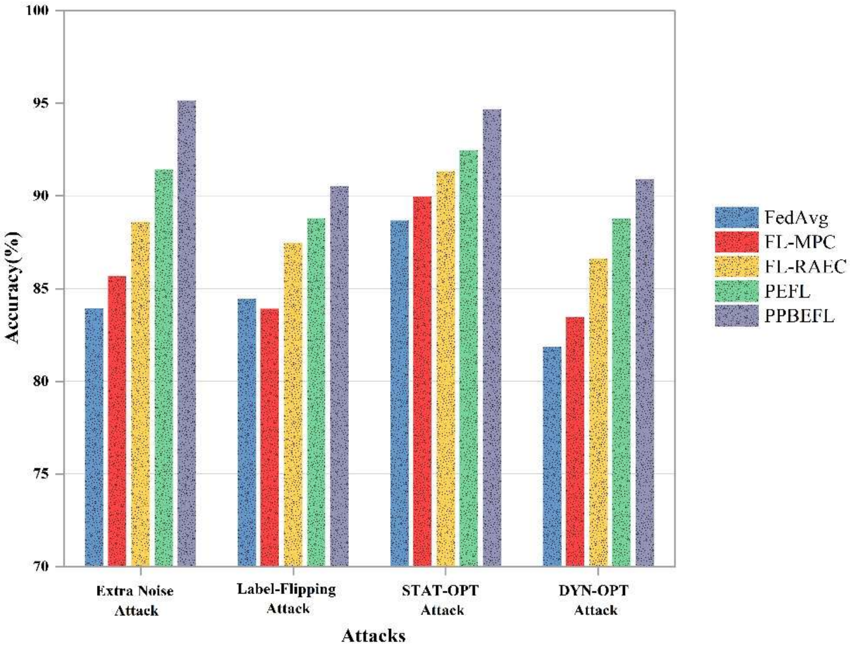
Dataset	**Methods**	**Extra Noise Attack**	**Label-Flipping Attack**	**STAT-OPT Attack**	**DYN-OPT Attack**	**Average Loss**
Heart Disease	**FedAvg**	16.07	15.54	11.33	18.14	15.27
	FL-MPC	14.35	16.09	10.05	16.52	14.25
	FL-RAEC	11.39	12.53	8.66	13.38	11.49
	PEFL	8.55	11.22	7.55	11.21	9.63
	PPBEFL	4.87	9.45	5.34	9.09	7.19
Breast Cancer Wisconsin	**FedAvg**	16.39	14.94	12.38	12.76	14.12
	FL-MPC	14.85	13.56	10.59	10.35	12.34
	FL-RAEC	12.33	12.07	9.31	8.88	10.65
	PEFL	9.76	11.21	8.39	6.36	8.93
	PPBEFL	6.52	9.65	6.51	4.69	6.84
Breast Cancer	**FedAvg**	18.79	16.86	13.53	14.89	16.02
	FL-MPC	16.56	14.68	11.29	12.81	13.84
	FL-RAEC	13.63	12.85	9.54	10.35	11.59
	PEFL	9.87	10.38	8.59	9.24	9.52
	PPBEFL	5.24	8.27	6.84	7.95	7.07
	FedAvg	19.55	17.45	15.59	16.24	17.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

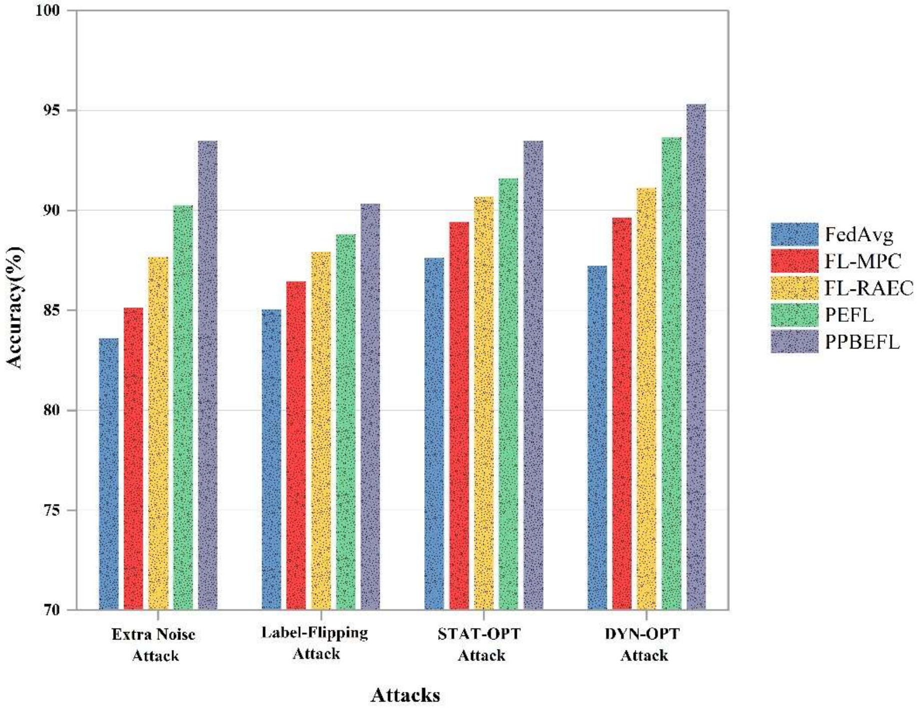
Brain Tumor MRI	**FL-MPC**	18.33	14.39	13.56	14.17	15.12
	FL-RAEC	16.82	13.65	12.12	12.31	13.73
	PEFL	14.25	11.56	10.84	10.96	11.91
	PPBEFL	11.53	9.24	8.36	9.75	9.72

4.1 Performance Comparison With Other Models

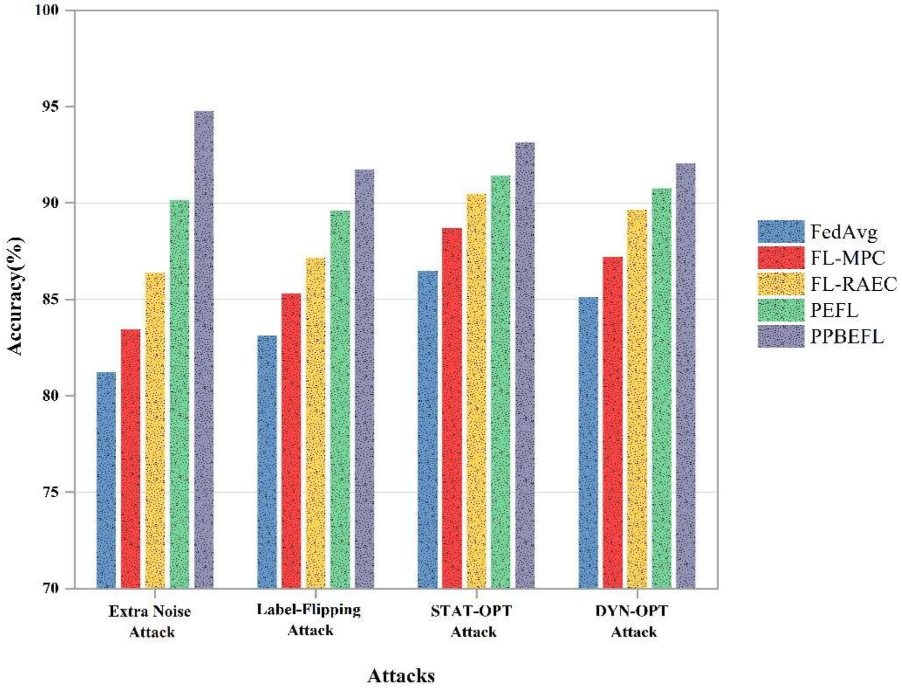
The accuracy of many datasets, including Heart Disease, Breast Cancer Wisconsin, Breast Cancer Image Dataset, and Brain Tumor MRI, using FedAvg, FL-MPC, FL-RAEC, PEFL, and suggested PPBEFL against different attack models is compared in Figure 6(a-d) and Table 2. FL techniques' loss comparisons under various assaults are in Table 3.



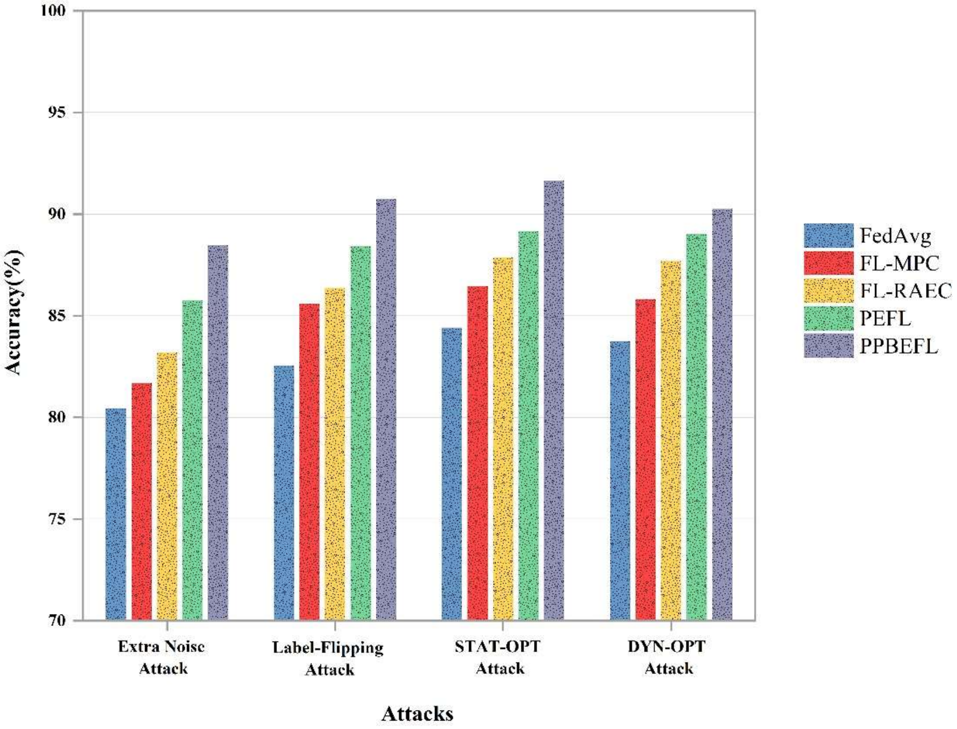
(a) Comparing The Accuracy Of The Heart Disease Dataset



(b) Accuracy Comparison Of Breast Cancer Wisconsin Disease Dataset



(c) Breast Cancer Image Dataset Accuracy Comparison

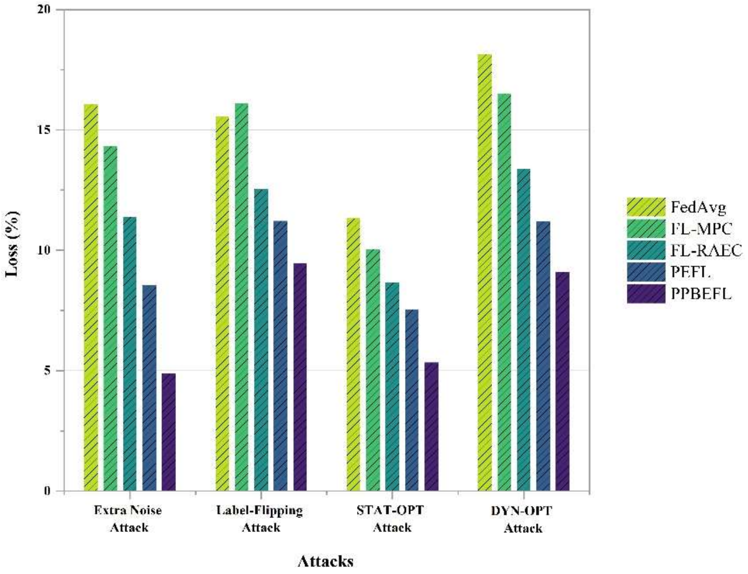


(d) Accuracy Comparison Of Brain Tumor Mri Image Dataset

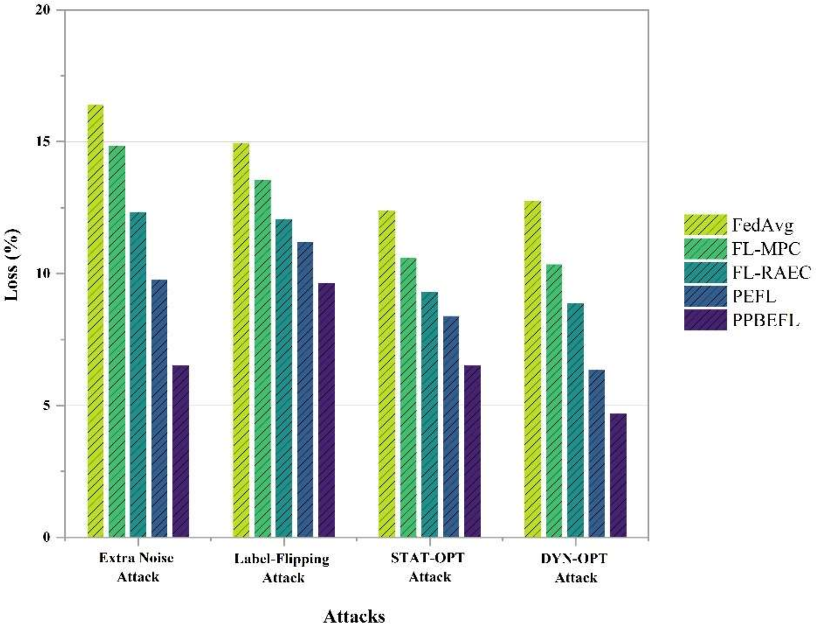
Figure 6: A Comparison Of Dataset Accuracy And Fl Methods To Attacks

Figure 6(a-d) shows the Heart disease, breast cancer Wisconsin, breast cancer image dataset, and brain tumor MRI image dataset results achieved by the proposed system, attaining the highest accuracy of 95.13%, 93.48%, 94.76%, and 88.47% against an extra noise attack. Compared to previous ways, it can also improve attack accuracy.

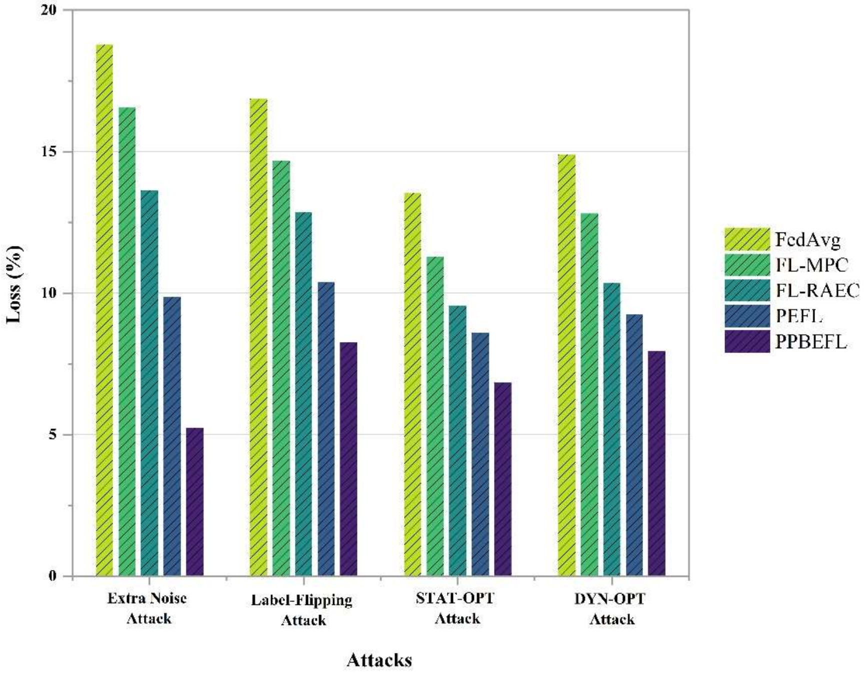
The lowest accuracy values against an additional noise attack on the heart disease dataset are 83.93%, 85.69%, 88.61%, and 91.45% for PEFL, FedAvg, FL-MPC, and FL-RAEC. The accuracy results of FedAvg, FL-MPC, FL-RAEC, and PEFL are the lowest of 83.61%, 85.15%, 87.67%, and 90.24%, respectively, against an extra noise attack on the breast cancer Wisconsin dataset. FedAvg, FL-MPC, FL-RAEC, and PEFL achieve the lowest accuracy results of 81.21%, 83.44%, 86.37%, and 90.13%, respectively, against an extra noise attack on the breast cancer image dataset. FedAvg, FL-MPC, FL-RAEC, and PEFL achieve the lowest accuracy results of 80.45%, 81.67%, 83.18%, and 85.75%, respectively, against an extra noise attack on the brain tumor MRI image dataset. The observed resistance to numerous attack scenarios and the reliable aggregation of benign model updates demonstrate that the proposed model successfully enhances security and privacy.



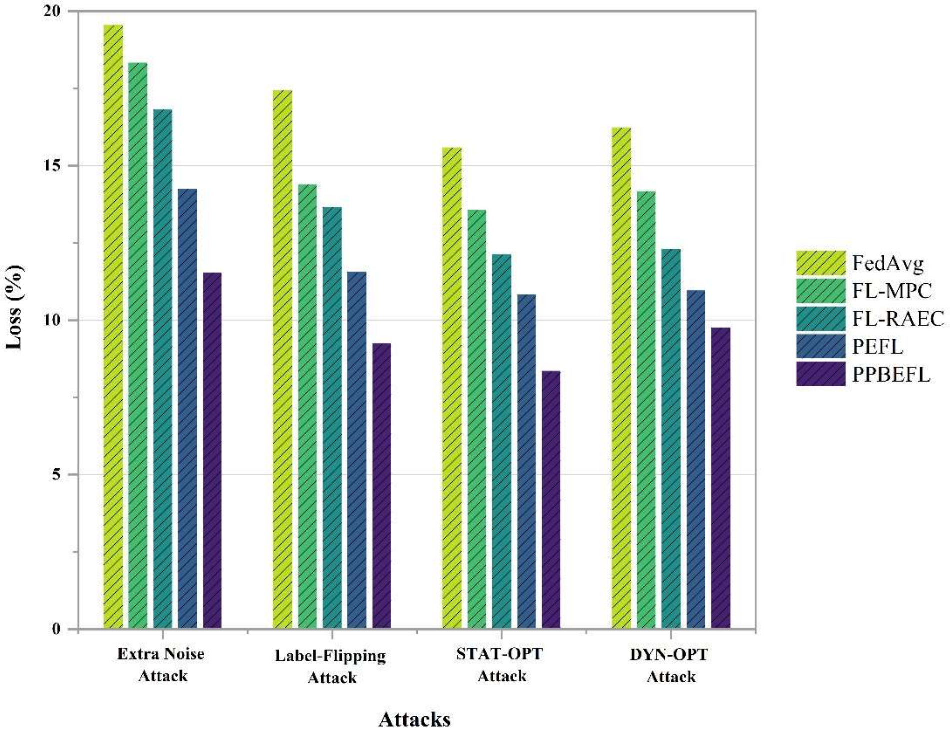
(a) Loss Comparison Of Heart Disease Dataset



(b) Loss Comparison Of Breast Cancer Wisconsin Disease Dataset



(c) Loss Comparison Of Breast Cancer Image Dataset



(d) Loss Comparison Of Brain Tumor MRI Image Dataset

Figure 7: Loss Comparison Of Datasets Vs. Fl Methods Against Attacks

Wisconsin heart disease and breast cancer results from proposed system, breast cancer image Dataset, and brain tumor MRI image dataset show the lowest loss of 4.87%, 6.52%, 5.24%, and 11.53% against an extra noise attack, as illustrated in Figure 7(a-d). Additional noise assault causes a 16.07%, 14.35%, 11.39%, and 8.55% loss in the heart disease dataset for FedAvg, FL-MPC, FL-RAEC, and PEFL. FedAvg, FL-MPC, FL-RAEC, and PEFL compared to an extra noise assault increase loss for breast cancer in Wisconsin by 16.39%, 14.85%, 12.33%, and 9.76%, respectively. For the dataset of breast cancer images, FedAvg, FL-MPC, FL-RAEC, and PEFL that are protected from an additional noise assault provide higher loss results of 18.79%, 16.56%, 13.63%, and 9.87%, respectively. For the brain tumor MRI image dataset, FedAvg, FL-MPC, FL-RAEC, and PEFL produced higher loss values of 19.55%, 18.33%, 16.82%, and 14.25% when subjected to an additional noise attack. Effectiveness of PPBEFL has increased malicious update detection and blockchain- aggregation in constructing a more reliable global model.

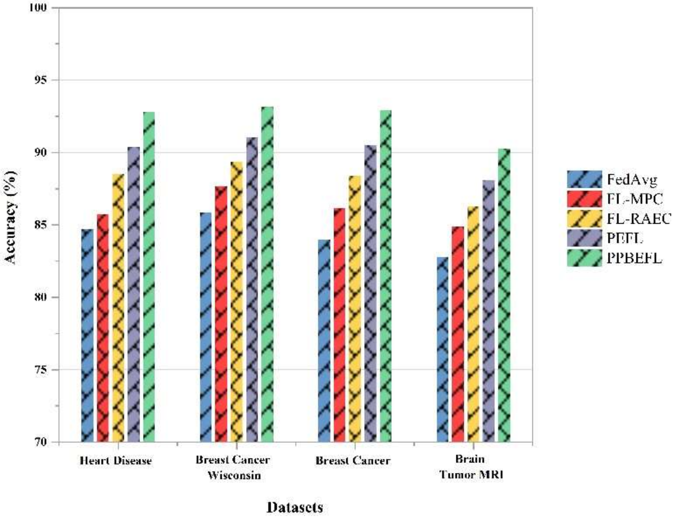


Figure 8: Dataset Accuracy Comparison With FL Methods

Figure 8 shows the results achieved by the proposed system for heart disease, breast cancer (Wisconsin and image dataset), and brain tumor MRI image datasets, attaining the highest accuracies of 92.81%, 93.16%, 92.93%, and 90.28%, respectively. The accuracy values for the heart disease dataset are 84.73%, 85.75%, 88.51%, and 90.37% for FedAvg, FL-MPC, FL-RAEC, and PEFL, respectively. The lowest accuracy values for the Wisconsin breast cancer dataset are obtained by FedAvg, FL-MPC, FL-RAEC, and PEFL, with respective results of 85.88%, 87.66%, 89.35%, and 91.07%. FedAvg, FL-MPC, FL-RAEC, and PEFL provide the lowest accuracy values for the dataset of breast cancer images (83.98%, 86.16%, 88.41%, and 90.48%, respectively). For the brain tumor MRI image dataset, FedAvg, FL-MPC, FL-RAEC, and PEFL provide the lowest accuracy values (82.79%, 84.88%, 86.27%, and 88.09%, respectively).

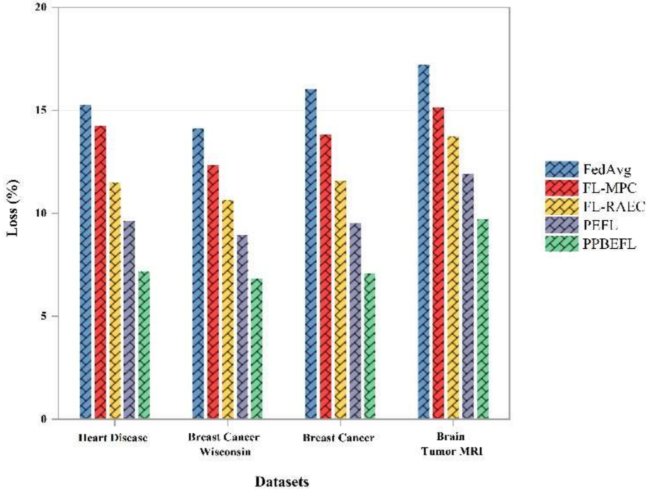


Figure 9: Dataset Loss Comparison With FL Methods

The suggested approach produces the lowest loss results for heart disease, breast cancer Wisconsin, breast cancer image dataset, and brain tumor MRI dataset, respectively, with 7.19%, 6.84%, 7.07%, and 9.72%, as shown in Figure 9. Comparing it to other techniques, it may also provide better results for other assaults. For the heart disease dataset, FedAvg, FL-MPC, FL-RAEC, and PEFL provide the lowest loss values, at 15.27%, 14.25%, 11.49%, and 9.63%, respectively. At 14.12%, 12.34%, 10.65%, and 8.93%, respectively, FedAvg, FL-MPC, FL-RAEC, and PEFL provide the greatest loss results for the Wisconsin breast cancer dataset. The compilation of breast cancer images shows the greatest loss results (16.02%, 13.84%, 11.59%, and 9.52%) for FedAvg, FL-MPC, FL-RAEC, and PEFL. FedAvg, FL-MPC, FL-RAEC, and PEFL yield the highest loss results of 17.21%, 15.12%, 13.73%, and 11.91%, respectively, for the brain tumor MRI image dataset. Validating the effectiveness, robustness, and applicability of the proposed framework.

4.2 Latency And Throughput Comparison With Other Models

The latency and throughput of common federated learning (FL) schemes, including FedAvg, FL-MPC, FL-RAEC, PEFL, and the suggested PPBEFL, are compared in this section with respect to the number of rounds, which ranges from 25 to 100. Throughput measures the system's capacity to process transactions, in this context, model update submissions, per second. Achieving approximately 97 transactions per second demonstrates the blockchain-enabled framework's capability to handle a substantial volume of client updates concurrently. High throughput supports scalability, allowing the system to maintain performance as the number of federated clients increases. However, throughput must be interpreted in conjunction with latency and resource utilization, as maximizing throughput under constrained computational or network resources may compromise individual transaction processing times or increase energy consumption, which are crucial considerations for deployment in distributed healthcare networks.

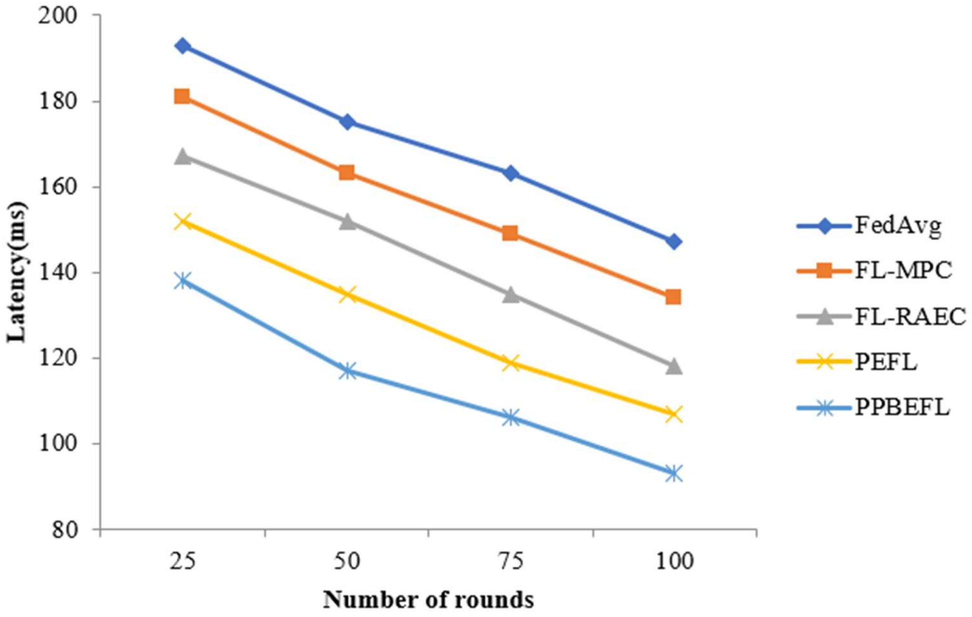


Figure 10: Fl Methods Comparison In Latency

Figure 10 compares the latency of various rounds, from 25 to 100, with a 25-round disadvantage, using FL methods. The findings demonstrate that for 25, 50, 75, and 100 rounds, respectively, the suggested system achieves the lowest latency of 138 ms, 117 ms, 106 ms, and 93 ms. For 100 rounds, the latency results for FedAvg, FL-MPC, FL-RAEC, and PEFL were 147 ms, 134 ms, 118 ms, and 107 ms, respectively. Low latency at BEFL stems from the maximum optimization of blockchain usage, ensuring fast yet reliable updates without performance delays. While the use of RAPPOR-based local differential privacy mechanisms introduces randomized response noise to obfuscate individual data contributions, the model's susceptibility to inference attacks under

varying adversarial capabilities is not quantified. A practical evaluation would involve simulating membership inference attacks across different values of the differential privacy parameter ε (e.g., ε ∈ {0.1, 1.0, 5.0}) to observe how attack accuracy changes with varying levels of privacy strength. Similarly, robustness quantification under adversarial perturbations, such as gradient-based poisoning or model replacement attacks, requires controlled experimentation where the attack strength (e.g., percentage of malicious clients, noise magnitude, or adversarial update frequency) is systematically varied. The resulting model degradation, accuracy drop, or anomaly detection precision (true positive and false negative rates) should be measured to validate the resilience of the aggregation algorithm and anomaly filter.

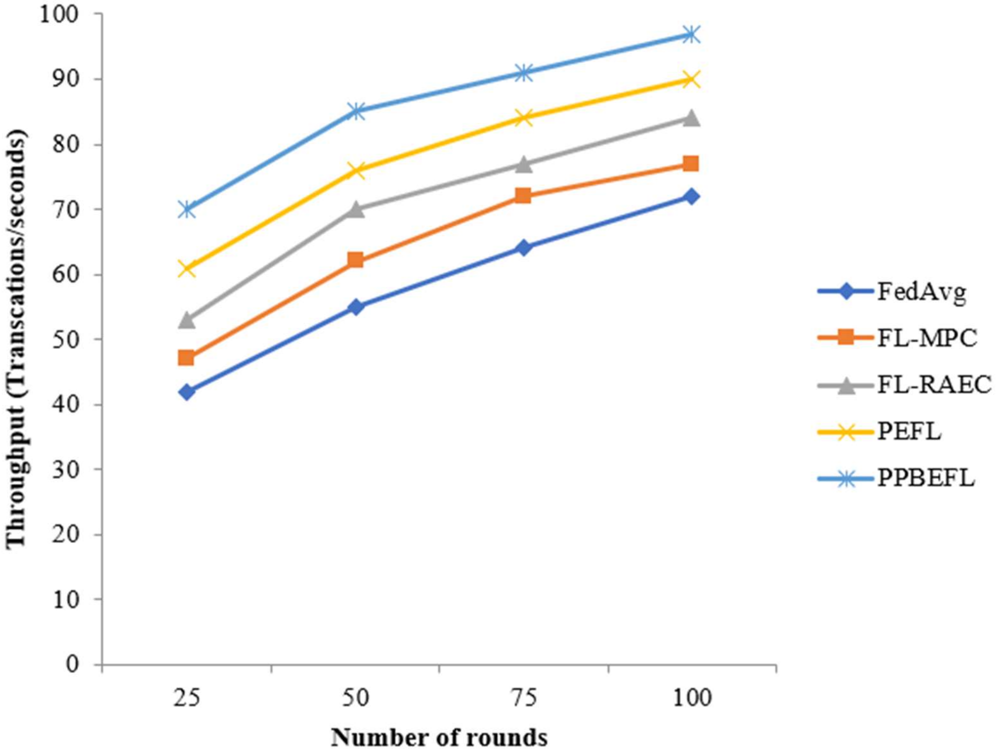


Figure 11: Throughput Comparison Of FL Methods

Figure 11, per second output with maximum throughput gained for PPBEFL, increasing from 70 to 97 transactions/sec. This testifies to the efficiency of update processes, as well as the integrity of the blockchain. FedAvg achieved a transaction rate of 72 transactions per second, which was attributed to its straightforward design; however, it attempted to provide the blockchain any dependability. The throughputs of FL-MPC and FL-RAEC are 77 and 84 transactions/sec, respectively, because of reliance on complicated blockchain and privacy-preserving protocols. PEFL at 90 transactions per second proves efficient in handling multiple clients concurrently. A better throughput of PPBEFL can be seen from the fact that it is based on the novelty concurrency mechanism, which removes latency at processing time, and it also utilizes resources efficiently, making it a suitable model for actual application usage cases.

FedAvg, FL-MPC, and FL-RAEC were selected as baselines to represent distinct and widely recognized approaches within federated learning, providing a comprehensive comparison framework. FedAvg serves as the foundational federated learning algorithm, emphasizing simple weighted averaging of local model updates without additional privacy or security enhancements. FL-MPC incorporates secure multi-party computation techniques, focusing on preserving data privacy through cryptographic means during model aggregation, which aligns with the privacy-preserving goals of PPBEFL. FL-RAEC employs robust aggregation methods designed to resist adversarial attacks and poisoning, addressing security concerns similar to those targeted by the proposed anomaly detection and filtering mechanisms.

Inferences analysis: Although the FL-integrated model with blockchain shows great promise, scalability is largely an uncharted area. Presently, most experiments have been conducted on very small-scale test beds, thereby providing limited insight into what happens when one expands the networks and increases the number of clients, as well as the complexity of the data. The significant issues are increasing the transactional load on the blockchain, managing communication between nodes to ensure efficiency, and addressing computational overhead due to aggregation. The mechanisms require much less computational power than Proof-of-Work (PoW). As environments update models so frequently, they are more efficient in such an environment. State channels and sidechains are examples of off-chain solutions. It could also be useful in reducing latency and processing requirements on the blockchain from conducting transactions off the main chain while retaining security and verifiability. The current architecture lacks performance optimization capabilities, especially in large-scale deployments with numerous clients and high data complexity, due to the lack of analysis of these alternatives. Future work can benchmark federated learning setups to assess their feasibility in various scenarios, aiming to reduce computationally expensive approaches without compromising integrity, transparency, or security. Thus, this opens a huge pathway to improve scalability and increase the applicability of the architecture. Delegated Proof-of-Stake (DPoS) reduces computational costs by involving only a few elected delegates who validate transactions, thereby making it more scalable for large networks with fast transaction times. However, in this case, some decentralization must be sacrificed. Practical Byzantine Fault Tolerance (PBFT), on the other hand, reduces the computational costs, as fewer nodes are required to agree on a transaction, making it highly efficient for smaller networks. However, this increases communication overhead with the increase in the number of nodes; therefore, the scalability is low. PBFT is more efficient for smaller federated learning systems, while DPoS is better suited for larger scalable systems.

Benchmarking PPBEFL across several dimensions, including accuracy, computational efficiency, communication overhead, and resistance to adversarial attacks, is necessary for evaluating it against state-of-the-art blockchain-integrated privacy-preserving systems and federated learning frameworks. To put current advancements into perspective, a comparative study should include models such as secure multi-party computation (MPC)-based federated learning (FL), hybrid blockchain-FL architectures, and Byzantine-resilient aggregation approaches. Experimental configurations should also account for real-world constraints, such as data distribution skew, regulatory compliance requirements, varied client device capabilities, and changing network conditions, to evaluate practical feasibility. Reliability and scalability are enlightened by stress testing under conditions of realistic latency, bandwidth constraints, and client drop-outs.

5. CONCLUSION AND FUTURE WORK

In this paper, privacy-preserving and blockchain-enabled federated learning (PPBEFL) is designed to jointly optimize efficiency, security, and data privacy in collaborative healthcare analytics. Datasets collected from Kaggle, the CNN and VGG19 architectures is developed for local model training, followed by secure aggregation through FL. PPBEFL training mechanism is described which combines blockchain system, local trainers, a server, and a consensus committee assuring transparency, traceability, and robustness throughout the learning process. Adaptive local differential privacy (LDP) strategy is incorporated for successfully balancing data utility and privacy protection. PPBEFL framework keeps track of the FL process, which is a two-layer blockchain that gives an accurate global model for learning in terms of security, privacy, and speed. PPBEFL gives a complete and scalable solution for safe and privacy-aware FL in sensitive domains. Extensive experiments are implemented on heart disease, Wisconsin breast cancer, breast cancer image, and brain tumor MRI datasets demonstrate the effectiveness of the proposed model. The accuracy values for the heart disease dataset are 84.73%, 85.75%, 88.51%, 90.37%, and 92.81% for FedAvg, FL-MPC, FL-RAEC, PEFL, and PPBEFL, respectively. Further work of the scalability of the proposed framework using larger datasets, deeper architectures for deep neural networks, and a variety of client environments can be conducted in future research. Advanced threat models remains an important direction against adversarial diversity and adaptive attacks for future research. Scalability under large-scale federated environments remains a difficult task. Secure multi-party computation (SMPC) and homomorphic encryption has highest computational complexity for future comparative analysis. Real-world applications in healthcare or finance, where the data is highly sensitive and collaborative in nature, will be further tested to demonstrate practical applicability and robustness.

REFERENCES:

[1] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated Learning of Deep Networks Using Model Averaging," *arXiv preprint arXiv:1602.05629*, Vol. 2, No. 2, 2016, pp. 1–11.

[2] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, Vol. 14, No. 1–2, 2021, pp. 1–210.

[3] Y. Liu, Y. Kang, T. Zou, Y. Pu, Y. He, X. Ye, Y. Ouyang, Y. Q. Zhang, and Q. Yang, "Vertical Federated Learning: Concepts, Advances, and Challenges," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 36, No. 7, 2024, pp. 3615–3634.

[4] A. M. Dirir, K. Salah, and D. Svetinovic, "Towards Blockchain-Based Fair and Trustworthy Federated Learning Systems," *Studies in Computational Intelligence*, Vol. 965, 2021, pp. 157–171.

[5] H. Wang, Q. Wang, and D. He, "Blockchain-Based Private Provable Data Possession," *IEEE Transactions on Dependable and Secure Computing*, Vol. 18, No. 5, 2019, pp. 2379–2389.

[6] M. Arun and G. Gopan, "Effects of Natural Light on Improving the Lighting and Energy Efficiency of Buildings: Toward Low Energy Consumption and CO$_2$ Emission," *International Journal of Low-Carbon Technologies*, Vol. 20, 2025, pp. 1047–1056.

[7] H. N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for Internet of Things: A Survey," *IEEE Internet of Things Journal*, Vol. 6, No. 5, 2019, pp. 8076–8094.

[8] H. Kim, J. Park, M. Bennis, and S. L. Kim, "Blockchained On-Device Federated Learning," *IEEE Communications Letters*, Vol. 24, No. 6, 2019, pp. 1279–1283.

[9] J. Huang, D. He, M. S. Obaidat, P. Vijayakumar, M. Luo, and K. K. R. Choo, "The Application of the Blockchain Technology in Voting Systems: A Review," *ACM Computing Surveys*, Vol. 54, No. 3, 2021, pp. 1–28.

[10] A. Qammar, A. Karim, H. Ning, and J. Ding, "Securing Federated Learning with Blockchain: A Systematic Literature Review," *Artificial Intelligence Review*, Vol. 56, No. 5, 2023, pp. 3951–3985.

[11] C. Ma, J. Li, L. Shi, M. Ding, T. Wang, Z. Han, and H. V. Poor, "When Federated Learning Meets Blockchain: A New Distributed Learning Paradigm," *IEEE Computational Intelligence Magazine*, Vol. 17, No. 3, 2022, pp. 26–33.

[12] D. C. Nguyen, M. Ding, Q. V. Pham, P. N. Pathirana, L. B. Le, A. Seneviratne, J. Li, D. Niyato, and H. V. Poor, "Federated Learning Meets Blockchain in Edge Computing: Opportunities and Challenges," *IEEE Internet of Things Journal*, Vol. 8, No. 16, 2021, pp. 12806–12825.

[13] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, "Federated Learning with Differential Privacy: Algorithms and Performance Analysis," *IEEE Transactions on Information Forensics and Security*, Vol. 15, 2020, pp. 3454–3469.

[14] X. Feng, K. Hu, and K. Han, "Blockchain-Based Asynchronous Federated Learning for Internet of Things," *IEEE Transactions on Computers*, Vol. 99, No. 1, 2021, pp. 1–9.

[15] Y. Li, H. Li, G. Xu, X. Huang, and R. Lu, "Efficient Privacy-Preserving Federated Learning with Unreliable Users," *IEEE Internet of Things Journal*, Vol. 9, No. 13, 2021, pp. 11590–11603.

[16] M. Asad, A. Moustafa, and M. Aslam, "CEEP-FL: A Comprehensive Approach for Communication Efficiency and Enhanced Privacy in Federated Learning," *Applied Soft Computing*, Vol. 104, 2021, pp. 1–12.

[17] Y. Wang, X. Zhang, J. Ma, and Q. Jin, "LDP-Fed+: A Robust and Privacy-Preserving Federated Learning Based Classification Framework Enabled by Local Differential Privacy," *Concurrency and Computation: Practice and Experience*, Vol. 35, No. 19, 2023, p. e7429.

[18] W. Liu, X. Xu, D. Li, L. Qi, F. Dai, W. Dou, and Q. Ni, "Privacy Preservation for Federated Learning with Robust Aggregation in Edge Computing," *IEEE Internet of Things Journal*, Vol. 10, No. 8, 2022, pp. 7343–7355.

[19] Y. Miao, Z. Liu, H. Li, K. K. R. Choo, and R. H. Deng, "Privacy-Preserving Byzantine-Robust Federated Learning via Blockchain Systems," *IEEE Transactions on Information Forensics and Security*, Vol. 17, 2022, pp. 2848–2861.

[20] B. Wang, Y. Chen, H. Jiang, Z. Zhao, Y. Chen, and Z. Zhao, "PPEFL: Privacy-Preserving Edge Federated Learning with Local Differential Privacy," *IEEE Internet of Things Journal*, Vol. 10, No. 17, 2023, pp. 15488–15500.

[21] I. Ullah, X. Deng, X. Pei, P. Jiang, and H. Mushtaq, "A Verifiable and Privacy-Preserving Blockchain-Based Federated Learning

Approach," *Peer-to-Peer Networking and Applications*, Vol. 16, No. 5, 2023, pp. 2256–2270.

[22] L. Tian, F. Lin, J. Gan, R. Jia, Z. Zheng, and M. Li, "PEFL: Privacy-Preserved and Efficient Federated Learning with Blockchain," *IEEE Internet of Things Journal*, Vol. 12, No. 3, 2024, pp. 3305–3317.

[23] M. Asad and S. Otoum, "BPPFL: A Blockchain-Based Framework for Privacy-Preserving Federated Learning," *Cluster Computing*, Vol. 28, No. 2, 2025, pp. 1–16.

[24] M. Abaoud, M. A. Almuqrin, and M. F. Khan, "Advancing Federated Learning through Novel Mechanism for Privacy Preservation in Healthcare Applications," *IEEE Access*, Vol. 11, 2023, pp. 83562–83579.

[25] S. Singh, S. Rathore, O. Alfarraj, A. Tolba, and B. Yoon, "A Framework for Privacy-Preservation of IoT Healthcare Data Using Federated Learning and Blockchain Technology," *Future Generation Computer Systems*, Vol. 129, 2022, pp. 380–388.

[26] Y. Qu, M. P. Uddin, C. Gan, Y. Xiang, L. Gao, and J. Yearwood, "Blockchain-Enabled Federated Learning: A Survey," *ACM Computing Surveys*, Vol. 55, No. 4, 2022, pp. 1–35.

[27] H. Cheng, Q. Youyang, W. Liu, G. Longxiang, and Z. Tianqing, "Decentralized Federated Learning for Private Smart Healthcare: A Survey," *Mathematics*, Vol. 13, No. 8, 2025, p. 1296.

[28] D. Wang, W. Liu, L. Gao, Y. N. Qu, H. Zhang, and J. Shi, "Modal-Centric Insights Into Multimodal Federated Learning for Smart Healthcare: A Survey," *Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*, 2024, pp. 145–160.

[29] M. Asad, A. Moustafa, and T. Ito, "FedOpt: Towards Communication Efficiency and Privacy Preservation in Federated Learning," *Applied Sciences*, Vol. 10, No. 8, 2020, pp. 1–11.

[30] J. Park and H. Lim, "Privacy-Preserving Federated Learning Using Homomorphic Encryption," *Applied Sciences*, Vol. 12, No. 2, 2022, pp. 1–17.

[31] P. Sanda, D. Pawar, and V. Radha, "Blockchain-Based Tamper-Proof and Transparent Investigation Model for Cloud VMs," *Journal of Supercomputing*, Vol. 78, No. 16, 2022, pp. 17891–17919.

[32] L. Kong, X. Y. Liu, H. Sheng, P. Zeng, and G. Chen, "Federated Tensor Mining for Secure Industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 3, 2020, pp. 2144–2153.

[33] H. Liang, Y. Zhang, and H. Xiong, "A Blockchain-Based Model Sharing and Calculation Method for Urban Rail Intelligent Driving Systems," *Proc. IEEE Int. Conf. on Intelligent Transportation Systems (ITSC)*, Rhodes, Greece, 2020, pp. 1–5.

[34] M. Shayan, C. Fung, C. J. M. Yoon, and I. Beschastnikh, "Biscotti: A Blockchain System for Private and Secure Federated Learning," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 32, No. 7, 2021, pp. 1513–1525.

[35] M. Fang, X. Cao, J. Jia, and N. Gong, "Local Model Poisoning Attacks to Byzantine-Robust Federated Learning," *Proc. 29th USENIX Security Symposium (USENIX Security 20)*, 2020, pp. 1605–1622.

[36] V. Shejwalkar and A. Houmansadr, "Manipulating the Byzantine: Optimizing Model Poisoning Attacks and Defenses for Federated Learning," *Proc. Network and Distributed System Security Symposium (NDSS)*, 2021, pp. 1–19.