Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.082310

ARTICLE

Privacy-Preserving Federated Learning for EEG-Based Biometric Recognition in AI-Enabled Epilepsy Detection

Qiuhao Xu[1,2], Chen Wang[1,3,*], Xi Wen[1], Lurong Jiang[1], Wenying Zheng[4,*] and Zhengkui Chen[1]

[1]School of Information Science and Engineering (School of Cyber Science and Technology), The Zhejiang Provincial Key Laboratory of Digital Fashion and Data Governance, and The Zhejiang Provincial International Cooperation Base for Science and Technology on Cloud Computing Security and Data Aggregation, Zhejiang Sci-Tech University, Hangzhou, China
[2]Faculty of Data Science, City University of Macau, Macau, China
[3]Zhejiang Provincial Innovation Center of Advanced Textile Technology, Shaoxing, China
[4]School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou, China

*Corresponding Authors: Chen Wang. Email: wangchen@zstu.edu.cn; Wenying Zheng. Email: zhengwy0501@126.com

Received: 13 March 2026; Accepted: 15 May 2026; Published: 23 July 2026

ABSTRACT: The convergence of Generative Artificial Intelligence and biometric recognition is reshaping modern healthcare. It enables more adaptive and intelligent human–machine interactions. Epilepsy, a common neurological disorder affecting millions worldwide, relies heavily on electroencephalography (EEG) signals for diagnosis and monitoring. Wearable consumer devices with EEG sensors support continuous physiological data collection. However, transmitting sensitive biometric data to centralized servers introduces serious privacy and security risks. Federated learning (FL) provides a distributed training framework that keeps raw data on local devices. Despite this advantage, existing FL methods remain vulnerable to gradient leakage attacks, where adversaries may infer private biometric information from shared model updates. To address this issue, we propose PFED, a privacy-preserving federated learning protocol that combines randomized group interaction with volunteer-assisted secure aggregation. The proposed method effectively obfuscates gradient information while maintaining model utility. Experiments on the public CHB-MIT EEG dataset show that PFED achieves reliable epilepsy detection performance and strong privacy protection. These results demonstrate its potential for secure AI-enabled biometric healthcare applications in consumer device environments.

KEYWORDS: Federated learning; secure aggregation; privacy-preserving; epilepsy detection; biometric recognition

1 Introduction

The convergence of Internet of Things (IoT) technologies, Artificial Intelligence (AI), and biometric recognition is reshaping intelligent healthcare systems. This paradigm not only improves diagnostic accuracy but also raises critical concerns regarding data privacy and ethical usage [1]. Biometric recognition, traditionally used for identity authentication, is increasingly being applied to health monitoring and disease diagnosis, where EEG signals serve as a distinctive form of physiological data.

Epilepsy is a neurological disorder characterized by sudden seizures, profoundly affecting the quality of life for millions of patients worldwide [2]. The complexity and diversity of epileptic seizure patterns pose significant challenges to the diagnosis and treatment of this condition. EEG, a non-invasive biological signal monitoring technology, captures subtle changes in brain electrical activity and plays a crucial role in

detecting epileptic seizures [3,4]. As a biometric modality, EEG signals exhibit unique characteristics across individuals, enabling not only seizure detection but also potential applications in patient identification.

Modern EEG monitoring systems enable continuous and non-invasive data collection, facilitating automated seizure detection and improving patient self-management. However, these systems introduce significant privacy concerns, as EEG data contains highly sensitive biometric information. Traditional approaches often rely on centralized data collection from multiple institutions to train high-performance models [5–7], which increases the risk of data leakage and limits data sharing across institutions.

In this context, FL, a distributed machine learning framework, has garnered significant attention in recent years [8]. FL eliminates the need to transmit data to a central server, instead allowing for local model training at individual clients, with collaboration occurring through the sharing of model updates rather than raw data. This innovative approach effectively mitigates the risks associated with data privacy breaches, positioning FL as a critical technology in privacy protection for healthcare applications [9–11]. When applied to EEG-based biometric recognition, FL enables collaborative learning across medical institutions without exposing individual patients' neural data. In addition to privacy preservation, FL reduces communication and storage overhead while leveraging distributed data to improve model generalization and recognition performance.

However, despite these advantages, existing FL-based methods in EEG-based applications primarily focus on improving model performance, while insufficiently addressing security risks associated with gradient information leakage [12,13]. Adversaries can infer sensitive information from model updates [14] or reconstruct input data through model inversion attacks [15]. Such vulnerabilities pose serious threats to patient privacy and hinder the practical deployment of FL in medical systems. As attack methods continue to evolve [16,17], protecting gradient information becomes a critical challenge in privacy-sensitive healthcare applications [18–20]. To tackle these challenges, we propose an innovative protocol that enables healthcare institutions and volunteer nodes to generate masks through randomized group interactions, thereby ensuring the privacy of gradient information in FL for EEG-based biometric recognition. The proposed protocol achieves a good balance between model performance and privacy protection, effectively enhancing the system's security and trustworthiness.

The core contributions of our work are as follows:

- **A privacy-preserving secure aggregation protocol for epilepsy detection in healthcare is proposed.** We design a randomized group interaction strategy involving clients and volunteer nodes to collaboratively generate masks, enabling secure aggregation of model updates while reducing communication and computational overhead.
- **The proposed PFED protocol achieves a balance between privacy protection and model performance.** The protocol effectively protects gradient information and sensitive EEG-based biometric data while maintaining high model accuracy in FL, meeting the requirements for security and accuracy in the healthcare domain.
- **The effectiveness of the proposed protocol is validated on real-world EEG data.** Experimental results on the CHB-MIT dataset demonstrate that PFED achieves robust epilepsy detection performance with strong privacy guarantees and minimal accuracy loss.

2 Related Work

Chaddad et al. [21] reviewed the research progress of FL in the healthcare domain, analyzed its advantages over traditional centralized models in terms of data privacy preservation, and summarized its applications in areas such as medical image analysis and remote health monitoring. Meanwhile, they pointed

out that FL still faced several challenges in healthcare scenarios, including potential security threats such as backdoor attacks, Byzantine attacks, and local poisoning attacks, which limited its practical deployment and reliability. Amin et al. [22] investigated the role of FL in the Healthcare 5.0 paradigm, highlighting its ability to enable privacy-preserving and personalized healthcare in Internet of Medical Things (IoMT) environments. They also evaluated key aspects such as resource awareness, security, scalability, and robustness, while identifying challenges including limited device resources, data heterogeneity, and model security risks. Li et al. [23] proposed a federated framework FedHMIR for device-free identity recognition that balances local personalization and global generalization through human-machine cooperative online updates and confidence-index-based aggregation. Given its focus on identity recognition using acoustic and Wi-Fi signals, this framework holds potential relevance for healthcare applications where patient identification is required across distributed sensing devices. Hou et al. [24] proposed FAGD, a federated diffusion-based framework for secure SAR image generation that addresses non-i.i.d. challenges through knowledge distillation, offering a relevant privacy-preserving paradigm for generative tasks in distributed sensing environments.

In the medical domain, various encryption techniques currently exist for FL systems to protect sensitive patient data [25], encompassing differential privacy and homomorphic encryption, among others. Otoum and Nayak [26] integrated differential privacy with FL for heart disease data analysis, introducing controlled noise to guarantee statistical privacy while leveraging FL to enable collaborative training across decentralized datasets without sharing raw data, achieving 85% test accuracy on heart disease prediction. Kumbhar and Rao [27] proposed a multi-key homomorphic encryption framework for healthcare data, combining a hybrid optimization algorithm with a deep residual network for privacy-preserving data classification in FL environments, demonstrating effective protection of sensitive medical data. Although these cryptographic approaches provide strong privacy guarantees, they may introduce significant computational overhead or require a trade-off between privacy and model accuracy.

Traditional epilepsy detection methods rely on centralized data collection and processing, which exposes patients' sensitive information. To address the issue, there is current research employing FL techniques to protect patient privacy while improving the accuracy of seizure detection. Baghersalimi et al. [28] proposed a personalized real-time FL framework for seizure detection on mobile platforms. This framework protected patient privacy and reduced communication energy consumption by training deep neural networks (DNN) [29] on local patient data and collaborating with other patient devices. The framework achieved a commendable detection performance and energy efficiency in ECG signals based on the EPILEPSIAE database [30]. In 2023, the same group of authors introduced a decentralized FL framework [13], addressing the challenge of non-independent and identically distributed (Non-IID) data distribution through adaptive ensemble learning and achieving the deployment of high-precision, low-power models on resource-constrained wearable devices via knowledge distillation. Ding et al. [31] adopted a lightweight spatiotemporal transformation network to effectively learn seizure patterns from EEG signals of patients in different geographical areas. They also proposed a greedy algorithm to select ideal fog nodes as global aggregation nodes, reducing reliance on central servers. To tackle data privacy issues in resource-constrained mobile health and wearable technologies, Aminifar et al. in [32] combined FL with secure multi-party computation (SMC) [33] techniques for model training, experimenting based on the CHB-MIT dataset [34] and intelligent mobile health applications, demonstrating the effectiveness of this framework in the medical field.

The aforementioned frontier research demonstrates that FL holds considerable potential in epilepsy detection. However, the focus of these studies has primarily been on enhancing the model's representation

capabilities and developing decentralized architectures. Nonetheless, the risk of gradient leakage within FL remains a concern, as attackers may exploit gradient information to reconstruct patients' private data.

3 Preliminaries

3.1 Key Agreement

The key agreement protocol proposed in this paper is based on the Diffie-Hellman key exchange mechanism, providing a secure and efficient method for key agreement between the client and volunteers.

3.1.1 Private Key Generation

A predefined large prime number p and a generator g are used. Each party independently selects a private key, c^{sk} for the client and v^{sk} for the volunteer, both randomly chosen and less than p.

3.1.2 Public Key Generation

Based on their respective private keys, each party computes their public key.

- The client computes $c^{pk} = g^{c^{sk}} \mod p$.
- The volunteer computes $v^{pk} = g^{v^{sk}} \mod p$.

3.1.3 Select Volenteers

All participants broadcast their public keys. The client randomly selects a number of volunteers, with $1 \leq |num_v| \leq |V|$, to prepare for shared key generation.

3.1.4 Shared Key Generation

Once both parties have received each other's public key, they compute the shared key $k_{c,v}$ as:

$$k_{c,v} = \left(c^{pk}\right)^{v^{sk}} \mod p = \left(v^{pk}\right)^{c^{sk}} \mod p. \tag{1}$$

3.2 Pseudo-Random Generator

In this protocol, a pseudo-random generator (**PRG**) [35] is used to generate random values required for cryptographic operations. The **PRG** produces a sequence of numbers that approximates the properties of random numbers but is deterministically generated from a small initial value known as the seed.

3.3 Federated Averaging

Federated Averaging (FedAvg) is a classic and widely used model aggregation algorithm in FL, proposed by McMahan et al. in 2017 [36]. FedAvg enables collaborative modeling of distributed data through an iterative process of local training on the clients and global model parameter aggregation. The algorithm allows clients to jointly train a model while maintaining privacy, ensuring that clients do not share their raw data with other clients or the server. The core idea of FedAvg is that in each communication round, clients independently perform several optimization steps based on their local data, and then send the resulting local model parameters to a central server. The server aggregates these local parameters using a weighted average based on the number of samples on each client, generating new global model parameters, which are then broadcast to all clients. FedAvg exhibits robust convergence properties even in Non-IID data scenarios. We refer to a more generalized version proposed by Reddi et al. [37] (FedOpt), which differs from FedAvg in that FedOpt allows for the selection of customizable optimizers on both the client and server sides, thereby

facilitating a more flexible and efficient approach to data processing. In this work, we combine the client-side flexibility of FedOpt with simple server-side aggregation. As shown in Algorithm 1, we use the average gradients from clients to directly update the global model parameters, rather than employing a customized server-side optimizer.

Algorithm 1: FED aggregation algorithm

Input: Initial model weights w_0, CLIENTOPT
for $t = 0, \ldots, T-1$ **do**
  $w_c^t = w_t$
  for each client $c \in \mathcal{C}$ **in parallel do**
    for $b \in \mathcal{B}_c$ **do**
      $g_c^t = \nabla f_c(w_c^t; b)$
      $w_c^t = \text{CLIENTOPT}(w_c^t, g_c^t, \eta_l, t)$
    end for
    $\Delta_c^t = w_c^t - w_t$
  end for
  $\Delta_t = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \Delta_c^t$
  $w_{t+1} = w_t + \Delta_t$
end for
Output: Final global model weights w^T

4 System Archiatecture

To better understand the protocol, we outline the system model and threat model in this section.

4.1 System Model

As shown in Fig. 1, our system model consists of three entities: hospitals, volunteer nodes, and the server.

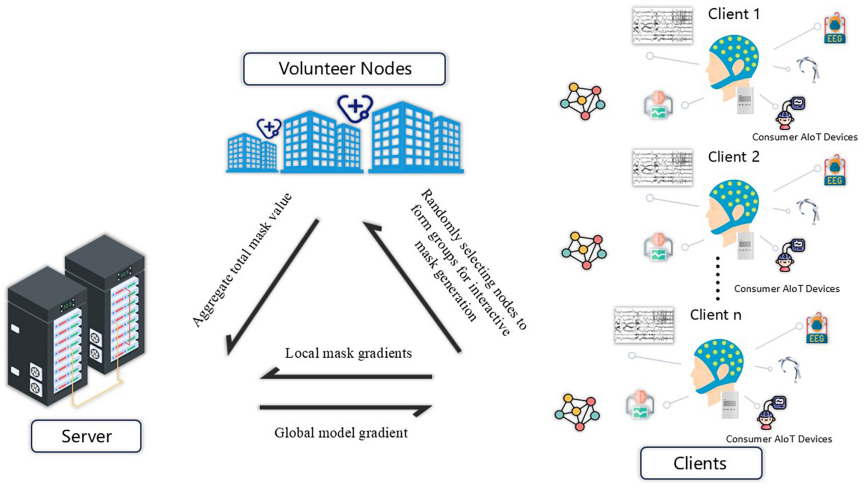


Figure 1: PFED system model for secure federated learning in AI-driven epilepsy detection.

- **Hospitals:** In clinical research, a single hospital often cannot collect sufficient EEG data from patients. Therefore, data is typically gathered and analyzed across multiple institutions in a multicenter manner. Each hospital acts as a client $c \in \mathcal{C}$, using edge IoT medical devices to collect local EEG data for model

training. To improve the global model's representational capacity, hospitals collaborate in FL with other institutions.

- **Volunteer Nodes:** The set of volunteer nodes $\mathcal{V}$ can represent hospitals or other medical entities that conduct research but do not collude with clients or participate in the actual model training. Their sole role is to assist in privacy protection.
- **Server:** The cloud server consolidates the masked gradients transmitted by all active clients and the aggregated sums of random numbers provided by the volunteer nodes.

In each training round, every client first performs local training and selects a subset of volunteer nodes. Using Diffie-Hellman key exchange, the client and each selected volunteer establish a shared secret, which serves as the seed of a **PRG** to generate identical random masks. The client adds the sum of these masks to its local model update before uploading it to the server, while each volunteer uploads the aggregated mask sum corresponding to the clients that selected it. After receiving all masked updates and aggregated masks, the server removes the masks to recover the correct global model update without accessing any client's raw gradient.

4.2 Threat Model

All entities correctly execute the prescribed protocol but may attempt to infer sensitive information from observed messages. Our threat model is defined as follows:

- All clients follow the protocol correctly but may attempt to infer private information from exchanged messages.
- The server is honest-but-curious, correctly performing aggregation while attempting to recover local gradients from masked updates.
- Volunteer nodes are honest-but-curious, correctly generating masks and uploading R_v, but may attempt to infer client information from communication data.

5 Proposed Scheme

In this section, we provide a detailed explanation of the key steps of the protocol, including the client's masked gradients, the server's process of removing the masks and aggregation, as well as handling client dropout.

5.1 Mask Generating

In the current training round t, each client trains on local EEG data to obtain local model weights w_c^t via CLIENTOPT, and computes the private gradient $\triangle_c^t = w_c^t - w_t$, which represents the update relative to the global model. To protect this gradient, each device randomly selects a number of volunteer nodes, denoted as $|num_v|$, and each selected node $v \in num_v \in \mathcal{V}$ establishes a shared key $k_{c,v}$ with the device using Eq. (1). These shared keys are used to generate pseudo-random masks via $\mathbf{PRG}(k_{c,v})$, and the masked gradient uploaded by the device is computed as:

$$\hat{\triangle}_c^t = \triangle_c^t + \sum_{v=1}^{num_v} \mathbf{PRG}(k_{c,v}). \tag{2}$$

The masking generating is only applied during the training phase for secure aggregation of model updates, ensuring that no plaintext local gradient is exposed during transmission.

5.2 Unmasking Input and Aggregation

For each selected volunteer node, it may interact with multiple clients. Suppose the set of clients that choose this volunteer node is denoted as num_c. Then, the volunteer node calculates the sum of its random numbers Rv using the following formula:

$$R_v = \sum_{c=1}^{num_c} \mathbf{PRG}(k_{c,v}). \tag{3}$$

Clients are unaware of others' selected volunteers, and volunteers do not collude or engage in training, the server can safely aggregate the uploaded gradients. Once it collects all $\hat{\triangle}_c^t$ and R_v, it removes the masks and computes the true global gradient:

$$\triangle_t = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \triangle_c^t = \frac{1}{|\mathcal{C}|} \left(\sum_{c \in \mathcal{C}} \hat{\triangle}_c^t - \sum_{v \in \mathcal{V}} R_v \right). \tag{4}$$

After computing the $\triangle_t$, the server can use the optimizer SERVEROPT to obtain the global model weights required for the next iteration and broadcast them to all online clients. Upon completion of T rounds of global collaborative training, the server will obtain the final global model weights w^T.

5.3 Dealing with Dropouts

Due to network instability or other issues, clients may drop out during the protocol. If, at a specific point in time, the server does not receive the masked gradient from a particular client, the server considers that client to have exited the protocol. Volunteer nodes are only required to compute the sum of the masks interacted with the online clients, thereby mitigating the impact of client dropout. The protocol ensures forward security and protects the privacy of the dropped-out clients' data, as their data remains masked.

6 Correctness and Security Analysis

We first describe the correctness of the server's aggregation results and then analyze the security of the protocol from an overall perspective in this section.

6.1 Correctness Analysis

Since the proposed masking strategy is linear and reversible, it does not alter the mathematical equivalence between the aggregated masked gradients and the true gradients. Specifically, as shown in Eq. (4), the server recovers the true global gradient by subtracting the total volunteer mask sum from the total client masked gradient sum. This operation is exactly identical to standard FedAvg aggregation, as the added masks cancel out completely. Although each client selects different volunteer nodes and each volunteer node interacts with different clients, the sum of the mask values from all clients is equal to the sum of the random values from all volunteer nodes (i.e., $\sum_c^{\mathcal{C}} M_c = \sum_v^{\mathcal{V}} R_v$, where M_c is the mask value of client c). The equation holds due to the properties of the **PRG**, where random values generated from the same seed are consistent (i.e., $\mathbf{PRG}(k_{c,v}) = \mathbf{PRG}(k_{v,c})$). Therefore, we can infer that as long as the total seed generated by the volunteer nodes matches that of the clients, the server can correctly eliminate the impact of all mask values during aggregation.

6.2 Security Analysis

In our protocol, the clients and volunteer nodes exchange their respective public keys c^{pk} and v^{pk}, and then independently compute the shared secret key $k_{c,v}$. Specifically, given a cyclic group G with generator

g, and two randomly chosen private exponents a and b, each party first generates its public key: the client computes $c^{pk} = g^{c^{sk}} \mod p$, and the volunteer computes $v^{pk} = g^{v^{sk}} \mod p$. Afterward, both parties exchange these public keys over a public channel and use the other party's public key along with their own private key to compute the same shared secret key $k_{c,v} = (c^{pk})^{v^{sk}} \mod p = (v^{pk})^{c^{sk}} \mod p$.

The security of this protocol reduces to the Computational Diffie-Hellman (CDH) problem. Specifically, any efficient algorithm that breaks the protocol and computes $k_{c,v}$ could also solve the CDH problem, i.e., compute $g^{c^{sk}v^{sk}}$ from g, $g^{c^{sk}}$, and $g^{v^{sk}}$. Therefore, as long as the CDH problem remains hard, our protocol provides strong security guarantees, ensuring that an attacker cannot easily compute the shared secret key, thereby protecting the security of the communication.

Furthermore, each session key (e.g., the shared key $k_{c,v}$) depends solely on the public and private key pairs used in the current session, rather than on long-term keys. Once the session ends, even if one party's private key (such as c^{sk} or v^{sk}) is compromised, an attacker cannot recover any previously established session keys from past session data, further enhancing the system's security.

7 Experimental Settings and Results

This section primarily evaluates the performance of our proposed FL secure aggregation scheme, PFED, on a standard dataset, and thoroughly demonstrates the protocol's applicability, efficiency, and security on epilepsy medical data.

7.1 Experimental Data

The MNIST dataset is a collection of handwritten digit images, containing 60,000 training images and 10,000 test images, each of size 28 × 28 pixels and in grayscale. The images represent handwritten digits from 0 to 9 and are used to evaluate the performance of image classification models.

The CIFAR-10 dataset consists of 60,000 color images of size 32 × 32 pixels, divided into 10 categories (such as airplane, automobile, cat, dog, etc.). Of these, 50,000 images are used for training, and 10,000 images are used for testing.

The epilepsy data used in this study comes from the CHB-MIT [34] Scalp EEG Database collected by Boston Children's Hospital. It includes records from 23 cases, following the International 10–20 Bipolar System for EEG electrode placement and naming conventions. This study selected 18 channels for the experiments: FP1-F7, F7-T3, T3-T5, T5-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T4, T4-T6, T6-O2, Fz-Cz, and Cz-Pz.

- **Data Preprocessing:** In the data preprocessing stage, to remove noise and retain effective signals, a band-pass filter was applied to the EEG data with a frequency range set between 1–100 Hz. Additionally, to suppress power-line interference, a notch filter targeting 60 Hz was employed. Furthermore, to effectively extract features from epileptic signals, the data were segmented using a sliding window of 4 s, focusing on ictal and interictal signals.
- **Discrete Wavelet Transform (DWT):** DWT is widely used in the processing of epileptic data due to its ability to decompose EEG signals into subbands with different frequency and time resolutions [38], thereby revealing subtle changes in epileptic brain activity. In this study, we utilized the 5th-order Daubechies wavelet function (`db5`) to perform a 4-level DWT decomposition of the EEG signals, resulting in subbands as shown in Fig. 2.
- **Feature Extraction:** In the context of epileptic data processing, selecting the energy of the signal as a feature is an effective approach. Energy features reflect the intensity distribution of the signal across different frequency components, thus revealing dynamic changes and abnormal patterns in the signal.

After DWT decomposition, detail coefficients were selected, and their energies were calculated. The energy calculation method involves summing the squares of the subband signals, as given by the following formula:

$$E_i = \sum_{j=1}^{N} \left| x_{i,j} \right|^2, \tag{5}$$

where E_i represents the energy of the i-th subband, $x_{i,j}$ denotes the j-th sample in the i-th subband, and N is the total number of samples in the i-th subband.

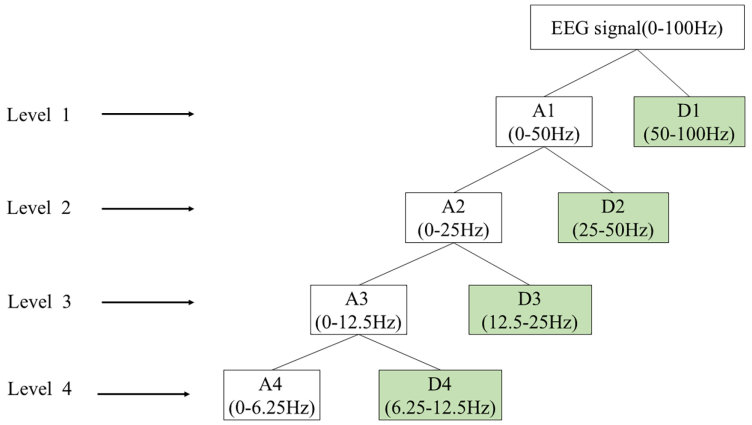


Figure 2: Four-level DWT decomposition of EEG signals using db5 wavelet.

7.2 Experimental Setup

The experimental environment was established on a desktop computer equipped with a 2.5 GHz Intel i5-12400F CPU and 63.8 GB of RAM, using the PyTorch framework for implementation. Clients utilized the Adam optimizer with a learning rate $\eta_l = 0.001$ to update network parameters.

For the MNIST dataset, a typical Multi-Layer Perceptron (MLP) model was used. The 28 × 28 pixel handwritten digit images are flattened into a 784-dimensional vector and processed through two fully connected layers, each containing 256 neurons, with a ReLU activation function applied after each layer. Finally, the network outputs the probability of each digit class through an output layer containing 10 neurons, and classification is performed using the softmax function.

For the CIFAR-10 dataset, a Convolutional Neural Network (CNN) was used. The input images pass through three convolutional layers, each followed by a ReLU activation function and a max-pooling layer to extract features and gradually reduce spatial dimensions. The output of the convolutional layers is flattened and then passed through two fully connected layers for classification. Finally, the output layer uses the softmax function to convert the results into a probability distribution across the 10 categories.

For the CHB-MIT dataset, we designed a one-dimensional CNN tailored for processing EEG time-series data. The architecture includes:

- **Input Layer:** Accepts tensors of shape $[B, 1, N]$, where B is the batch size and N is the number of features. Each input tensor corresponds to a batch of EEG signal segments.
- **Convolutional Layer:** Applies 32 filters with a kernel size of 3, stride of 1, and padding of 1, followed by ReLU activation ($\sigma(x) = \max(0, x)$) and max pooling with a kernel size of 2 and stride of 2. The output can be expressed as:

$$Y = \text{MaxPool}(\sigma(W * X + b)), \tag{6}$$

where W and b are learnable weights and biases, and $*$ denotes the convolution operation.

- **Flatten Layer:** Transforms the two-dimensional feature maps into a one-dimensional vector of size $[B, CHA \times L]$, where CHA is the number of channels (filters) and L is the length of the feature maps.
- **Fully Connected Layers:** Two fully connected layers follow. The first has 128 neurons with ReLU activation and dropout (rate 0.5), while the second has 2 neurons for binary classification. The transformations can be represented as:

$$\begin{aligned} H_1 &= \sigma(W_1 \cdot \text{Flatten}(Y) + b_1), \\ H_2 &= \sigma(W_2 \cdot H_1 + b_2), \end{aligned} \tag{7}$$

where H_1 and H_2 are hidden states, and W_1, W_2, b_1, and b_2 are learnable parameters.

- **Output Layer:** Uses softmax to generate probability distributions over two classes. Training uses cross-entropy loss:

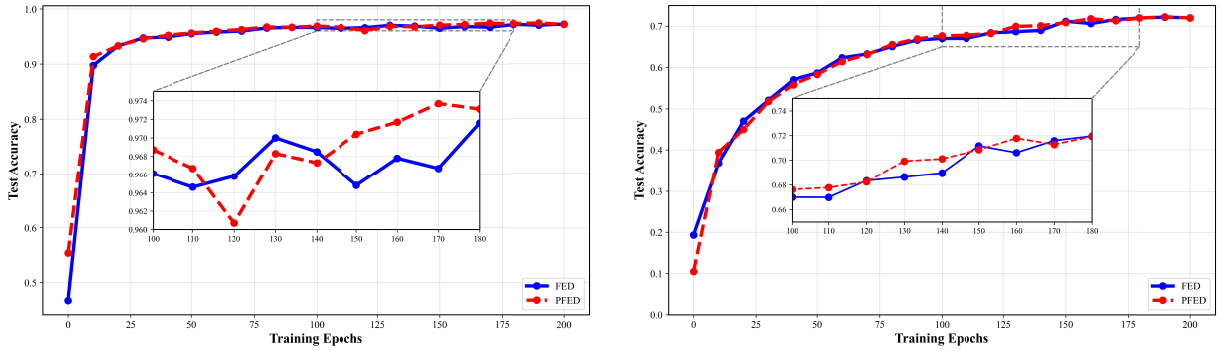
$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right], \tag{8}$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability for the i-th sample in the batch.

7.3 Experimental Results

7.3.1 Performance Evaluation

To demonstrate that our protocol ensures data privacy without significantly compromising model performance, we compare it with standard aggregation algorithms that employ no security mechanisms. The results are shown in Fig. 3. We test model accuracy on a conventional dataset and find that the gradient protection approach with masking does not affect the convergence of model accuracy, thus not sacrificing model performance. In the context of EEG-based biometric recognition for medical data, the proposed protocol still performs well, as shown in Table 1. With a single client (simulating centralized training), FL exhibits a certain degree of performance loss across multiple metrics. A vertical comparison between FED and PFED reveals that PFED closely matches FED across all evaluation metrics, demonstrating that our privacy-preserving approach effectively maintains the performance of EEG-based biometric identification and epilepsy detection. Fig. 4 illustrates the changes in model accuracy for both approaches. The gap between PFED and FED remains small in each round, indicating that privacy-preserving security can be achieved with minimal performance cost when training biometric recognition models on client devices. To investigate the impact of the number of volunteer nodes on the protocol, we conducted an ablation study, as shown in Fig. 5. The results indicate that the number of $\mathcal{V}$ does not affect model training, demonstrating that it is a controllable security parameter within the protocol that does not impact the accuracy of biometric recognition.



a MNIST. **b** CIFAR-10.

Figure 3: Comparison of accuracy between privacy preserving PFED model with 5 clients and conventional model based on the MNIST and CIFAR10 dataset. Clients were set to train locally for 1 and 10 epochs, with a batch size of 64.

Table 1: Performance comparison with different client numbers (CHB-MIT dataset). Clients were set to train the model locally for a single epoch per round with a batch size of 64.

Clients	Method	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	MCC (%)
Single	–	90.21	89.25	92.37	90.78	80.40
N = 5	FED	88.75	86.99	92.23	89.53	77.54
	PFED	87.77	86.23	91.11	88.60	75.56
N = 10	FED	86.17	83.82	91.07	87.30	72.47
	PFED	85.86	84.03	90.01	86.92	71.77

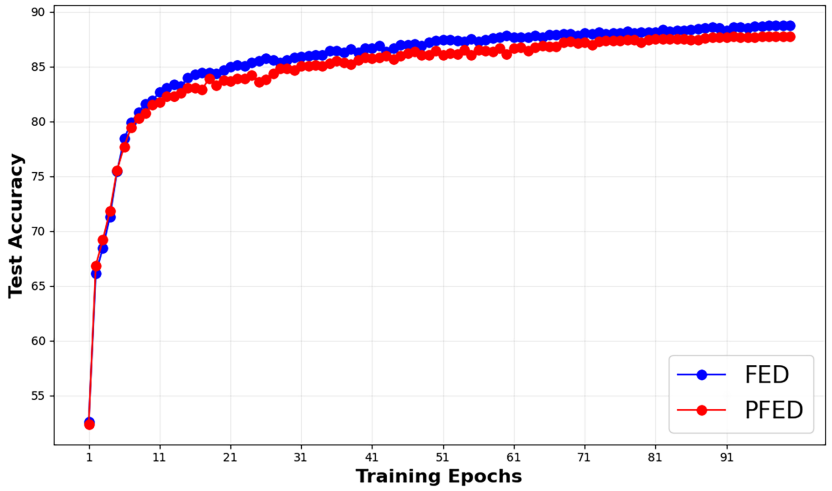


Figure 4: Comparison of accuracy between privacy preserving PFED model with 5 clients and conventional model based on the CHB-MIT dataset. Clients were set to train the model locally for a single epoch per round with a batch size of 64.

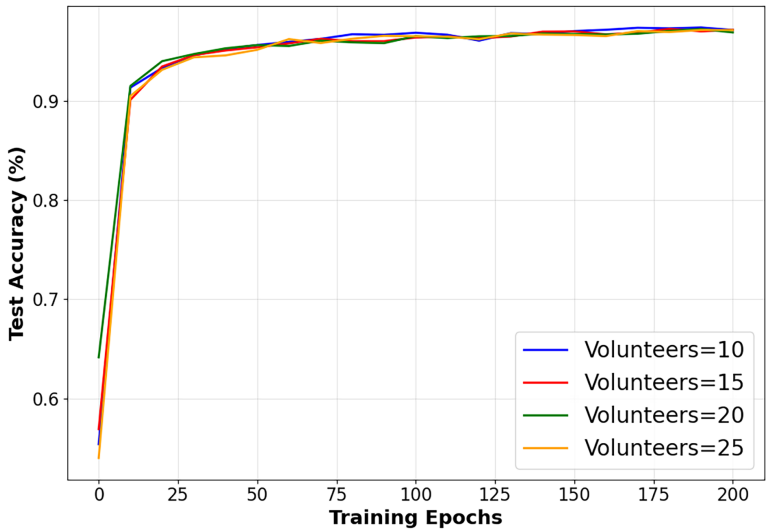


Figure 5: Comparison of model accuracy trends under different numbers of volunteer nodes based on the MNIST dataset. The total number of clients is 5, and clients were set to train the model locally for a single epoch per round with a batch size of 64.

7.3.2 Time Evaluation

As illustrated in Fig. 6, the total time for the two primary steps when executing the protocol with different numbers of clients based on medical data is presented. Each client performs one round of local training, and the number of volunteer nodes is set to 10, the reported results correspond to the time cost of the two stages within one global training round. The local model training time in PFED is identical to that of standard FED, since both methods use the same model architecture, optimizer, and number of local epochs. *Key Generate* represents the time taken for clients to interact with volunteers to generate shared keys, while *Mask Generate* denotes the time clients spend generating and aggregating mask values via the **PRG**. For **PRG**, a SHA-256-based generator is employed. The time cost for clients to protect EEG-based biometric gradients is almost negligible, with the majority of time still dedicated to training biometric recognition models. This observation indirectly underscores the efficiency of the protocol for biometric recognition applications.

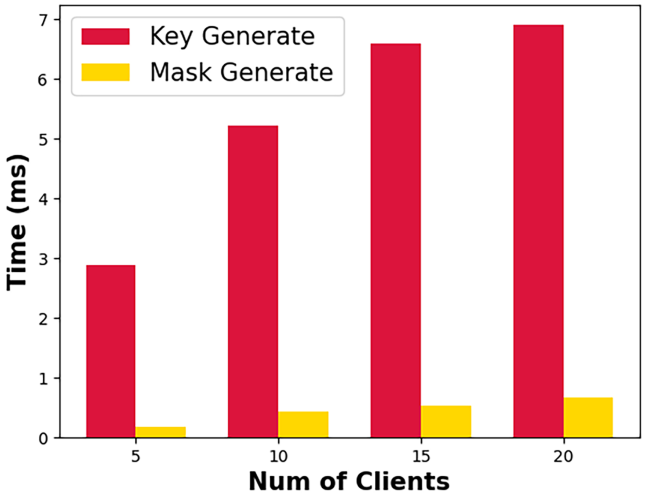


Figure 6: The time cost of different operations with different numbers of clients.

8 Conclusion

This paper proposes PFED, a privacy-preserving FL protocol for EEG-based biometric recognition in AI-driven epilepsy detection. To our knowledge, PFED is the first method to introduce a randomized multi-party interactive masking mechanism into FL. It effectively protects sensitive EEG biometric information while maintaining high recognition accuracy. PFED designs a new secure aggregation mechanism. Volunteer nodes interact through random grouping to collaboratively generate gradient masks. This provides dual protection for biometric data and gradients during federated training. Theoretical analysis and experimental results on the CHB-MIT EEG dataset show that PFED offers strong privacy protection with only a small impact on model performance. Therefore, it is suitable for EEG biometric recognition and epilepsy detection in real medical scenarios.

Several future directions are worth exploring. For example, generative AI techniques such as diffusion models or GANs can be combined with PFED to enable privacy-preserving EEG data augmentation. This may alleviate data scarcity and improve model robustness. PFED can also be extended to cross-modal physiological signal analysis. Integrating EEG with other biomedical signals like ECG or EMG can support more comprehensive patient health monitoring. Furthermore, inspired by works such as FAGD, future research can explore combining privacy-preserving FL with secure generative frameworks. This could lead to trustworthy medical data generation and multimodal fusion.

Acknowledgement: Not applicable.

Funding Statement: The work is supported by the the National Natural Science Foundation of China (Nos. 62302457, 62441228, 62402444), the Zhejiang Provincial Natural Science Foundation of China (Nos. LQ24F020008, LQ24F020012), the Program for Leading Innovativ Research Team of Zhejiang Province (No. 2023R01001), the Fundamental Research Funds of Zhejiang Sci-Tech University (No. 22222266-Y) and the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (Nos. 2025C02033, 2023C01119).

Author Contributions: All the authors made a great contribution to this work: Qiuhao Xu: writing—original draft preparation, validation, visualization. Chen Wang: supervision, writing—review and editing, resources, funding acquisition. Xi Wen: writing—original draft preparation, visualization, data curation. Lurong Jiang: visualization, writing—review and editing. Wenying Zheng: funding acquisition, supervision. Zhengkui Chen: supervision. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alaskar NM, Hussain M, Almheiri SJ, Khan A, Adnan KM. Big data-driven federated learning model for scalable and privacy-preserving cyber threat detection in IoT-enabled healthcare systems. Comput Mater Contin. 2026;87(1):074041. doi:10.32604/cmc.2025.074041.
2. Sander JW. The epidemiology of epilepsy revisited. Curr Opin Neurol. 2003;16(2):165–70.
3. Petukhov IV, Glazyrin AE, Gorokhov AV, Steshina LA, Tanryverdiev IO. Being present in a real or virtual world: a EEG study. Int J Med Inform. 2020;136(7):103977. doi:10.1016/j.ijmedinf.2019.103977.
4. Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol. 2018;17(3):279–88.

5. Avcu MT, Zhang Z, Chan DWS. Seizure detection using least EEG channels by deep convolutional neural network. In: Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019 May 12–17; Brighton, UK. p. 1120–4.
6. Vidyaratne L, Glandon A, Alam M, Iftekharuddin KM. Deep recurrent neural network for seizure detection. In: Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN); 2016 Jul 24–29; Vancouver, BC, Canada. p. 1202–7.
7. Zhang M, Chen S, Shen J, Susilo W. Privacyeafl: privacy-enhanced aggregation for federated learning in mobile crowdsensing. IEEE Trans Inf Forensics Secur. 2023;18:5804–12.
8. Mahmud F, Mahmud F, Rahman RM. A survey of federated learning: advances in architecture, synchronization, and security threats. Comput Mater Contin. 2026;86(3):073519. doi:10.32604/cmc.2025.073519.
9. Li J, Meng Y, Ma L, Du S, Zhu H, Pei Q, et al. A federated learning based privacy-preserving smart healthcare system. IEEE Trans Ind Inform. 2021;18(3):2021–31. doi:10.1109/tii.2021.3098010.
10. Lian Z, Wang W, Han Z, Su C. Blockchain-based personalized federated learning for internet of medical things. IEEE Trans Sustain Comput. 2023;8(4):694–702. doi:10.1109/tsusc.2023.3279111.
11. Wang C, Yang Q, Shen J, Wu QJ, Bm-pda He D. Blockchain based multifunctional private-preserving data aggregation for e-health systems. IEEE Trans Dependable Secur Comput. 2025;22(6):7600–13. doi:10.1109/tdsc.2025.3598867.
12. Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans Inf Forensics Secur. 2020;15:3454–69.
13. Baghersalimi S, Teijeiro T, Aminifar A, Atienza D. Decentralized federated learning for epileptic seizures detection in low-power wearable systems. IEEE Trans Mob Comput. 2023;23(5):6392–407. doi:10.1109/tmc.2023.3320862.
14. Liu Y, Ma S, Aafer Y, Lee WC, Zhai J, Wang W, et al. Trojaning attack on neural networks. In: Proceedings of the 25th Annual Network and Distributed System Security Symposium (NDSS 2018); 2018 Feb 18–21; San Diego, CA, USA.
15. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security; 2015 Oct 12–16; Denver, CO, USA. p. 1322–33.
16. Yazdinejad A, Dehghantanha A, Karimipour H, Srivastava G, Parizi RM. A robust privacy-preserving federated learning model against model poisoning attacks. IEEE Trans Inf Forensics Secur. 2024;19:6693–708. doi:10.1109/tifs.2024.3420126.
17. Han S, Buyukates B, Hu Z, Jin H, Jin W, Sun L, et al. Fedsecurity: a benchmark for attacks and defenses in federated learning and federated LLMs. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2024 Aug 25–29; Barcelona, Spain. p. 5070–81.
18. Zhang Y, Zeng D, Luo J, Xu Z, King I. A survey of trustworthy federated learning with perspectives on security, robustness and privacy. In: Proceedings of the Companion Proceedings of the ACM Web Conference 2023; 2023 Apr 30–May 4; Austin, TX, USA. p. 1167–76.
19. Zhang G, Liu B, Zhu T, Ding M, Zhou W. PPFed: a privacy-preserving and personalized federated learning framework. IEEE Internet Things J. 2024;11(11):19380–93.
20. Chang W, Zhu T. Gradient-based defense methods for data leakage in vertical federated learning. Comput Secur. 2024;139:103744. doi:10.1016/j.cose.2024.103744.
21. Chaddad A, Wu Y, Desrosiers C. Federated learning for healthcare applications. IEEE Internet Things J. 2024;11(5):7339–58. doi:10.1109/jiot.2023.3325822.
22. Amin MS, Ahmad S, Loh WK. Federated learning for Healthcare 5.0: a comprehensive survey, taxonomy, challenges, and solutions. Soft Comput. 2025;29(2):673–700.
23. Li Q, Cao Y, Wang Q, Yao L, Yu Z, Cui J. Fedhmir: unified framework for federated human-machine synergy in personalization-generalization balancing identity recognition. IEEE Trans Mob Comput. 2025;24(8):7406–22. doi:10.1109/tmc.2025.3549925.
24. Hou Y, Wang Y, Xia X, Tian Y, Li Z, Quek TQ. Toward secure sar image generation via federated angle-aware generative diffusion framework. IEEE Internet Things J. 2025;13(2):2713–30. doi:10.1109/jiot.2025.3630329.

25. Mantey EA, Zhou C, Anajemba JH, Arthur JK, Hamid Y, Chowhan A, et al. Federated learning approach for secured medical recommendation in internet of medical things using homomorphic encryption. IEEE J Biomed Health Inform. 2024;28(6):3329–40. doi:10.1109/jbhi.2024.3350232.
26. Otoum Y, Nayak A. Differential privacy-driven framework for enhancing heart disease prediction. In: Proceedings of the ICC 2025-IEEE International Conference on Communications; 2025 Jun 8–12; Montreal, QC, Canada. p. 5456–62.
27. Kumbhar HR, Rao SS. Federated learning enabled multi-key homomorphic encryption. Expert Syst Appl. 2025;268(1):126197. doi:10.1016/j.eswa.2024.126197.
28. Baghersalimi S, Teijeiro T, Atienza D, Aminifar A. Personalized real-time federated learning for epileptic seizure detection. IEEE J Biomed Health Inform. 2021;26(2):898–909. doi:10.1109/jbhi.2021.3096127.
29. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. doi:10.1038/nature14539.
30. Ihle M, Feldwisch-Drentrup H, Teixeira CA, Witon A, Schelter B, Timmer J, et al. EPILEPSIAE—a European epilepsy database. Comput Methods Programs Biomed. 2012;106(3):127–38.
31. Ding W, Abdel-Basset M, Hawash H, Abdel-Razek S, Liu C. Fed-ESD: federated learning for efficient epileptic seizure detection in the fog-assisted internet of medical things. Inf Sci. 2023;630(3):403–19. doi:10.1016/j.ins.2023.02.052.
32. Aminifar A, Shokri M, Aminifar A. Privacy-preserving edge federated learning for intelligent mobile-health systems. Future Gener Comput Syst. 2024;161(2):625–37. doi:10.1016/j.future.2024.07.035.
33. Yao AC. Protocols for secure computations. In: Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982); 1982 Nov 3–5; Chicago, IL, USA. p. 160–4.
34. Guttag J. CHB-MIT scalp EEG database. PhysioNet, Version 1.0.0. 2010 [cited 2026 Jan 1]. Available from: https://doi.org/10.13026/C2K01R.
35. Impagliazzo R, Levin LA, Luby M. Pseudo-random generation from one-way functions. In: Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing; 1989 May 14–17; Washington, DC, USA. p. 12–24.
36. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. Proc Mach Learn Res. 2017;54:1273–82.
37. Reddi S, Charles Z, Zaheer M, Garrett Z, Rush K, Konečnỳ J, et al. Adaptive federated optimization. arXiv:2003.00295. 2020.
38. Raghu S, Sriraam N, Temel Y, Rao SV, Hegde AS, Kubben PL. Performance evaluation of DWT based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using SVM classifier. Comput Biol Med. 2019;110(4):127–43. doi:10.1016/j.compbiomed.2019.05.016.