Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 6, pp. 2263–2279
DOI : 10.47974/JIOS-2323

Neuro shield : Securing neural data with privacy-preserving algorithms for brain computer interface

Kaku Riya Dharmendra [†]
Deepak Kumar Verma *
Department of Computer Engineering
Marwadi University
Rajkot 360003
Gujarat
India

Satvik Vats [§]
Department of Computer Science & Engineering
Madan Mohan Malaviya University of Technology
Gorakhpur 273010
Uttar Pradesh
India

Abstract

Signal classification via electroencephalography (EEG) is essential for smart healthcare systems and human-machine interaction systems. Centralized machine learning approaches are problematic because they require collecting sensitive neural data from multiple participants, which raises seriously privacy concerns. This paper proposes a privacy-preserving federated learning approach for classifying EEG signals in a distributed manner. The framework uses Convolutional Neural Networks (CNN) and two federated aggregation techniques: Federated Averaging (FedAvg) and Federated Proximal (FedProx) to allow the collaborative training of models without revealing individual raw EEG datasets. The approach is tested on the PhysioNet EEG Motor Imagery dataset with 109 subjects. The performance of the federated framework is comparable to the centralized framework, thus, successfully preserving privacy. More specifically, the model based on FedProx, in a specified

[†] *ORCID-ID:* **0009-0006-2662-5019**
* *ORCID-ID:* **0000-0002-4745-1752**
[§] *ORCID-ID:* **0000-0002-9422-4915**

[†] *E-mail:* **riya.kaku119353@marwadiuniversity.ac.in**
* *E-mail:* **deepak.verma1980@gmail.com (Corresponding Author)**
[§] *E-mail:* **svcse@mmmut.ac.in**

© Taru Publications

Interval of Data (IID), classification achieved 90.36%, exceeding many current federated EEG learning methods. The findings reinforce the value of federated optimization methods for confidential distributed learning over diverse EEG datasets.

Subject Classification: *68T07, 68T05.*

Keywords: *Distributed EEG learning system, EEG, Motor imagery, FL, FedAvg, FedProx, Feature selection.*

1. Introduction

With the development of data-driven applications, the protection of privacy in distributed learning systems has been a growing area of research. In many instances, the sharing of sensitive biomedical signals (e.g., EEG signals) among institutions is not possible due to privacy and ethical considerations. A solution to this problem is potential with federated learning, an approach that allows several clients to train machine learning models in a collaborative manner without having to share their data. From a computational standpoint, federated learning is characterized as a distributed optimization problem that involves secure aggregation of the model parameters and communication efficiency. EEG is the most widely used modality due to its non-invasive nature, low cost, and high temporal resolution. Distributed EEG learning systems frequently employ Motor Imagery (MI) tasks, such as imagining left or right-hand movements, to translate neural activity into control commands. Despite progress in deep learning, traditional centralized EEG learning systems face two significant challenges: (i) the scarcity of annotated datasets and high inter-subject variability, which complicates robust training and model generalization [1], and (ii) data privacy concerns, since centralizing raw EEG signals [20] risks exposing sensitive neurological information [2]. Regarding neural data, privacy-sensitive cases are beginning to view federated learning as a helpful method. Unlike traditional learning frameworks, FL enables decentralized model training across multiple EEG clients without sharing raw signals. FL also enhances generalization across subjects. In EEG-based motor imagery tasks, client heterogeneity and non-IID data distributions degrade model convergence stability and overall accuracy [2], [3]. Consequently, FL designs should target frameworks that balance the need for efficient coverage of inter-subject variability and real-world data privacy within Distributed EEG learning system applications [4], [12]. As for studies focused on federated frameworks, not many have

looked into large-scale motor imagery datasets, particularly PhysioNet, which is realistically subject to variability conditions.

1.1 *Contribution*

- The proposed framework uses a privacy-focused approach to classifying brainwave (EEG) activity for motor imagery (MI) tasks. It employs two common teamwork strategies, FedAvg and FedProx, for sharing knowledge among groups while minimizing the exposure of sensitive data.
- A federated feature optimization strategy to improve model stability for non-IID EEG datasets.
- To test this framework, we used data from 109 individuals performing tasks with their left and right hands. There were two tests: in the first, we compared learning centrally versus in groups; in the second, we varied the group size (1–5 clients) to assess how well the system performs for different group sizes.

1.2 *Novelty of the Work*

Unlike existing federated EEG studies that focus on limited subject sizes or rely solely on FedAvg aggregation, the proposed framework introduces a federated feature selection strategy combined with FedProx to minimize client drift and improve model stability for non-IID EEG data. By experimenting with 109 subjects, this work demonstrates large-scale federated MI classification, which has not been explored extensively in prior studies. Additionally, the framework achieves higher accuracy than existing FL-based Distributed EEG learning system methods while simultaneously preserving data privacy and improving scalability in decentralized environments.

2. Related Work

The researchers developed a hierarchical FL model that incorporates CNNs into meta-learning, highlighting the effectiveness of adaptive personalization as a technique to improve cross-subject generalization [4]. Similarly, [5] applied federated transformers to EEG-based motor imagery (MI) problems using the MI OpenBCI dataset. In [6], a privacy-preserving method for FL was proposed to incorporate different levels of confidentiality in FedAvg with RNNs. [7] proposed federated contrastive

learning (SimCLR + FedSimCLR) on the EEG Eye State dataset. Personalization is directly addressed and implemented by [8], using subject-specific FL with the structure SVM + FedProx. A graph-based FL approach using GNNs to infer EEG channel interactions in TUH EEG dataset records [9]. MobileNet+FedAvg was applied to the Emotiv Epoc EEG, which enabled real-time EEG classification on edge devices [10]. Similarly, a hybrid FL plus transfer learning approach by [11] on PhysioNet utilized CNNs that leveraged pre-trained features to expedite convergence. A new method based on local batch-specific BatchNorm sharpness-minimizing (SAM) is proposed for classifying EEGs in a federated setting [12]. A federated seizure diagnosis model that allows clients to train collaboratively without sharing raw EEG data was proposed in [13].

3. Federated Optimization Framework

The purpose of the research is to evaluate the effectiveness of FL in analysing EEG-based motor imagery tasks.

A. *Dataset*

This work utilized the EEG Motor Movement/Imagery Dataset [15]. It is a collection of EEG data from 109 normal individuals, recorded using the BCI2000 system on a 64-channel system (1020 system) with a sampling rate of 160 Hz in EDF+ format. We selected this dataset for our work because it is the most appropriate for assessing federated learning with realistic Distributed EEG learning systems. The dataset has well-defined motor imagery tasks, which makes it very aligned with my classification-based Distributed EEG learning system research.

B. *Data Pre-processing*

Figure 1 illustrates the FL process for EEG motor imagery classification. This process begins with the preparation of the dataset, including preprocessing, subject filtering, removal of bad channels, and Epoching of the EEG signal. Preprocessing is the first step in the dataset preparation process, where raw EEG signals are filtered, resampled, re-referenced, and normalized to eliminate noise and ensure data consistency. This is followed by subject filtering to remove subjects with incomplete, corrupted, or poor-quality recordings, which can hinder model training. Then, bad channel elimination is performed to identify and remove destructive or noisy electrodes through automated tests (e.g., flatline, high variance, low

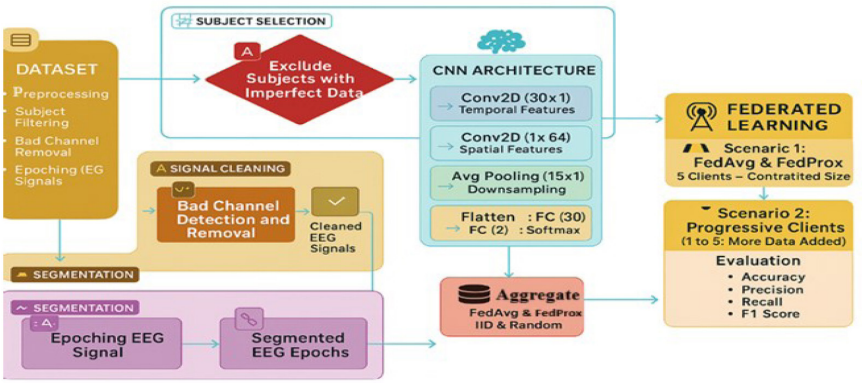

Figure 1

EEG Motor Imagery Classification using Federated CNN Framework

correlation), which may be followed by interpolating the bad channels. Lastly, epoching is performed on the cleaned continuous EEG, whereby the signals are partitioned into fixed time windows, synchronized to event markers, and subsequently baseline-adjusted and labeled to produce structured input for use in additional analysis and training models. A CNN architecture identifies both time and space characteristics using convolutional layers, downsizes the data with average pooling for down-sampling, and transfers information through fully connected layers. Two scenarios are investigated in the FL phase: (1) training on fixed clients with FedAvg and FedProx, and (2) gradually increasing the number of clients with additional data. Lastly, the performance of aggregated models is measured by accuracy, precision, recall, and F1-score.

C. *Methodology*

In traditional EEG classification, high-dimensional and noisy feature spaces often lead to increased model variance, which degrades generalization.

In the federated learning setting, the global training objective can be formulated as a distributed optimization problem. Let K denote the number of clients and n_k represent the number of samples at client k. The overall objective is to minimize the weighted sum of local empirical losses across all clients.

$$\min F(w) = \sum_{k=1}^{K} \frac{nk}{n} Fk(w)$$

Where $Fk(w)$ represents the loss function at client k, n_k is the number of data samples at client k and $n = \Sigma_{k=1}^{K} nk$ denotes the total number of samples across all clients. This formulation provides the basis for federated optimization algorithms such as FedAvg and FedProx, which iteratively aggregate locally trained model updates to obtain a global model.

The federated feature selection module in the proposed framework mitigates this by eliminating redundant and irrelevant EEG features. Furthermore, the FedProx optimization algorithm introduces a proximal regularization term given by

$$\text{Proximal regularization} = \frac{\mu}{2} \| w_k - w \|^2$$

Where w_k denotes the local client parameters and w represents the global model parameters, this term penalizes excessive divergence between local and global weights, thereby constraining client drift during non-IID updates.

D. *CNN Model & FL*

The network is designed according to the CNN architecture [16], with a few modifications. The most significant is the replacement of ReLU with Leaky ReLU to address the dying ReLU problem [18] and ensure proper gradient flow. The model utilizes temporal convolution, spatial filtering, and depth-wise convolution to extract rich temporal-spatial features directly from the raw EEG signal layer [19], eliminating the need for manual feature design. Stability and overfitting are addressed through the combination of batch normalization and dropout. FedAvg and FedProx have been considered as aggregation strategies because both works well in federated optimization [21] when clients are heterogeneous. FedAvg remains the basic algorithm used in distributed training as it simply averages the local model updates. Because of this, it is the most cited algorithm in the FL literature [14]. However, training with EEG data presents challenges because they are non-IID. This is due to inter-subject variability, which results in client drift and unstable convergence during training.

Algorithm 1: Federated Averaging (FedAvg)

1. Initialize global model parameters w_0
2. For each round t = 1, 2, ……., T
3. For each client $k \in N$ in parallel do

- Train a local model starting from *wtk* using the provided data.
- Optimize local objective function $F_k(w)$ using dataset D_k.
- Obtain updated local parameters w_t^k.
  Aggregate client updates at the server:

$$w_t = \sum_{k=1}^{N} \frac{n_k}{n} w_t^k \tag{1}$$

4. End For
5. Output: Final global model weights w_T.

At communication round t, the global model parameters are denoted by w_t. Each client k possesses n_k local data samples, while the total number of data samples across all clients is

$$N = \sum_{k=1}^{N} n_k$$

where n_k represents the number of samples at client *k,* and *N* is the total sample count across all clients.

Algorithm 2: Federated Proximal (FedProx)

1. Initialize global model parameters w0.
2. For each round t=1,2,…,T
   For each client k ∈ N in parallel:

- Solve the local optimization problem:

$$w_t^k = \arg\min(F_k(w) + \frac{\mu}{2} \| w - w_t \| 2) \tag{2}$$

  Where F_k (w) is the empirical loss on client k.
- Here, w_t^k represents the parameters of the local model for client k during round t.

Server Aggregation:

$$w_t = \sum_{k=1}^{n} \frac{n_k}{n} w_t^k \tag{3}$$

Where w_t are the updated global parameters.

3. End For
4. Output: Final global model parameters w_T

μ is the regularization coefficient; FedAvg is FedProx with μ = 0.

4. Result and Discussion

Federated learning is used to evaluate the performance of the proposed framework EEG motor imagery classification, namely, IID and random. The data from 109 subjects were divided into five groups, each with approximately 20% of the total samples. These samples were then divided into training and testing sets, with 80% allocated to the training set and 20% to the testing set, respectively. Conversely, the random condition used the same sample size for each individual. However, there was an unequal distribution of classes, with some clients having more samples of one type compared to the other. After each client updated its local model twice on a given round, the process was repeated 10 times of intercommunication. FL with centralized training using two methods of aggregation has been compared: FedAvg and FedProx [14], [17]. The global model performed the best overall in FedAvg experiments in the IID setting with four clients, with an overall accuracy of 88.10% in round 10. Nonetheless, the maximum accuracy of 89.80% was at round 9 with three clients. The IID arrangement also resulted in more rapid and uniform convergence, even with varying client numbers. Additionally, all cases were evaluated based on precision, recall, and F1 score, as shown in Figures 2 and 3.

In the IID case, the highest accuracy of 90.36 was obtained with four clients using FedProx (μ = 0.3), surpassing all other paradigms and methods. In the random environment, the greatest accuracy of 89.57% was achieved with five clients using FedProx, whereas FedAvg performed worse in all assessment measures in the same environment. Figures 4 and 5 demonstrate the trends of precision, recall, and F1 score.

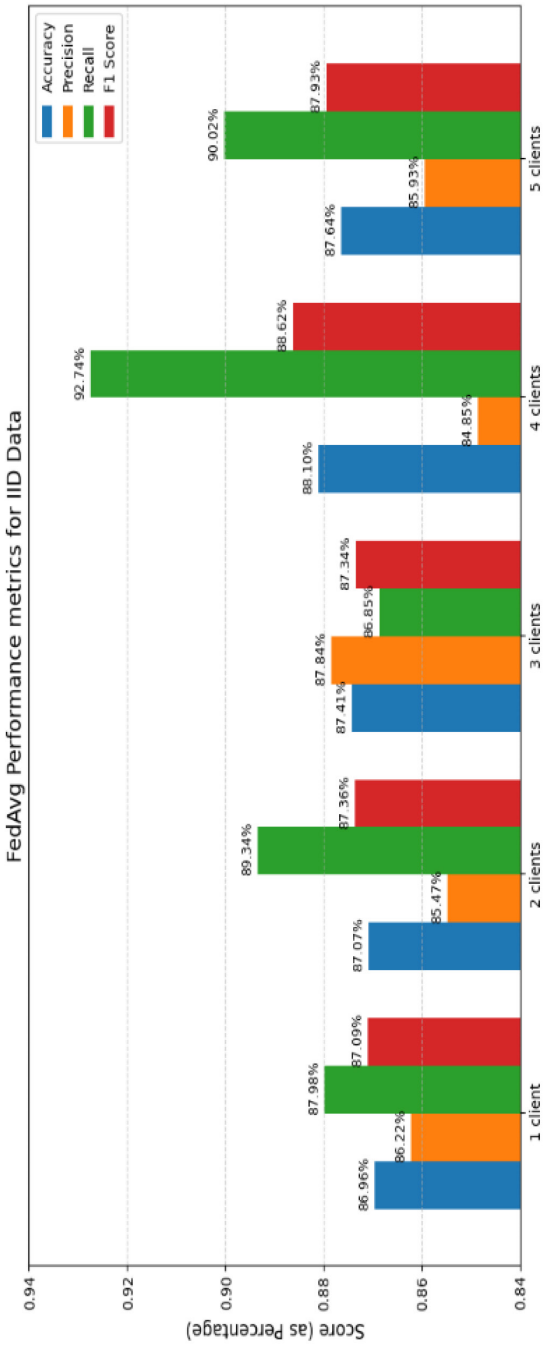

Figure 2

FedAvg Performance metrics for IID data

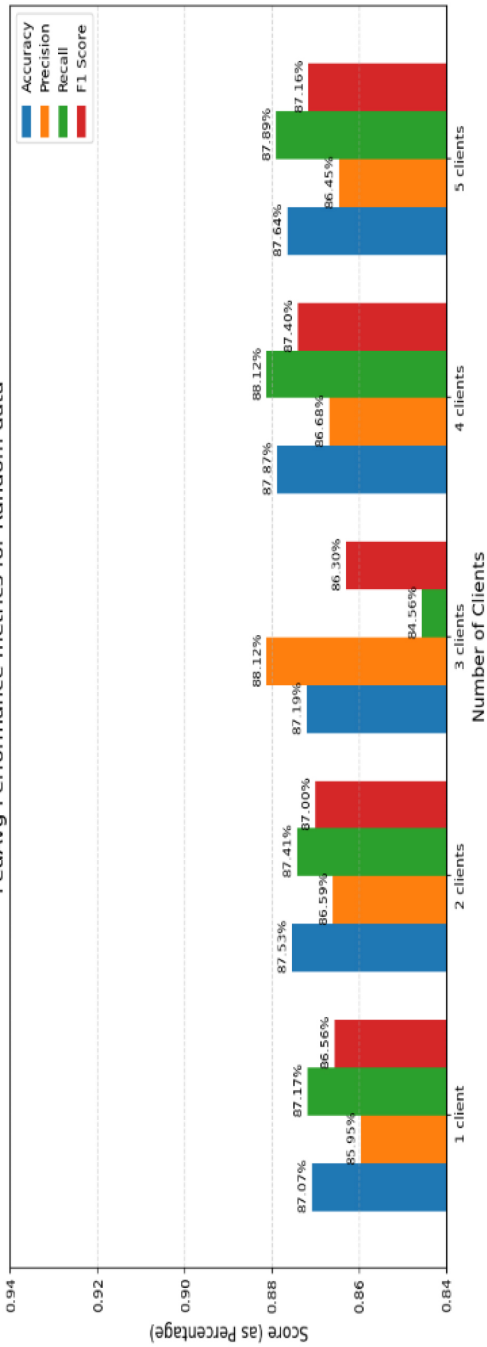

Figure 3

FedAvg Performance metrics for Random data

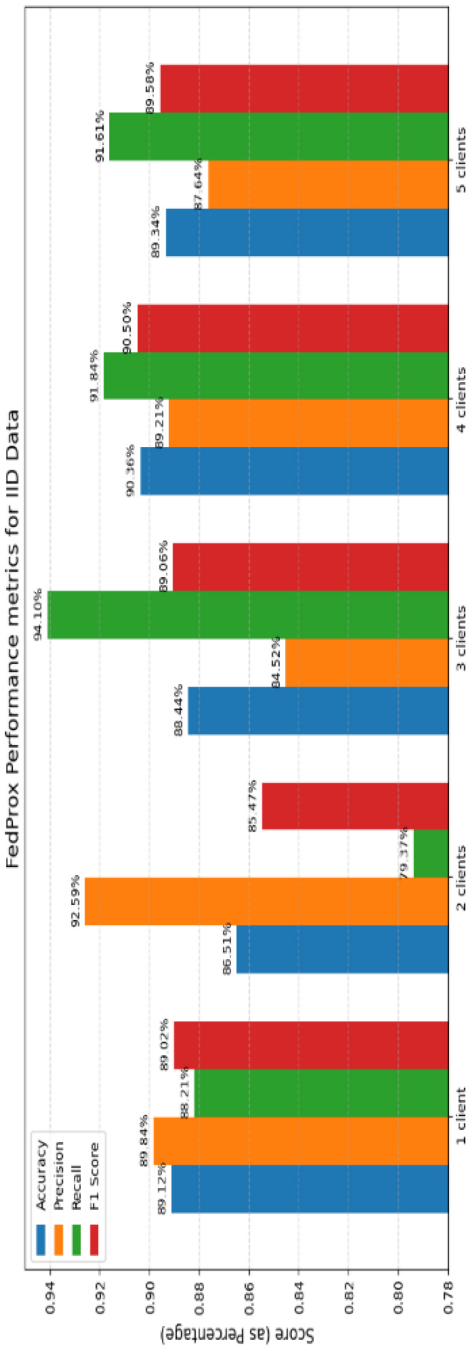

Figure 4

FedProx Performance Metrics for IID Data

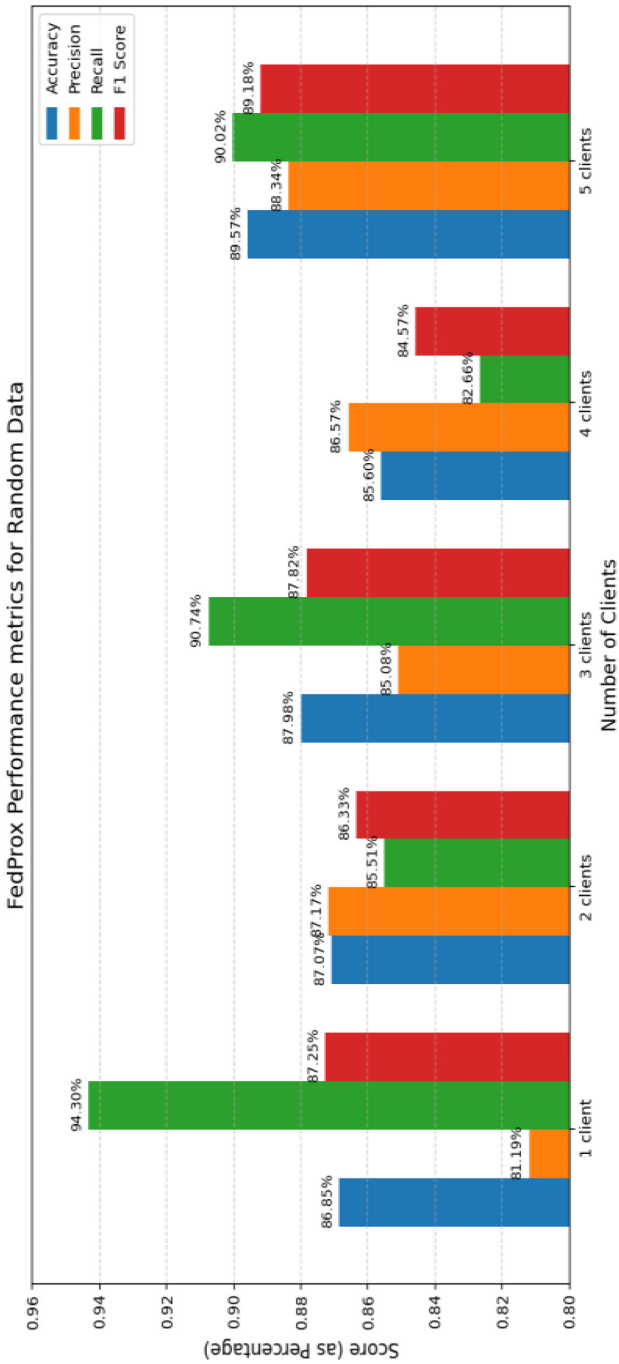

Figure 5

FedProx Performance Metrics for Random Data

In the second case, we assumed an equal data distribution, with each of the five clients possessing one-fifth of the total data. Learning was performed using FedAvg and FedProx over 10 communication rounds in five different conditions. Initially, one client was present, and then one client was added at a time until a total of five clients were present. The findings showed that the more clients involved, the more accurate the global model became. However, the situation with FedProx was different, as there was a slight decrease in accuracy between the two- and three-client cases, which was later compensated for as the number of clients continued to increase. The highest total performance was achieved in terms of the number of clients under FedProx, as the model achieved an accuracy of 88.32 percent during the sixth round. In the meantime, FedAvg performed better in the majority of the evaluation measures when there were five clients. In addition, recall and F1 score showed a positive trend as the number of clients increased, indicating that the classification quality improved. The accuracy steadily increases with the number of clients, highlighting the advantages of greater involvement in the federated learning process. The visual summary of these findings is presented in Figure 6.

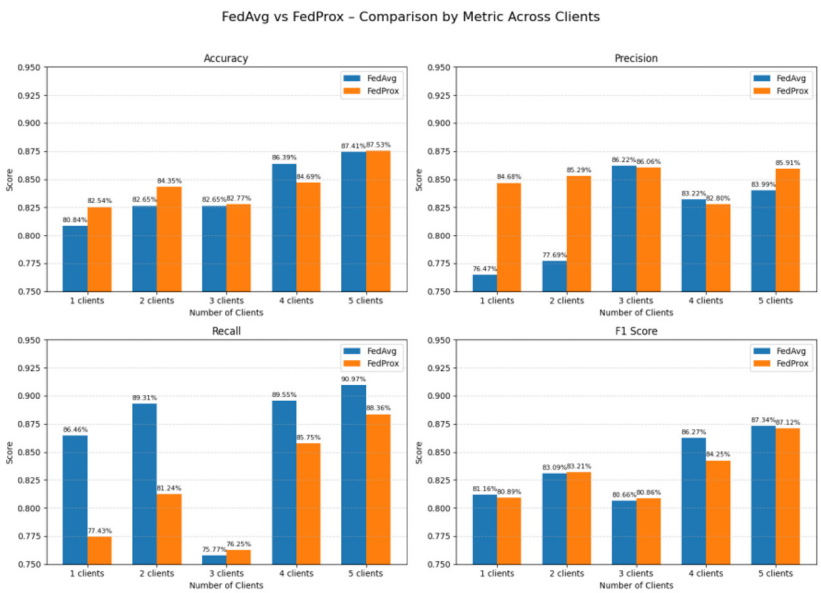

Figure 6

FedAvg vs. FedProx – Comparison

Table 1

Comparison between Existing and Proposed Work

Ref. No.	Dataset Used	Algorithm / Method	Accuracy (%)
[4]	BCI-IV 2a	CNN + Meta-learning (Hierarchical FL)	85.5
[5]	MI OpenBCI	Transformer + FedAvg	89.2
[6]	Physio Net EEG	DP-FedAvg + RNN	87.1
[7]	EEG Eye State	SimCLR + FedSimCLR	82.3
[8]	TUH EEG	GNN + FedAvg	84.6
Our Work	PhysioNet EEG Motor Imagery (109 subjects)	CNN + FedAvg + FedProx	90.36

Table 1 provides a comparison between existing approaches and the proposed framework, highlighting its superior accuracy. Previous studies with CNN, Transformer, RNN, GNN, and SimCLR had achieved accuracies of 82.3% to 89.2% with personalization, privacy, or self-supervised learning as their objectives. Conversely, our CNN with FedAvg and FedProx on the PhysioNet EEG Motor Imagery dataset (109 subjects) achieved the highest accuracy of 90.36%. Figure 7 presents the same graphically.There are two main reasons for the superiority of the proposed framework. First,

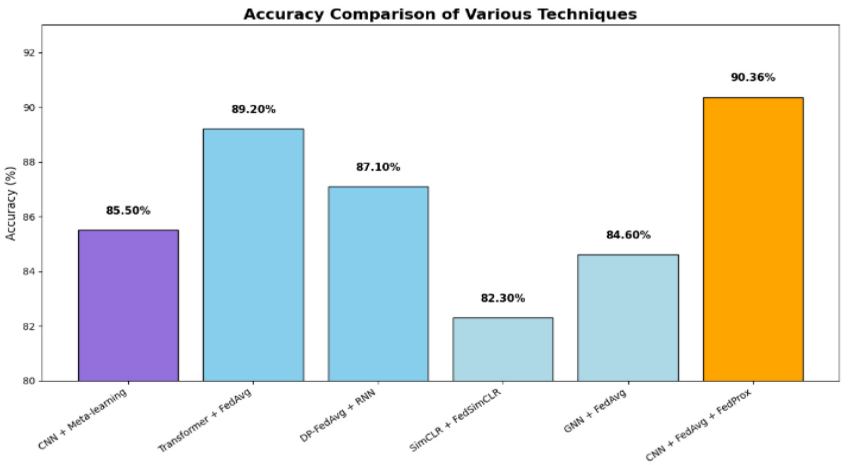

Figure 7

Comparison study of Results

employing CNNs provides the framework with the ability to capture spatio-temporal features of EEG signals. This allows to learn more discriminative features as opposed to shallow FL techniques such as SVM or RNN. Second, using FedProx to implement customized FedAvg systems effectively manages the convergence instability caused by the fragmentation of non-IID distributions, thus boosting the performance of the proposed framework. Together, these characteristics enable the proposed framework to manage subject variability better, thereby improving generalization and robustness across decentralized EEG clients.

5. Conclusion

Even with its strong accuracy and privacy-preserving characteristics, the proposed framework has its shortcomings. One, the research only analyzes the scenario of a single binary-class motor imagery task (left and right hand) and does not analyze performance on more complicated multiclass MI tasks. Two, the simulated federated environment still does not consider possible real-world communication restrictions, such as latency and bandwidth. Three, while privacy has undoubtedly improved, highly refined decentralization techniques, such as homomorphic encryption and differential privacy, which can be incorporated in future works, have not been explored. For future research, a multi-class MI classification may be added to accommodate more advanced Distributed EEG learning system applications. The addition of meta-learning and subject adaptation techniques will enable the personalization of neuro-adaptive systems for new, unseen users.

References

[1] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, and Y. Fukuda, "Intra-and inter-subject variability in EEG-based brain–computer interface: A review," *Frontiers in Computational Neuroscience,* vol. 13, Art. no. 87 (2019), doi: 10.3389/fncom. 2019.00087.

[2] T. Jia, L. Meng, S. Li, J. Liu, and D. Wu, "Federated Motor Imagery Classification for Privacy-Preserving Brain-Computer Interfaces," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 32, pp. 3442-3451 (2024), doi: 10.1109/TNSRE.2024.3457504.

[3] R. Liu, Y. Chen, A. Li, Y. Ding, H. Yu, and C. Guan, "Aggregating intrinsic information to enhance BCI performance through feder-

ated learning," *Neural Networks*, vol. 172, Art. no. 106100 (2024), doi: 10.1016/j.neunet.2024.106100.

[4] J. Wu, Y. Chen, and M. Zhang, "Hierarchical FL for EEG-based BCI: A Personalization Framework," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 450–460 (2023).

[5] Y. Jia, L. Zhou, and F. Lin, "Federated Transformers for EEG Motor Imagery Classification in BCI Applications," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1234–1238 (2024).

[6] H. Kim, D. Lee, and S. Park, "Differential Privacy-Aware FL for EEG-based Brain–Computer Interfaces," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 155–165 (2023).

[7] X. Lin, Y. Qian, and L. Zhao, "Contrastive FL for EEG Classification with Limited Labels," in *Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, pp. 998–1003 (2024).

[8] R. Zhou, A. Sun, and J. Li, "Subject-Specific FL for BCI with EEG Signals," *IEEE Sensors Journal*, vol. 23, no. 5, pp. 6789–6796 (2023).

[9] A. Rao, V. Srinivasan, and P. Roy, "Graph Neural Network-Based FL for EEG Signal Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 215–224 (2024).

[10] M. Hasan, F. Ahmed, and N. Rahman, "Mobile-Friendly FL with Lightweight CNN for Real-Time BCI Applications," *IEEE Access*, vol. 13, pp. 45678–45688 (2025).

[11] S. Patel, R. Kumar, and A. Sinha, "Hybrid Transfer Learning and FL for EEG Classification," in *Proc. IEEE Int. Conf. on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 145–150 (2024).

[12] Y. M. Huang, H. N. Hung, and V. S. Tseng, "Privacy-preserving attention-weighted multi-source domain adaptation for EEG motor imagery," in *Proc. 2024 IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, pp. 7470–7474 (Apr. 2024), doi: 10.1109/ICASSP49456.2024.10098321.

[13] R. Singh, A. Mehta, and P. Verma, "Cluster-Based FL for Heterogeneous EEG Motor Imagery Data," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 2, pp. 112–120 (2025).

[14] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from

decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282 (Apr. 2017).

[15] G. Schalk, "EEG Motor Movement/Imagery Dataset," *PhysioNet* (Sep. 2009). [Online]. Available: https://physionet.org/content/eegmmidb/1.0.0/. [Accessed: Feb. 21, 2025].

[16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013 (2018), doi: 10.1088/1741-2552/aace8c.

[17] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429-450 (2020).

[18] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420 (2017).

[19] Z. Hu, L. Chen, Y. Luo, and J. Zhou, "EEG-based emotion recognition using a convolutional recurrent neural network with multi-head self-attention," *Applied Sciences*, vol. 12, no. 21, Art. no. 11255 (2022), doi: 10.3390/app122111255.

[20] M. A. Abdulhussein, X. An, A. A. Alsakaa, and D. Ming, "Lack of habituation in migraine patients and Evoked Potential types: Analysis study from EEG signals," *Journal of Information and Optimization Sciences*, vol. 43, no. 4, pp. 855–891 (2022).

[21] S. Gautam, A. Khunteta, and D. Ghosh, "Adaptive optimization framework for multimodal software defect prediction using reinforcement learning," *Journal of Information and Optimization Sciences*, vol. 46, no. 7, pp. 2351–2362 (2025).

Received August, 2025