Exploiting Federated Learning for EEG-based Brain-Computer Interface System

1st Mohammadnavid Ghader
Department of Computer Science, Shahid Beheshti University
Tehran, Iran
m_ghader@sbu.ac.ir

2nd Bahar Farahani
Cyberspace Research Institute, Shahid Beheshti University
Tehran, Iran
b_farahani@sbu.ac.ir

3rd Zahra Rezvani
[a]*Institute for Research in Fundamental Sciences (IPM)*
[b]*Department of Computer Science, Alzahra University*
Tehran, Iran
zrezvani@ipm.ir

4th Mahyar Shahsavari
Donders Institute, Radboud University
Nijmegen, The Netherlands
mahyar.shahsavari@ru.nl

5th Mahmood Fazlali
Cybersecurity and Computing Systems Research Group, University of Hertfordshire
Hatfield, United Kingdom
m.fazlali@herts.ac.uk

***Abstract*—Motor imagery (MI) is a widely used technique in brain-computer interface (BCI) systems, which allows users to control external devices using their brain signals. Electroencephalogram (EEG) signals are commonly used to detect and classify MI tasks. However, the lack of annotated data hampers the performance of machine learning (ML) algorithms. Federated learning (FL) is a promising approach to address these challenges, as it allows models to be trained on decentralized datasets without exchanging data. This study proposes an FL approach for MI-EEG signal classification using a convolutional neural network (CNN) on the PhysioNet dataset containing EEG recordings of left and right-hand imagery movements. We evaluate the performance of the FL approach using two different aggregation methods, namely FedAvg and FedProx, and compare it to the centralized ML approach. Furthermore, we explore the effect of increasing the number of clients who participate in the learning process on the performance of the model. Our findings demonstrate that FL maintains consistent classification accuracy comparable to the centralized ML approach while reducing data leakage risks. Thus, FL shows great promise as an essential instrument for MI-EEG signal classification and BCI systems, enabling distributed training of a comprehensive model when privacy-sensitive data is dispersed across multiple clients.**

***Index Terms*—Motor Imagery (MI), Electroencephalography (EEG), Convolutional Neural Network (CNN), Federated Learning (FL)**

I. Introduction

A Brain-Computer Interface (BCI) is a technology that enables the communication between the human brain and a computer system [1]. BCIs typically use electroencephalography (EEG) signals to detect brain activity and translate it into commands for a computer system [2]. The EEG signals are obtained by placing electrodes on the scalp and recording the brain's electrical activity. The signals are then processed and analyzed by a computer system, which translates them into commands for a device or application [3], [4]. There are several types of BCI systems, including invasive and non-invasive systems. Invasive BCI systems involve the implantation of electrodes directly into the brain tissue, while non-invasive BCI systems use external electrodes placed on the scalp to record EEG signals [5].

A common method for BCI is to use Motor Imagery (MI) as a way for individuals to communicate their intentions to the computer or device. The brain generates activity patterns specific to the imagined movement during motor imagery. BCI systems can use electroencephalography (EEG) or other brain imaging techniques to detect these activity patterns and translate them into commands for the computer or device [6]. For example, in a BCI system that uses motor imagery, an individual may imagine moving their left or right hand to select a particular item on a computer screen.

Several Machine Learning (ML) techniques have been applied to this task, but the lack of annotated data hinders their performance. Federated learning (FL) is a relatively new approach to ML that allows models to be trained on decentralized datasets without exchanging data [7], [8], [9]. This technique can potentially address the challenges of MI-EEG signal classification, including data privacy, security, and scalability.

In 2016, Google researchers introduced the concept of Federated Learning (FL) [10]. They proposed a framework for training deep neural networks on mobile devices without compromising user privacy. Instead of sending data to a central server, FL involves training the model on local data and only sending model updates to a central server for aggregation. This

979-8-3503-4647-3/23/$31.00 ©2023 IEEE

means FL can enable researchers to train a machine-learning model using data from multiple clients without requiring them to share their sensitive health data with others. Another advantage of using FL for motor imagery EEG signal classification is that it can enable the model to be updated and improved over time as more data becomes available. This can result in a more accurate and robust model for classifying motor imagery EEG signals.

In this paper, we applied FL as MI-EEG signal classification on the PhysioNet dataset for left-hand and right-hand imagery movement. This dataset contains EEG recordings of 109 healthy subjects performing MI tasks with their left and right hands. Our contribution is that we evaluate the performance of FL with two aggregation methods, FedAvg and FedProx, in classifying MI-EEG signals for left and right-hand imagery movements. The results showed that the performance of the FL model improved as more clients and data were added to the system. However, there was also a point of diminishing returns, where adding more clients or data did not significantly improve performance. This suggests that there may be a practical limit to the number of clients that can effectively contribute to the FL system.

The study includes five sections. First, we provide related work on MI-EEG signal classification and FL in Section II. Section III describes the PhysioNet dataset and the preprocessing steps we applied to the data. Also, we present our methodology for FL. In Section IV, we present our experimental results, which demonstrate the effectiveness of the proposed method. Section V discusses the implications of our results and potential future research directions.

II. Related Work

Numerous techniques have been employed to accomplish the MI-EEG classification process across multiple phases. During the preprocessing phase, the primary objective is to eliminate noise and artifacts from the EEG signal to ensure precision and significance in subsequent analyses. This stage involves diverse operations such as electrode selection, referencing, filtering, artifact removal, epoching, downsampling, and interpolation. Feature extraction can be performed manually by domain specialists or implicitly managed within succeeding machine learning stages. Three primary categories of features- temporal, spectral, and spatial features- are typically observed in EEG studies, each requiring distinct feature extraction approaches based on their unique characteristics. Common Spatial Patterns (CSP) [11] is a popular manual spatial feature extraction technique that has been used in several studies. Some variations of CSP include Sparse CSP, Stationary CSP, Divergence CSP, Probabilistic CSP, and Filter-Bank CSP. Filter-Bank CSP (FBCSP) [12] has shown the best performance on MI classification out of all the other methods that rely on manual feature extraction.

Different algorithms have been used in the studies in the model selection phase. Support vector machine (SVM) [13], K-Nearest Neighbor (KNN) [14], and linear discriminant analysis (LDA) [15] as supervised machine learning algorithms observed in studies that have been carried out further in the past. With the widespread use of deep learning and artificial neural networks in various fields, they were also used in the field of MI-EEG signal classification. CNN (Convolutional Neural Network) models are a deep learning architecture with promising results in classifying MI-EEG (Motor Imagery Electroencephalography). CNN models are well suited for MI-EEG classification because they can automatically learn both temporal and spatial features from the raw EEG data without the need for manual feature extraction, and this can be done automatically. This is achieved through the use of convolutional layers, which apply filters to the EEG data to extract relevant features. In MI-EEG classification, CNN models typically consist of several convolutional layers followed by one or more fully connected layers. The convolutional layers extract relevant features from the EEG data, while the fully connected layers classify the input into different motor imagery tasks. Some end-to-end models have been introduced for classifying MI-EEG signals with CNN models. Dose et al. [16] implemented an artificial neural network comprising two convolutional layers, designed such that the first layer extracts temporal features while the second focuses on spatial features. For the final classification, they employed a fully connected layer. Similarly, Lun et al. [17] suggested employing a CNN structure with separate temporal and spatial filters. This allows for feature extraction without the need for any manual data processing or explicit feature extraction. However, they have used a relatively deeper CNN with five convolutional layers and a fully connected layer as a classifier. Hwaidi et al. [18] combined deep autoencoder (DAE) and CNN in their study. Since DAE can extract nonlinear features and eliminate superfluous information from EEG signals, they utilized it before importing data into CNN.

In recent years, with the introduction of transformer architecture and its effective applications in fields such as natural language processing, Xie et al. [19] studied using this architecture in EEG signal classification. They used five types of Transformer-based models, and the obtained results demonstrate that they provided good performance for MI-EEG classification, especially for the 3s data.

Regarding insufficient training data for EEG signal classification, Ju et al. [20] proposed a federated transfer learning method for subject-adaptive MI-EEG signal classification. The result of this study shows that for mitigating low-appropriate data, covariance-based representation of EEG signals can help achieve better deep learning architecture.

In this paper we employed two mainstream federated learning techniques, FedAvg and FedProx and considered the model's performance changes concerning the number of federated learning clients and the size of local data. Also we investigate how the model's performance and quality can be improved as the number of clients and the amount of data contributions increased.

TABLE I: CNN architecture used for classification.

	neurons	kernel size	pool size	activation	padding	params	output shape
Conv2d	40	(30,1)	-	leaky relu	same	1240	(497,64,40)
Conv2d	40	(1,64)	-	leaky relu	valid	102440	(497,1,40)
AveragePooling2D	-	-	(15,1)	-	valid	0	(33,1,40)
Flatten	-	-	-	-	-	0	1320
FC	80	-	-	relu	-	105680	80
FC	2	-	-	softmax	-	162	2

III. Methodology

The purpose of this study is to evaluate the efficacy of federated learning in the classification of motor imagery (MI) from EEG signals. Firstly, we will introduce the dataset used and detail the preprocessing applied to it. We will then present the architecture of our deep learning model, followed by an explanation of the implementation process for the federated learning methodology.

A. Dataset Description

The Physionet EEG dataset is a publicly available collection of electroencephalogram (EEG) recordings obtained from 109 volunteers, which the BCI2000 system developers recorded [21]. Each subject has performed 14 runs, including two baseline runs and three rounds of four motor imagery/movement tasks. The eyes should be opened and closed in the first and second baseline runs. Motor tasks consist of:

- left or right fist movement
- left or right fist imagery movement
- both fist or feet movement
- both fist or feet imagery movement

The EEG recordings were obtained using a standard 64-electrode placement system, and the signals were amplified and digitized. The task labels are provided for each recording, indicating the subject's task during the EEG recording. The signals obtained from each electrode were sampled at 160 HZ and stored in EDF+ format.

This study used the left or right fist imagery movement task for binary classification.

B. Preprocessing

In this phase, we use the MNE [22], an open-source Python package for exploring, visualizing, and analyzing human neurophysiological data. The set of actions we performed on the data before the model training process are:

- Subject selection: 105 subjects from the dataset were selected (subjects with the id numbers 88, 89, 92 and 100 were not used because of the problems they had in the data stored)
- Bad channels removal: bad channels were eliminated using the functionality included in the MNE package
- Data epoching: 6s data were used for classification (this included the entire motor imagery period as well as two seconds after it)
- Data unit conversion: data were converted from volt to microvolt

C. CNN Model

In this study, CNN architecture has been used for the base model, and it is derived from the structure that was presented in the paper of Dose et al. [16]. However, the activation functions of convolutional layers were changed from relu to leaky relu. In the first layer, a Conv2D layer with 40 filters and a kernel size of 30×1 was placed for temporal feature extraction. In the second layer, a Conv2D layer with 40 filters and a kernel size of 1×64 was placed for spatial feature extraction. After the these two convolutional layers, an AveragePooling2D layer with a pool size of 15×1 is placed for down-sampling. For the classification, data were flattened, and after passing through a fully connected layer with 80 neurons, they were classified into two classes. Table I shows details of the CNN architecture that was chosen and its parameters.

D. Federated Learning Framework

In this study, we incorporate two distinct Federated Learning (FL) techniques, namely FedAvg and FedProx. As shown in Algorithm 1, FedAvg operates iteratively such that in each round, The server randomly selects K clients from a pool of N clients and distributes the generic model parameters to those chosen individuals. Each client, in turn, trains a local model by optimizing the loss function F_k using their private local data for E epochs, then uploads the corresponding parameters back to the server. These parameters are then averaged to refine the model. The aggregated parameters thus define the global model for the subsequent round [10]. FedProx is an extension of FedAvg designed to handle the issues of non-Independent and Identically Distributed (non-IID) data and system heterogeneity by incorporating a proximity term to the optimization objective (See Algorithm 2) [23]. The idea behind the introduction of the new term is to penalize large deviations from global model parameters during local training.

This paper examines two distinct scenarios to conduct a thorough investigation of various aspects related to EEG-based BCI systems in Federated Learning.

- Scenario 1: The fundamental concept behind scenario 1 is to evaluate the performance of FL in comparison to centralized learning. In the FL setup, we distribute the entire dataset across five clients. Nonetheless, the cumulative data across all clients matches the size of the data utilized in centralized learning.
- Scenario 2: The central concept of Scenario 2 is to assess the impact of data size and the number of clients on FL performance. In this scenario, we start by distributing

Algorithm 1 Federated Averaging (FedAvg)

1: **Initialize** w_0 ▷ Initial global model parameters
2: **for** each round $t = 1, 2, \dots$ **do**
3: **for** each client $k \in N$ **in parallel do**
4: $w_t^k \leftarrow$ (Train a local model using w_{t-1}, optimizing the loss F_k with local data on client k)
5: **end for**
6: $w_t \leftarrow \sum_{k \in N} \frac{n_k}{n} w_t^k$ ▷ Update global model
7: **end for**

Note that w_t represents the global model parameters at round t. n_k is the number of local data points at client k, and n is the total number of data points across all clients.

Algorithm 2 Federated Proximal (FedProx)

1: **Initialize** w_0 ▷ Initial global model parameters
2: **for** each round $t = 1, 2, \dots$ **do**
3: **for** each client $k \in N$ **in parallel do**
4: $w_t^k \leftarrow$ (Solve $\min_w F_k(w) + \frac{\mu}{2}||w - w_{t-1}||^2$ with local data on client k) ▷ w_t^k are the local model parameters of client k at round t
5: **end for**
6: $w_t \leftarrow \sum_{k \in N} \frac{n_k}{n} w_t^k$ ▷ w_t are the updated global model parameters at round t
7: **end for**

Note that μ represents the regularization coefficient. FedAvg is a special case of FedProx where $\mu = 0$.

the entire data among five clients. The learning process begins with a single client, and with each subsequent step, we incrementally add another client until we reach a total of five. It is crucial to note that with each added client, an additional set of data is incorporated into the system. As such, when we only have one client, the size of the data in FL is one-fifth that of centralized learning. Conversely, when all five clients are participating, the data size in FL equals that of centralized learning.

IV. Experimental Results

We conduct our experiments on the Physionet dataset described in the previous section. In the experiments, only the imagery movement part of the dataset corresponding to the right or left fist was employed for the binary classification task.

As mentioned earlier, we designed two scenarios to explore the proposed method. In the experiments, we report our results in two paradigms for data distribution.

1) IID paradigm: To simulate IID (independent and identically distributed) environment, the preservation of an equal number of samples from each class is manually maintained.

2) Random paradigm: In this paradigm, samples were assigned to the clients accidentally. So, the number of instances may vary in each class and client. This paradigm was designed to simulate more closely to real-world behavior.

It should be noted that the sample size in all clients was considered equal in both scenarios. For model evaluation, the data was divided into 80% for training and 20% for testing. We used Adam as the optimizer with an initial learning rate of 1e-4 and the presumed values for the number of local epochs and aggregation rounds are equal to 2 and 10, respectively. A batch size of 60 was utilized in all experiments. In the following, the results of the two scenarios are given in separate subsections.

A. First Scenario

In the first scenario, we examine the centralized training across decentralized ones aggregated with FL. So, FL can manifest itself in this scenario. In this experiment, we carefully examined the FL approach using two aggregation methods with varying numbers of clients. To aggregate the clients' model weights, the investigation process utilized FedAvg and FedProx, with the number of clients ranging from one to five. Also, data were distributed among clients in both IID and random paradigms. The evaluation results were summarized in TABLE II and III.

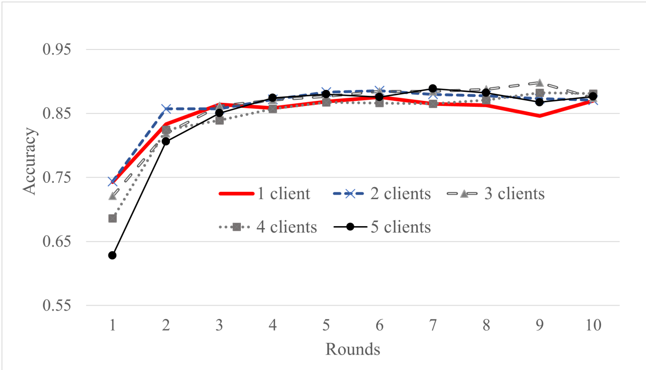
TABLE II: FL classification evaluation metrics of FedAvg in the first scenario.

Data	Method	Accuracy	Precision	Recall	F1 score
IID	1 client	0.8696	0.8622	0.8798	0.8709
	2 clients	0.8707	0.8547	0.8934	0.8736
	3 clients	0.8741	0.8784	0.8685	0.8734
	4 clients	**0.8810**	0.8485	**0.9274**	**0.8862**
	5 clients	0.8764	0.8593	0.9002	0.8793
Random	1 client	0.8707	0.8595	0.8717	0.8656
	2 clients	0.8753	0.8659	0.8741	0.8700
	3 clients	0.8719	**0.8812**	0.8456	0.8630
	4 clients	0.8787	0.8668	0.8812	0.8740
	5 clients	0.8764	0.8645	0.8789	0.8716

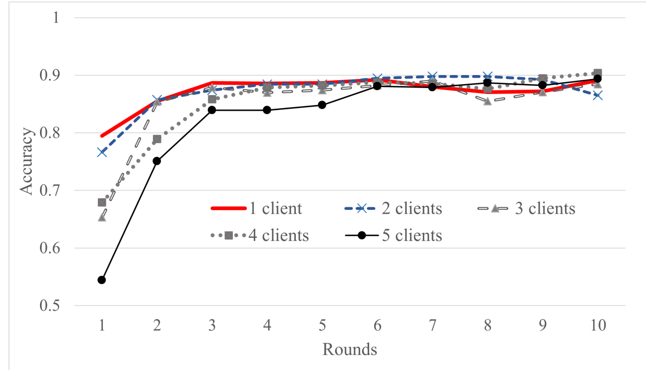
TABLE III: FL classification evaluation metrics of FedProx in the first scenario.

Data	Method	Accuracy	Precision	Recall	F1 score
IID	1 client	0.8912	0.8984	0.8821	0.8902
	2 clients	0.8651	**0.9259**	0.7937	0.8547
	3 clients	0.8844	0.8452	0.9410	0.8906
	4 clients	**0.9036**	0.8921	0.9184	**0.9050**
	5 clients	0.8934	0.8764	0.9161	0.8958
Random	1 client	0.8685	0.8119	**0.9430**	0.8725
	2 clients	0.8707	0.8717	0.8551	0.8633
	3 clients	0.8798	0.8508	0.9074	0.8782
	4 clients	0.8560	0.8657	0.8266	0.8457
	5 clients	0.8957	0.8834	0.9002	0.8918

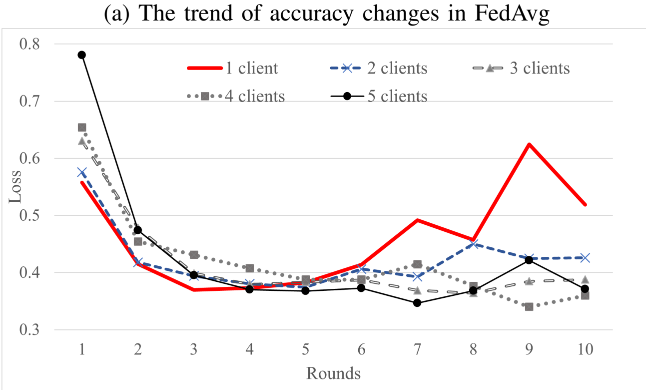
1) Aggregation using FedAvg: During our experimentation with FedAvg, we observed that in the IID paradigm, when the number of clients was four, we achieved superior results compared to other scenarios in terms of most evaluation metrics. In this particular case, the accuracy of the global model at the end of the tenth round was 88.10%. However, if we consider the highest accuracy achieved across all rounds, we found that with three clients, the accuracy of the global model reached 89.80% at the end of the ninth round. When we compared the accuracy figure of the global model in the IID paradigm, the global model reached convergence earlier, and the convergence processes in different states of the number



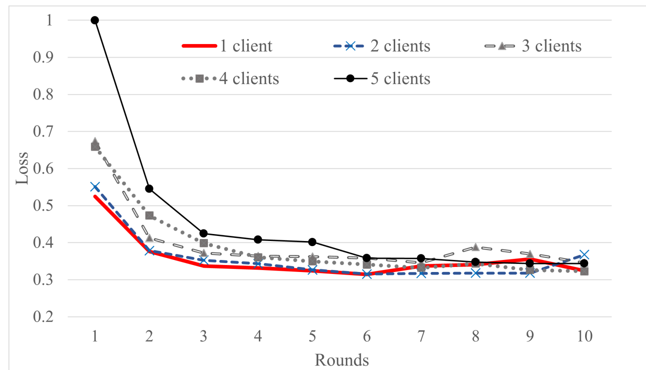
(a) The trend of accuracy changes in FedAvg



(b) The trend of accuracy changes in FedProx



(c) The trend of loss changes in FedAvg



(d) The trend of loss changes in FedProx

Fig. 1: The accuracy and loss of the global model in IID labels distribution.

of clients were more consistent. We calculated classification evaluation metrics such as precision, recall, and F1 score for all cases of participating clients, provided in TABLE II.

2) Aggregation using FedProx: When FedProx was used with a value of $\mu = 0.3$, the highest accuracy was achieved with four clients with the IID paradigm. In this scenario, the clients were able to increase the accuracy of the global model to 90.36%, which was the highest accuracy value obtained across both paradigms, the number of clients, and the aggregation algorithms used. When comparing the accuracy values obtained in the IID paradigm, FedAvg showed a higher convergence speed than FedProx, but FedProx resulted in higher final accuracy values. Additionally, the changes in the loss were much smoother with FedProx compared to FedAvg in the IID paradigm.

In the case of random paradigm, the highest accuracy value for the global model was achieved with five clients in the learning process with FedProx, reaching a value of 89.57%. When analyzing the highest values obtained in different evaluation metrics, it was found that the values obtained in FedProx were larger than those obtained in FedAvg. Also, random paradigm with five clients, FedProx outperformed FedAvg in all evaluation metrics under similar conditions. The trends of the global model accuracy and loss in the IID paradigm with FedAvg and FedProx are illustrated in Fig. 1, while the classification evaluation metrics, such as precision, recall, and F1 score for all cases of client counts, are provided in TABLE III.

B. Second Scenario

As previously discussed, federated learning can be applied when data scarcity arises due to privacy concerns. In such instances, the learning process can be executed using federated learning without transferring data from its original source. We introduced an additional scenario to explore the impact of the number of clients participating in the learning process.

In this second scenario, we assumed that the data were evenly distributed among the clients, with each client holding one-fifth of the total data volume. The learning process was then conducted using two aggregation algorithms, FedAvg and FedProx, each featuring five distinct conditions based on the number of participating clients.

At first, we assume that only one client has participated. Then after checking its performance, we assume two clients were present in the learning process. After that, the number of participating clients increases one by one until their number reaches five. In this scenario, the learning process in all conditions continued up to 10 rounds. By observing the results obtained at the end of the 10th round, it can be seen that more accuracy has been obtained for the model with the increase in the number of participating clients. However, in FedProx, when there were three clients compared to two clients, we saw a tiny drop in accuracy. We can also see that the highest accuracy obtained among all rounds and all scenarios is related to the sixth round of FedProx when there were five clients, where the global model reached an accuracy of 88.32%.

The classification evaluation metrics of the second scenario are summarized in TABLE IV. According to this table, in FedAvg, we see superiority in most metrics when five clients participated. Also, in most cases in both aggregation methods, the trend of recall and F1 score values have been ascending with the increase in the number of clients, and the accuracy always has an upward trend across the number of clients.

TABLE IV: FL classification evaluation metrics of FedAvg and FedProx in the second experiment

Aggregation	Method	Accuracy	Precision	Recall	F1 score
FedAvg	1 client	0.8084	0.7647	0.8646	0.8116
	2 clients	0.8265	0.7769	0.8931	0.8309
	3 clients	0.8265	**0.8622**	0.7577	0.8066
	4 clients	0.8639	0.8322	0.8955	0.8627
	5 clients	0.8741	0.8399	**0.9097**	**0.8734**
FedProx	1 client	0.8254	0.8468	0.7743	0.8089
	2 clients	0.8435	0.8529	0.8124	0.8321
	3 clients	0.8277	0.8606	0.7625	0.8086
	4 clients	0.8469	0.8280	0.8575	0.8425
	5 clients	**0.8753**	0.8591	0.8836	0.8712

V. Conclusion

EEG-based BCIs allow users to interact with devices using brain signals directly. Despite advancements in Motor Imagery signal classification, data privacy presents a significant challenge. To tackle this, our study suggests the application of a federated learning approach, demonstrated through two specific scenarios. The first scenario aimed to compare the functional differences between the model in a centralized context and two mainstream federated learning techniques, FedAvg and FedProx. Notably, the data size between the centralized model and clients kept consistent. The second scenario's key objective was to investigate how the model's performance changes concerning the number of federated learning clients and the size of local data. The results indicated a comparable performance between the models trained in a federated manner and those trained in centralized conditions. Notably, the use of the FedProx algorithm outperformed FedAvg, especially with increased client participation. The second scenario demonstrated that the model's performance and quality improved as the number of clients and the amount of data contributions increased. One crucial finding was the significant impact of data distribution among clients on the model's learning process. For future research directions, we anticipate focusing on the impact of heterogeneous data distribution and the non-independently and identically distributed (non-IID) problem among clients on the performance of federated learning models.

References

[1] Kawala-Sterniuk A, Browarska N, Al-Bakri A, Pelc M, Zygarlicki J, Sidikova M, Martinek R, Gorzelanczyk EJ, "Summary of over Fifty Years with Brain-Computer Interfaces—A Review." Brain Sciences. 2021.

[2] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," Sensors, vol. 12, no. 2. pp. 1211–1279, 2012.

[3] Z. Rezvani, R. Khosrowabadi, A. Seyedebrahimi, G.-H. Meftahi, and B. Hatef, 'Alteration of brain functional network and cortisol level during induction and release of stress: An EEG study in young male adults', Basic and Clinical Neuroscience, 2020.

[4] V. Sobhani, Z. Rezvani, G. H. Meftahi, F. Ghahvehchi-Hosseini, and B. Hatef, 'The Linear and Nonlinear Indices of Electroencephalography Change in the Stroop Color and Word Test', Journal of Modern Rehabilitation, 2022.

[5] S.Waldert, "Invasive vs. Non-Invasive Neuronal Signals for Brain-Machine Interfaces: Will One Prevail?." Frontiers in neuroscience vol. 10 295. 27 Jun. 2016.

[6] S.-M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface," Inf. Sci., vol. 178, pp. 1629–1640, Mar. 2008.

[7] F. Firouzi et al., "Fusion of IoT, AI, Edge–Fog–Cloud, and Blockchain: Challenges, Solutions, and a Case Study in Healthcare and Medicine," in IEEE Internet of Things Journal, vol. 10, no. 5, pp. 3686-3705, 1 March1, 2023.

[8] F. Firouzi, B. Farahani, M. Barzegari and M. Daneshmand, "AI-Driven Data Monetization: The Other Face of Data in IoT-Based Smart and Connected Health," in IEEE Internet of Things Journal, vol. 9, no. 8, pp. 5581-5599, 15 April15, 2022.

[9] A. Nadian-Ghomsheh, B. Farahani, and M. Kavian, "A hierarchical privacy-preserving IoT architecture for vision-based hand rehabilitation assessment - Multimedia Tools and Applications," SpringerLink, Feb. 13, 2021.

[10] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," Proc. 20th Int. Conf. Artif. Intell. Stat. AISTATS 2017, 2017.

[11] Wang, Yijun, Shangkai Gao, and Xiaorong Gao. "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface." In 2005 IEEE engineering in medicine and biology 27th annual conference, pp. 5392-5395. IEEE, 2006.

[12] Ang, Kai Keng, et al. "Filter bank common spatial pattern (FBCSP) in brain-computer interface." 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008.

[13] D. Planelles, E. Hortal, Á. Costa, A. Úbeda, E. Iáez, and J. Azorín, "Evaluating classifiers to detect arm movement inten- tion from EEG signals," Sensors, vol. 14, no. 10, pp. 18172–18186, Oct. 2014.

[14] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. Tibarewala, "Performance analysis of LDA, QDA and KNN algo- rithms in left-right limb movement classification from EEG data," in Proc. Int. Conf. Syst. Med. Biol. Syst. (ICSMB), Dec. 2010, pp. 126–131.

[15] S. Bhattacharyya, A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar, "Performance analysis of left/right-hand movement classification from EEG signal by intelligent algorithms," in Proc. IEEE Symp. Comput. Intell., Cogn. Algorithms, Mind, Brain (CCMB), Apr. 2011, pp. 1–8.

[16] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," Expert Syst. Appl., vol. 114, pp. 532–542, Dec. 2018.

[17] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, "A simplified CNN classification method for MI-EEG via the electrode pairs signals," Front. Hum. Neurosci., vol. 14, 2020.

[18] J. F. Hwaidi and T. M. Chen, "Classification of Motor Imagery EEG Signals Based on Deep Autoencoder and Convolutional Neural Network Approach," in IEEE Access, vol. 10, pp. 48071-48081, 2022.

[19] J. Xie et al., "A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2126-2136, 2022.

[20] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu and C. Guan, "Federated Transfer Learning for EEG Signal Classification," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Montreal, QC, Canada, 2020.

[21] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering 51(6):1034-1043, 2004.

[22] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, et al., MEG and EEG data analysis with MNE- Python (2013).

[23] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," 2020.