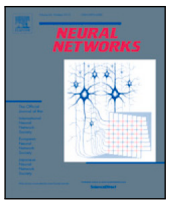
Neural Networks 172 (2024) 106100

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

Aggregating intrinsic information to enhance BCI performance through federated learning

Rui Liu, Yuanyuan Chen, Anran Li, Yi Ding, Han Yu, Cuntai Guan *

School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore

ARTICLE INFO

Keywords:
Federated learning
Brain–computer interface
Heterogeneous datasets
Motor imagery

ABSTRACT

Insufficient data is a long-standing challenge for Brain–Computer Interface (BCI) to build a high-performance deep learning model. Though numerous research groups and institutes collect a multitude of EEG datasets for the same BCI task, sharing EEG data from multiple sites is still challenging due to the heterogeneity of devices. The significance of this challenge cannot be overstated, given the critical role of data diversity in fostering model robustness. However, existing works rarely discuss this issue, predominantly centering their attention on model training within a single dataset, often in the context of inter-subject or inter-session settings. In this work, we propose a hierarchical personalized Federated Learning EEG decoding (FLEEG) framework to surmount this challenge. This innovative framework heralds a new learning paradigm for BCI, enabling datasets with disparate data formats to collaborate in the model training process. Each client is assigned a specific dataset and trains a hierarchical personalized model to manage diverse data formats and facilitate information exchange. Meanwhile, the server coordinates the training procedure to harness knowledge gleaned from all datasets, thus elevating overall performance. The framework has been evaluated in Motor Imagery (MI) classification with nine EEG datasets collected by different devices but implementing the same MI task. Results demonstrate that the proposed framework can boost classification performance up to 8.4% by enabling knowledge sharing between multiple datasets, especially for smaller datasets. Visualization results also indicate that the proposed framework can empower the local models to put a stable focus on task-related areas, yielding better performance. To the best of our knowledge, this is the first end-to-end solution to address this important challenge.

1. Introduction

Brain–computer interface (BCI) is a crucial technology that establishes a connection between the human brain and external devices, which has clinical and non-clinical applications in many areas, such as movement capability recovery and assistance (Mane, Chouhan, & Guan, 2020), cognitive health (Lee et al., 2013), and entertainment (Nijholt, Contreras-Vidal, Jeunet, & Väljamäe, 2022). Electroencephalogram (EEG) is one of the most commonly used signals in BCI to decode brain activities. In recent years, deep learning algorithms have been employed in EEG decoding and classification tasks. Since the success of deep learning algorithms is largely attributed to the availability of large amounts of data, collecting enough data is important for EEG decoding. However, EEG data collection from humans can be challenging and costly. The physiological limitations of subjects limit the number of samples collected from one person. And the significant cost and complicated usage of the devices limit the number of participants in one dataset. Thus, most datasets are of middle size, collected from dozens of subjects with hundreds of samples per subject at most.

Existing works enhance EEG decoding model performance mainly by sharing knowledge between subjects or sessions within one single dataset (Wan, Yang, Huang, Zeng, & Liu, 2021; Wei, Ortega, & Faisal, 2021). However, there are some datasets designed with the same task and protocol. Drawing upon the widely acknowledged principle that larger training datasets tend to yield enhanced classification outcomes in deep learning approaches, it is anticipated that advancements in model performance could be realized through the amalgamation of knowledge from these datasets as a larger virtual training set. Unfortunately, EEG data collected by various devices have heterogeneous formats, in terms of the number and location of EEG channels, sampling rates, and amplifiers. This device heterogeneity problem prevents knowledge sharing among datasets. A few works made initial attempts to solve this problem by dropping channels or padding with zeros, which may lose information or add noises to the data (Bakas et al., 2022; Gu et al., 2022; Saeed, Grangier, Pietquin, & Zeghidour, 2021). This important issue remains unresolved.

* Corresponding author.
E-mail addresses: rui.liu@ntu.edu.sg (R. Liu), yuanyuan.chen@ntu.edu.sg (Y. Chen), anran.li@ntu.edu.sg (A. Li), ding.yi@ntu.edu.sg (Y. Ding), han.yu@ntu.edu.sg (H. Yu), ctguan@ntu.edu.sg (C. Guan).

https://doi.org/10.1016/j.neunet.2024.106100
Received 14 August 2023; Received in revised form 20 November 2023; Accepted 3 January 2024
Available online 9 January 2024

Federated learning (FL) is an emerging collaborative machine learning paradigm to train models across multiple data owners and enables local models to benefit from each other while keeping local data private (Goebel, Yu, Faltings, Fan, & Xiong, 2023; Kairouz et al., 2021; Li, Zhang, Wang, Han, & Li, 2021; Yang, Liu, Cheng, Kang, Chen, & Yu, 2020). FL has been applied in many medical and healthcare applications, such as disease prediction (Peng, Wang, Dvornek, Zhu, & Li, 2022), and brain template estimation (Bayram & Rekik, 2021), to facilitate cross-silo cooperation with privacy protection. Personalized Federated Learning (PFL) (Tan, Yu, Cui, & Yang, 2022) is a branch of FL dealing with the heterogeneity issue across clients, including the heterogeneity in data distributions, network structures, and data formats. It inspires us to design the proposed framework to solve the device heterogeneity issue in the EEG decoding application.

In this work, we provide a new solution to enlarge the training set by including more related datasets. We design a hierarchical personalized Federated Learning EEG decoding (FLEEG) framework to solve the device heterogeneity issue. The framework makes use of the PFL architecture to facilitate cooperative training among multiple device-heterogeneous EEG datasets. It consists of a server and several clients. The server orchestrates local models' training in clients to obtain the optimal personalized models for each dataset assigned to the clients. The personalized model compromises a local module followed by a global one. The local module is responsible for extracting features with the same formats from device-heterogeneous datasets, while the global module transfers the knowledge between datasets to improve model performances. We evaluated the proposed framework with nine real Motor Imagery (MI) EEG datasets collected by multiple institutes. It improves the model performances on most datasets compared to independent training, especially for smaller datasets with an improvement of up to 8.4%. This framework provides a new general learning paradigm for the BCI community to train higher-performance models with multiple datasets, instead of one.

In summary, to the best of our knowledge, this is the first endeavor to tackle the device heterogeneity issue among multiple EEG datasets with a federated-learning-based end-to-end solution for training deep learning models. We briefly summarize the contributions of our work as follows:

- To obtain a higher-performance model, we provide a new learning paradigm for the BCI community to train the EEG decoding models with an enlarged training set consisting of multiple datasets, instead of one dataset.
- To solve the device heterogeneity issue, we propose a hierarchical personalized federated-learning-based framework, named FLEEG, enabling knowledge sharing between datasets during the model training process. Each client is assigned a dataset and trains a personalized model, consisting of a local module to align data formats and a global module to transfer knowledge between datasets.
- To validate the performance of the proposed framework, we evaluate the proposed framework on nine real EEG MI datasets. The results demonstrate remarkable improvements in the performance of local models across the majority of datasets, especially for small datasets with an improvement of up to 8.4%. We further analyze the factors for the improvements and provide visualized interpretation.

The paper is structured in the following way: Section 2 covers related works, Section 3 explains the proposed framework in detail, and Section 4 describes the experiment setups and reports the experiment results. Section 5 provides some analysis and discussions on the results. The paper concludes in Section 6.

2. Related works

2.1. Device-heterogeneity issue in EEG

The issue of device-heterogeneity in EEG applications has received limited attention in existing studies. Most of the research has primarily focused on transferring knowledge across subjects or sessions within one dataset (Ding, Robinson, Tong, Zeng, & Guan, 2023; Wan et al., 2021; Wei et al., 2021; Zhang et al., 2020). Only a few works have made attempts to address the heterogeneity across datasets in EEG applications. Some works (Bakas et al., 2022; Cui et al., 2023; Gu et al., 2022; Kuang et al., 2021; Saeed et al., 2021; Xu et al., 2020) make the data formats consistent by either deleting channels or padding with zeros. However, these methods can introduce disturbances into the EEG signal, either by losing valuable information or adding noises. Other researchers have explored the channel mapping methods as a separate feature extractor during pre-processing on heterogeneous datasets (Kostas, Aroca-Ouellette, & Rudzicz, 2021). Nevertheless, since this feature extractor is not integrated into the end-to-end training process, it lacks the flexibility to adapt to each dataset and may introduce noises into the analysis.

2.2. Federated learning

Federated learning (FL) is a collaborative machine learning paradigm to train models across distributed data owners and enables local models to benefit from each other. It allows privacy-preserving, especially for applications that use sensitive or personal data. In FL, data owners with local data can be referred to as *clients* if they are coordinated by a central entity referred to as the *server*.

Since the data is isolated in the clients, PFL has been developed to solve the heterogeneity issues across clients (Gao, Yao, & Yang, 2022; Tan et al., 2022). Because of the heterogeneity of data distributions, data formats, local model architectures, and computation abilities, various solutions have been proposed. For the data format heterogeneity issue, some works (Bica & van der Schaar, 2022; Feng, Li, Yu, Liu, & Yang, 2022; Liang et al., 2020) design local encoders to project the data to a common space first and then transfer knowledge between clients based on the aligned features.

FL algorithms have been applied to some BCI applications recently. Hang et al. (2023) applies FL to EEG decoding with the cross-subject task. Hu et al. (2021), Ju et al. (2020) borrow the manifold learning methods in transfer learning and apply them in the federated setting on the cross-subject and cross-session tasks. These works only focus on the data distribution heterogeneity issue limited to one dataset. Besides, privacy protection between subjects in one dataset is not a critical issue. EEG data is not as intuitive as images. People can understand the information in the image when they see it, but EEG information needs a professional device to collect and methods to decode. Thus, the ability to interpret and protect EEG data privacy is limited exclusively to research institutes rather than individuals. Thus, EEG data privacy protection for research institutes is more practical than the individuals.

3. Proposed methods

Leveraging multiple device-heterogeneous datasets presents a potential solution for expanding the training set for high-performance model training. However, effectively utilizing such datasets remains an unsolved but critical problem. In this section, we first describe this problem and then present our proposed solution.

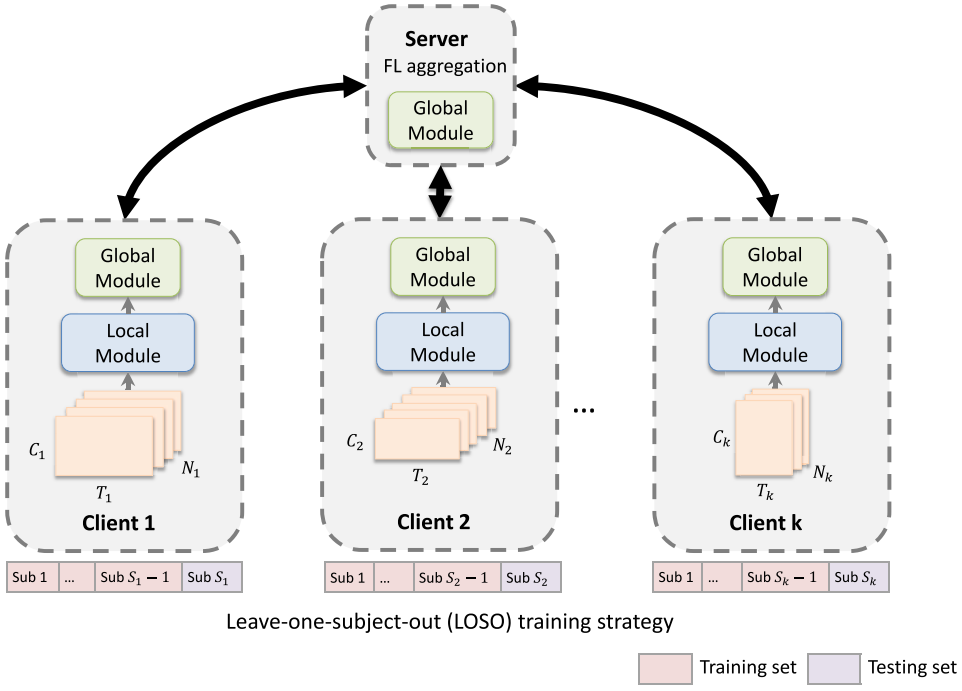


Fig. 1. The overview of the proposed hierarchical personalized federated learning framework. Each client is assigned one dataset with various formats defined by C_k, T_k, and $N_k, k \in \{1, 2, \ldots, K\}$. Clients train their local personalized model, including a local module and a global module, for the classification task. The server manages the cooperation between clients. Each dataset applies the LOSO strategy simultaneously.

3.1. Problem description

As illustrated in Section 2.1 that the device-heterogeneity issue in EEG model training leads to the small amount of available data and the low test accuracy of the trained model. There are two kinds of entities involved: a server S and K distributed clients (*i.e.*, EEG data collection devices). Each client possesses a dataset $D_k = \{\mathbf{x}_k, \mathbf{y}_k\}, k \in \{1, 2, \ldots, K\}$, where $\mathbf{x}_k \in \mathbb{R}^{C_k \times T_k \times N_k}$ represents EEG recordings and $\mathbf{y}_k \in \{0, 1\}^{N_k}$ indicates the corresponding labels. C_k, T_k, and N_k represent the number of channels, the number of time steps, and the number of samples or trials, respectively. Each dataset D_k has S_k subjects. The goal is to take full advantage of data from various devices to train high-performance models.

3.2. Overall framework

To utilize the device-heterogeneous datasets, we propose a personalized federated learning EEG decoding framework, named FLEEG. Following the classical FL framework, the proposed framework consists of one central server and several clients, as illustrated in Fig. 1. Each client is assigned one dataset and processes the dataset with its local personalized model for the classification task. The server manages the cooperation between clients. Note that the "*personalized model*" in this work indicates the specific network structures designed for the assigned datasets, instead of models trained for different subjects.

The personalized model in the client consists of a local module and a global module. The local module acts as a feature encoder to extract the embedding features from EEG data and map them into a unified format across clients. The global module is designed to transfer knowledge between clients by communicating the model weights of global modules in all clients via the FL aggregation in the server. The proposed framework makes each client train its personalized model not only using its own data but also employing knowledge transferred from other datasets, which tremendously enlarges its training set to get better performance. Next, we introduce the detailed designs of the proposed framework.

3.3. Personalized model in the clients

Clients train personalized local models on their corresponding datasets. As illustrated in Fig. 2, the input EEG is first processed by the local module to extract embedding features. Then, the extracted features are sent to the global module to get the prediction as the output. Inspired by the DeepConvNets (DCN) (Schirrmeister et al., 2017), we design the local module with a convolution-based temporal filter and a convolution-based spatial filter followed by two standard convolution-max-pooling layers to extract spatial–temporal information from EEG data, refer to the "*Local Module*" part in Fig. 2. To match the dataset formats, the network structure design of the local module is personalized to its dataset. By setting a suitable kernel size based on the format of input data, the extracted features from heterogeneous datasets can be unified. The detailed settings will be introduced in Section 4.1.2. The global module is designed with one standard convolution-max-pooling layer for high-level feature extraction, followed by a convolution-softmax layer for final classification, refer to the "*Global Module*" part in Fig. 2.

Since we evaluate the proposed algorithm on a classification task, the local model training is guided by the mean cross-entropy loss $\mathcal{L}_k$ on dataset D_k, which is defined as follows:

$$\mathcal{L}_k(D_k, \theta_k^g, \theta_k^l) = \frac{1}{N_k} \sum_{(\mathbf{x}_k^i, y_k^i) \in D_k} \Big[-y_k^i \log\Big(g(l_k(\mathbf{x}_k^i, \theta_k^l), \theta_k^g)\Big) + (1 - y_k^i) \log\Big(1 - g(l_k(\mathbf{x}_k^i, \theta_k^l), \theta_k^g)\Big) \Big] \tag{1}$$

where $l_k(\cdot, \theta_k^l)$ and $g(\cdot, \theta_k^g)$ describe the local module with its model weights θ_k^l and the global module with corresponding model weights θ_k^g, respectively. $\mathbf{x}_k^i$ and y_k^i represents the ith sample and its label in the dataset D_k. N_k indicates the number of samples in the dataset D_k.

3.4. FL aggregation in the server

Inspired by *FedAvg* (McMahan, Moore, Ramage, Hampson, & y Arcas, 2017), the entire framework is trained with the following overall

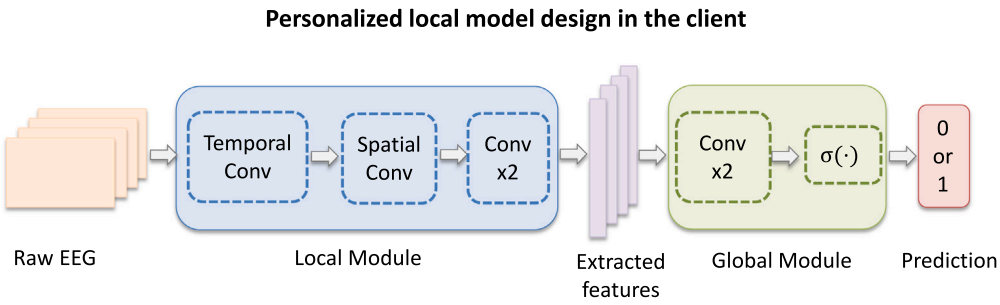


Fig. 2. A local personalized model with local and global modules in one client.

loss function:

$$\mathcal{L}(\theta^g, \theta^l) = \sum_{k=1}^{K} \frac{N_k}{N} \cdot \mathcal{L}_k(D_k, \theta_k^g, \theta_k^l) \tag{2}$$

where N_k denotes the number of samples in client k and N represents the total number of samples in all clients. $\mathcal{L}_k(D_k, \theta_k^g, \theta_k^l)$ is the local loss function in client k as illustrated in Section 3.3. Once the local personalized model finishes local training for the current round, the model weights in the global modules are uploaded to the server for the FL aggregation to transfer knowledge between datasets. The server updates the model weights of global modules in the server as below:

$$\theta^g = \sum_{k=1}^{K} \frac{N_k}{N} \cdot \theta_k^g \tag{3}$$

After the global module in the server is updated, the server distributes the updated model parameters to the clients. Combined with their local modules, clients continue to train the personalized model with their datasets for the next round. The training process stops until the whole system converges.

3.5. Training process

The overall training process of our framework is presented in Algorithm 1, which follows a classical FL system training process:

1. Each client trains its personalized local model with its individual dataset separately for E local epochs.
2. Each client uploads global module model weights to the server for FL aggregation.
3. The server updates model weights of global modules from all clients with FL aggregation, as described in function (3).
4. The server sends back the updated global module to all clients.
5. Each client updates their local model with the updated global module and continues local training, following step 1, for the next round.

The clients and the server repeat steps 1 to 5 for R rounds. Each client selects the model with the smallest validation loss as its final model for testing with its dataset.

Algorithm 1: FLEEG

Input: local training data from K clients, the number of round R, the number of local epochs E, and the learning rate η

Server executes :

initialize global modules with weights θ^g;
initialize K local modules with weights θ_k^l;
for *each round* $t = 1, \dots R$ **do**
  for *each client* k **in parallel do**
    $\theta_k^{g^{(t+1)}} \leftarrow ClientUpdate(k, \theta^{g^{(t)}})$;
  $\theta^{g^{(t+1)}} = \sum_{k=1}^{K} \frac{N_k}{N} \cdot \theta_k^{g^{(t+1)}}$; // aggregate updates

ClientUpdate(k, θ^g)**:**

for *each local epoch* $e = 1,\dots,E$ **do**
  for *each batch* $B \subset D_k$ **do**
    $\theta_k^l \leftarrow \theta_k^l - \eta \cdot \nabla_{\theta_k^l} \mathcal{L}_k(B, \theta_k^g, \theta_k^l)$; // update local module
    $\theta_k^g \leftarrow \theta_k^g - \eta \cdot \nabla_{\theta_k^g} \mathcal{L}_k(B, \theta_k^g, \theta_k^l)$; // update global module

4. Evaluations

In this section, we begin by introducing the experimental setup. Subsequently, we assess the efficacy of FLEEG through comparative analysis with baseline models, utilizing diverse EEG datasets.

4.1. Experiment settings

We evaluate the proposed FL framework using nine EEG MI datasets. We begin by introducing the experiment settings, including the selection of datasets, network structure settings, evaluation strategies, and the baseline model used for comparison.

4.1.1. Datasets

The objective of this work is to transfer knowledge across heterogeneous datasets with the same task but different data formats, including the number of subjects, channels, and sampling frequencies. According to this assumption, we select nine public EEG datasets for this study: the Korea University (KU) MI dataset (Lee et al., 2019), the Shanghai University (SHU) MI dataset (Ma et al., 2022), the Shin2017A dataset (Shin et al., 2016), the BCI-IV-2a dataset (Tangermann et al., 2012), the Weibo2014 (Yi et al., 2014), the MunichMI (Grosse-Wentrup, Liefhold, Gramann, & Buss, 2009), the High-Gamma Dataset (HGD) (Schirrmeister et al., 2017), the Cho2017 (Cho, Ahn, Ahn, Kwon, & Jun, 2017), and the Murat2018 (Kaya, Binli, Ozbay, Yanar, & Mishchenko, 2018) dataset. These datasets all focus on the hands' motor imagery task to classify subjects' imagery movements of their hands. All of them contain the left-hand and right-hand motor imagery classes. The statistical information of these datasets is presented in Table 1 with the number of subjects, the number of trials per subject, the total amount of trials, the number of channels, and sampling frequencies. It should be noted that some datasets have more than two classes (e.g. the BCI-IV-2a dataset has four categories including left hand, right hand, feet, and tongue), but this work only uses data related to the left and right hand.

The EEG data is band-pass filtered between 0.3 Hz and 40 Hz. To save the usage of RAM, we downsample the KU dataset from 1000 Hz to 250 Hz and the Shin2017A dataset from 1000 Hz to 200 Hz.

4.1.2. Local models

As illustrated in Section 3.3, we employed the DCN model as the backbone in the clients of the proposed framework. The network structures of the local modules were designed to accommodate different data

Table 1
Statistic information of the nine MI EEG datasets.

No.	Dataset	#Subjects	#Trials/sub	#Trials	#Channels	f (Hz)
1	KU	54	400	21 600	62	1000
2	SHU	25	500	12 500	32	250
3	Shin2017A	29	60	1740	30	1000
4	BCI-IV-2a	9	288	2592	22	250
5	Weibo2014	10	158	1580	60	200
6	MunichMI	10	300	3000	128	250
7	High-Gamma Dataset (HGD)	14	482	6742	128	500
8	Cho2017	52	190	9880	64	512
9	Murat2018	11	1593	17 515	22	200

formats, as outlined in Table 6. On the other hand, the model designs of the global modules remained consistent across all clients and the server, referring to Table 7.

4.1.3. Evaluation settings

We adopt a subject-independent setting for the cross-dataset training task and utilize the leave-one-subject-out (LOSO) strategy for evaluation. LOSO leaves one subject as the test set. The rest subjects are partitioned into a training set and a validation set to train one model. Once the model is well-trained, it is tested on the left subject to obtain classification accuracy. The overall performance is determined by averaging these accuracy values across all subjects.

Since our evaluation involves nine datasets comprising a total of 214 subjects, strictly following the LOSO approach would be extremely time-consuming. Consequently, we propose an approximate version of the LOSO evaluation strategy for the federated learning framework to enhance training efficiency. Each dataset applies the LOSO methodology independently but simultaneously, as illustrated in Fig. 1. Compared to the strict LOSO approach, the modified version leaves out the data of nine subjects – one from each dataset – as nine test sets for nine datasets correspondingly. However, due to variations in the number of subjects across datasets, subjects from datasets with fewer individuals will undergo the LOSO multiple times. This repetition is aimed at facilitating the training for datasets with more subjects. Ultimately, the final classification results for these repeated subjects are obtained by averaging the values.

In the experiment, the partition of the training set and the validation set follows a trial-wise way with a ratio of 9:1. We set the maximum training round R to 250 and local epoch E to 1. Adam optimizer is applied in the training process. Once the model training finishes, the model with the smallest validation loss is selected as the best model and applied to the test set to get classification accuracy for the subject. The final result for the target client is the averaged classification accuracy of all subjects. Batch sizes B and learning rates η for different datasets are set differently, as shown in Table 2.

4.1.4. Comparison baselines

Given that no prior research has addressed this issue in BCI field, we compare the proposed algorithm with two baseline approaches that train the models independently using a single dataset with a subject-independent setting and a subject-dependent setting. To ensure a fair comparison, we maintain consistency between the network structures, maximum training rounds, learning rates, and batch sizes for both the proposed algorithm and the baseline approaches. Regarding the training strategy, we employ a LOSO approach in the subject-independent setting. In the subject-dependent scenario, when the datasets encompass multiple sessions, we adopt a Leave-One-Session-Out strategy to partition the data into training (including the validation set) and test sets. In cases where the datasets consist of only one session, we perform trial-wise partitioning of the training and test data using a 5-fold cross-validation strategy. The network structures in the baseline setting are the same as the local personalized model structures (including the local module and the global module) in the clients for the corresponding datasets in the proposed framework.

Table 2
The learning rates and batch sizes used in the training of each dataset.

Dataset	Learning rate	Batch size
KU	0.01	512
SHU	0.01	512
Shin2017A	0.001	512
BCI-IV-2a	0.005	512
Weibo2014	0.05	512
MunichMI	0.01	128
HGD	0.01	128
Cho2017	0.05	512
Murat2018	0.01	512

4.2. Experiment results

In this section, we present the results of the proposed framework on nine distinct datasets, compared with two baseline approaches: one for subject-dependent settings, denoted as "Sub-dep. DCN", and the other for subject-independent settings, denoted as "Sub-ind. DCN".

Table 3 presents the average classification accuracies along with their corresponding standard deviation values for our method and the baseline models on each dataset. The datasets are arranged in ascending order based on the improvements FLEEG exhibits over the subject-independent baseline results. The enhancement values in comparison to the two baseline methods are listed in the "Δ_{sd}" and "Δ_{si}" columns.

According to Table 3, it is evident that the proposed framework has led to substantial enhancements in the classification accuracy of multiple datasets. In comparison to the subject-dependent setting baseline, FLEEG consistently improves accuracy across all datasets. In the subject-independent setting, most datasets exhibit notable improvements. Specifically, the Shin2017A, Weibo2014, BCI-IV-2a, MunichMI, Murat2018, KU, and Cho2017 datasets have shown increases of 5.35, 4.29, 2.63, 1.63, 1.18, 0.60, and 0.01 in their classification accuracies respectively. On the other hand, the SHU and HGD datasets did not exhibit noticeable improvements but rather maintained similar performance levels.

To present the results in a more visually intuitive manner, we have graphed the data from Table 3 in Fig. 3. It also includes a plot illustrating the number of trials for each dataset. By examining the classification accuracy plot in conjunction with the number of trials plot, we can observe that as the number of trials increases, the improvement achieved by the FLEEG becomes less pronounced. This observation suggests that the proposed framework effectively assists datasets, particularly smaller ones, in leveraging information from datasets with different data formats to enhance their model training and achieve better performance.

SHU and HGD datasets exhibit similar performance levels between FLEEG and baselines. One possible reason for the similar performance is that the number of trials of these datasets is large enough. Besides, for HGD datasets, the similar performances may also be due to its high baseline performance, which already exceed 80%. Thus, it cannot benefit significantly from other lower-quality datasets. As for the SHU dataset, it may be attributed to the significant disparity between the

Table 3
The accuracy (%) results of the proposed framework and baselines.

Dataset	Sub-dep. DCN	Sub-ind. DCN	FLEEG	Δ_{sd}	Δ_{si}
Shin2017A	51.95 ± 0.06	63.79 ± 10.27	**69.14** ± 12.66	+17.19	+5.35
Weibo2014	53.00 ± 0.04	63.03 ± 8.14	**67.32** ± 9.44	+14.32	+4.29
BCI-IV-2a	58.95 ± 0.10	80.63 ± 7.27	**83.26** ± 7.84	+24.31	+2.63
MunichMI	60.07 ± 0.07	73.43 ± 12.04	**75.06** ± 11.78	+14.99	+1.63
Murat2018	79.61 ± 0.12	81.19 ± 8.37	**82.37** ± 6.89	+2.76	+1.18
KU	59.79 ± 0.11	84.82 ± 9.25	**85.42** ± 9.50	+25.63	+0.60
Cho2017	64.84 ± 0.14	76.98 ± 12.06	**76.99** ± 11.8	+12.15	+0.01
SHU	53.86 ± 0.05	**60.60** ± 9.33	60.37 ± 8.17	+6.51	−0.23
HGD	83.38 ± 0.11	**87.59** ± 7.67	87.21 ± 7.44	+3.83	−0.38

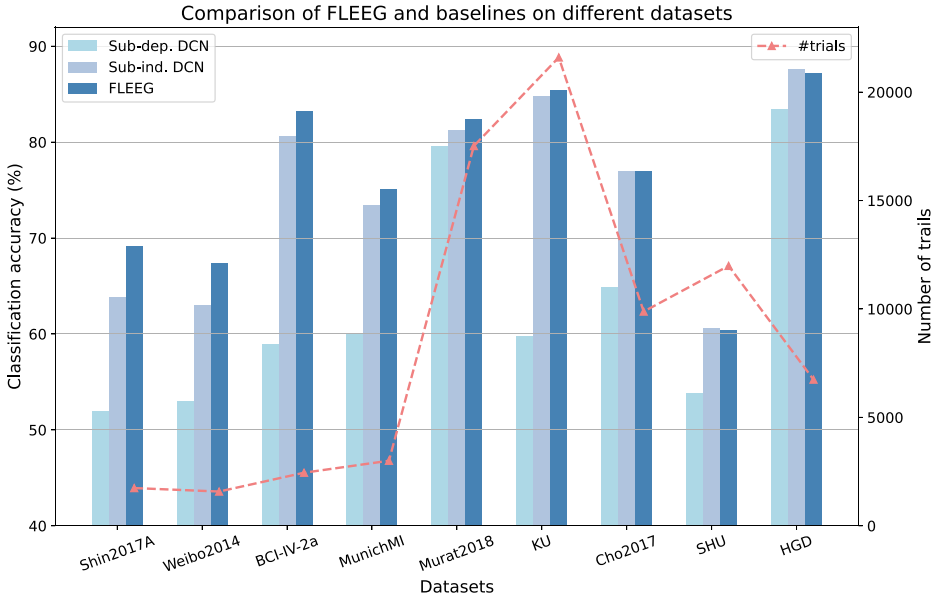


Fig. 3. The proposed algorithm and the baselines were compared on nine MI EEG datasets, with the classification accuracy results presented in ascending order based on their respective improvements. The number of trials for each dataset is also shown. The proposed framework effectively supports datasets, especially those that are smaller.

data distribution of the SHU dataset and the others, as evidenced by its lowest baseline performance among the nine datasets, indicating a large disparity.

5. Discussions

In this section, we proceed to examine the factors influencing the performance of FLEEG, followed by a visual representation illustrating the improvements achieved by the proposed framework.

5.1. Sensitivity analysis

In this section, we delve into an examination of the factors that impact the performance of FLEEG. Specifically, we explore the influence of dataset sizes and model training hyperparameters.

5.1.1. Number of trails per subject matters

Based on the results shown in Fig. 3, it indicates that the performance of FLEEG is significantly affected by the number of trials. The number of trials is determined by two factors — the number of subjects and the number of trials collected from each subject. To gain a better understanding of how these factors impact the proposed framework, we plot a bubble chart for all datasets in Fig. 4. The x-axis represents the number of trials per subject, while the y-axis shows the number of subjects in the datasets. The size of each bubble reflects the absolute value of the "*Improvement*" metric, which is defined as $(Acc_{FLEEG} - Acc_{Baseline})/Acc_{Baseline}$. It indicates the performance of FLEEG: the larger the better for blue bubbles representing positive changes, and the smaller the better for red bubbles representing negative changes. The numerical values corresponding to Fig. 4 are provided in Table 4a and Table 4b. Table 4a lists the "*Improvement*" values of all datasets in ascending order of their corresponding numbers of trials per subject. Similarly, Table 4b lists the total number of subjects in all datasets in ascending order.

In Fig. 4, blue bubbles are mainly located on the left side of the x-axis while red bubbles are located on the right side. It indicates that datasets with fewer trials per subject can benefit more from the proposed framework. For instance, the Shin2017A dataset, with 60 trials collected for each subject, shows an 8.4% improvement with FLEEG. Meanwhile, the HGD and SHU dataset, with more than 400 trials per subject, rarely gains any improvement from the system. Similarly, for a fixed value on the x-axis, the bubbles at the bottom of the y-axis are larger than the ones at the top. This suggests that datasets with fewer subjects can benefit more from the proposed framework, given a similar number of trials per subject. For example, the Weibo2014 and Cho2017 datasets have around 150–200 trials per subject. But the improvement on the Weibo2014 dataset, with 10 subjects, reaches 6.81%, compared to only 0.01% on the Cho2017 dataset, which has 52 subjects.

Comparing these two factors, the number of trials per subject is more important. Even if a dataset involves many subjects, FLEEG can still improve the performance if the number of trials collected from one subject is small, such as Shin2017A dataset. Therefore, the number of trials per subject is the primary factor, while the number of subjects is a secondary factor in determining the FLEEG performance. Thus, the proposed framework can be applied to small datasets to train high-performance models. Additionally, with the help of the FLEEG framework, the model can achieve good results trained with a small number of trials collected from one subject.

Table 4
Improvements of the FLEEG with each MI dataset.

(a) Improvements on nine datasets in ascending order of the number of trials per subject.

Dataset	#trials/sub	Improvement
Shin2017A	60	**8.39%**
Weibo2014	158	**6.81%**
Cho2017	190	**0.01%**
BCI-IV-2a	288	**3.26%**
MunichMI	300	**2.22%**
KU	400	**0.71%**
HGD	482	−0.43%
SHU	500	−0.38%
Murat2018	1593	**1.45%**

(b) Improvements on nine datasets in ascending order of the number of subjects.

Dataset	#sub	Improvement
BCI-IV-2a	9	**3.26%**
Weibo2014	10	**6.81%**
MunichMI	10	**2.22%**
Murat2018	11	**1.45%**
HGD	14	−0.43%
SHU	25	−0.38%
Shin2017A	29	**8.39%**
Cho2017	52	**0.01%**
KU	54	**0.71%**

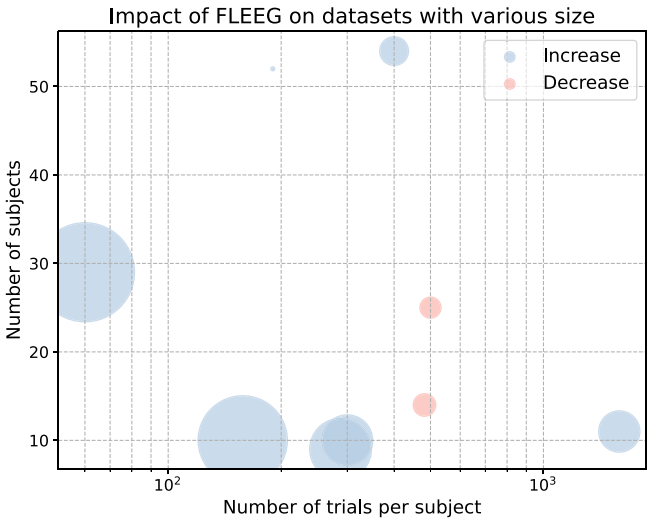


Fig. 4. The impact of the framework on the datasets with various numbers of subjects and numbers of trials per subject. The bubble color indicates the performance of FLEEG with the blue bubbles illustrating an increase and the red bubbles representing a decrease. The bubble size reflects the absolute values of changes. The number of trials per subject is the primary factor that affects FLEEG performance, with the number of subjects being a secondary factor.

5.1.2. *Hyperparameter sensitivity analysis*

In this section, we evaluate hyperparameter sensitivity, specifically focusing on the learning rate for each dataset. The proposed framework operates under the assumption of client isolation, meaning clients do not need to consider other clients' circumstances. During the training process, they solely focus on their local training procedure and share model weights with the server. Consequently, adjusting hyperparameters for FLEEG system is not necessary. Instead, we tune learning rates independently for each dataset. We experiment with learning rates of $0.05, 0.01, 0.005$, and 0.001 for the subject-independent baseline setting with local personalized models, as stated in Section 4.1.2, presenting the resulting accuracies in Table 5 for each dataset.

According to the results in Table 5, we select 0.05 as the learning rate for Weibo2014 and Cho2017 dataset, 0.01 as learning rate for KU, SHU, MunichMI, HGD, and Murat2018 dataset, 0.005 as learning rate for BCI-IV-2a dataset, and 0.001 as learning rate for Shin2017A dataset. These learning rates are applied to the corresponding local model training in FLEEG and the subject-dependent setting as well.

5.2. *Interpretability and visualization*

In this section, we employ saliency maps (Simonyan, Vedaldi, & Zisserman, 2014) to visualize the informative regions within the data.

Table 5
The accuracy (%) results of subject-independent baseline with different learning rates.

Dataset	Learning rate			
	0.05	0.01	0.005	0.001
KU	83.37	**84.82**	84.63	84.75
SHU	59.14	**60.60**	59.75	59.67
Shin2017A	62.70	60.11	61.09	**63.79**
BCI-IV-2a	60.60	80.32	**80.63**	80.32
Weibo2014	**63.03**	61.29	59.96	59.54
MunichMI	73.03	**73.43**	73.17	72.33
HGD	87.32	**87.59**	87.30	85.36
Cho2017	**76.98**	76.48	76.29	76.84
Murat2018	80.85	**81.19**	80.85	80.35

For enhanced visualization, the original saliency map is averaged across the time dimension, resulting in each subject's topological map of the EEG channels.

We plot the saliency maps of the Shin2017A and HGD datasets. Shin2017A gains the largest improvement from the framework, meanwhile, HGD has the largest decrease. We also plot the individual accuracy comparison between FLEEG and the subject-independent baselines for each subject in these two datasets. For Shin2017A dataset, the accuracy comparison for each subject is presented in Fig. 5 with a descending sequence of accuracy improvement. Due to the space limitation, we only plot the saliency map of subjects with the top 5 improvements and bottom 5 improvements in Figs. 6 and 7, correspondingly. For HGD dataset, the accuracy comparison for each subject is presented in Fig. 8, sorted with a descending sequence of accuracy improvement. Due to the space limitation, we also plot the saliency map of the top 5 and bottom 5 subjects in Figs. 9 and 10, correspondingly. The detailed classification accuracy comparison results for the rest of the datasets are presented in Fig. 11.

Compared with the baseline, the proposed framework can stably catch the features from the most informative areas related to the motor cortex regions in the brain, even for small datasets. According to Fig. 6, F8, and P7 contribute more to the predictions of the baseline method, whereas, CCP5h, CCP3h, Cz, CCP4h, and CCP6h provide more information to FLEEG. This indicates that FLEEG learns neurophysiologically meaningful features from the EEG signals originating from motor cortex regions (Pfurtscheller & Neuper, 1997). Although the improvements are relatively lower for the bottom 5 subjects shown in Fig. 7, FLEEG stably learns from motor cortex regions (CCP5h and CCP3h) compared to the baseline method that focuses on non-motor areas for some subjects, i.e., subject 2 and subject 26. As the number of samples increases, both FLEEG and the baseline method concentrate more on EEG from the motor-related areas. As shown in Figs. 9 and 10, both methods focus on CCP3h, CCP4h, and C2 which are located in the motor area of the brain when they are trained on the HGD dataset which has more data samples than Shin2017A.

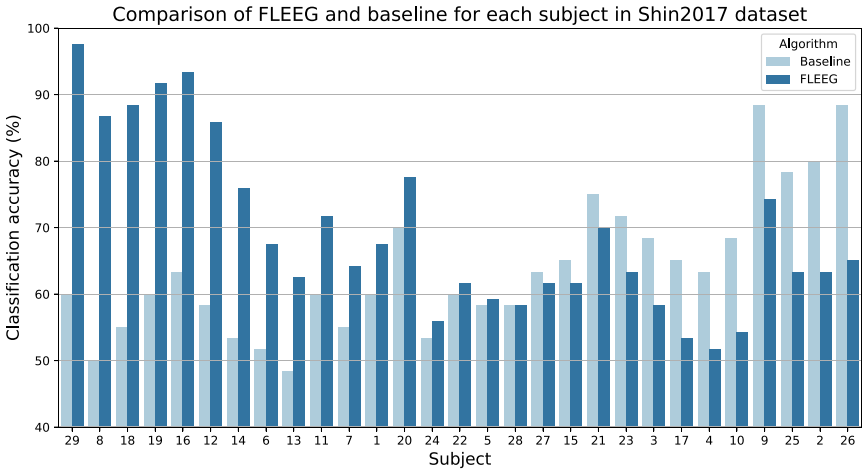
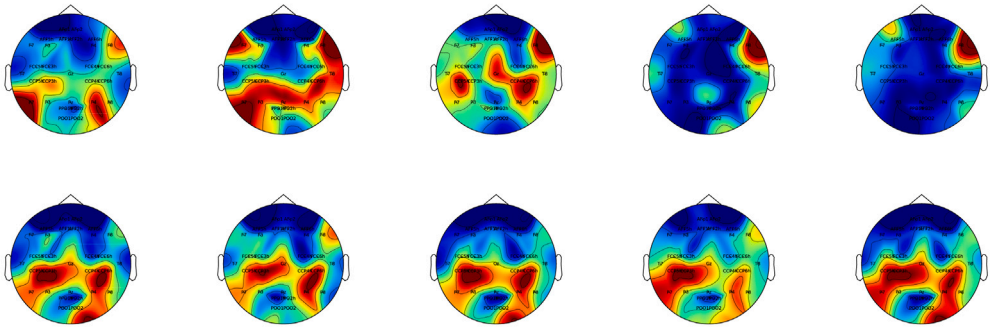
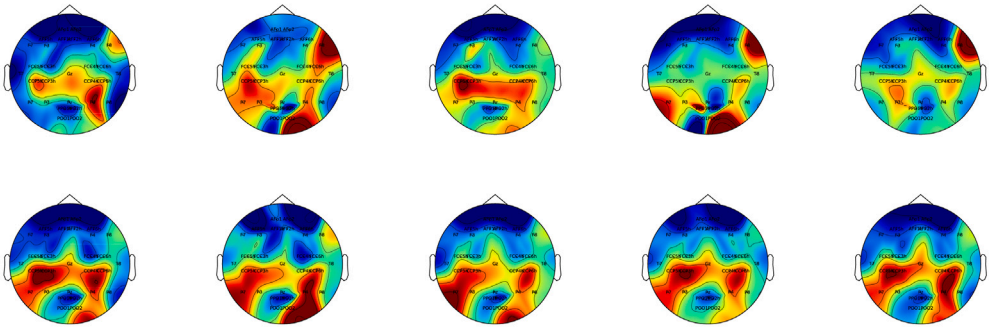


Fig. 5. The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the Shin2017A MI EEG datasets.



(a) Subject 29 (b) Subject 8 (c) Subject 18 (d) Subject 19 (e) Subject 16

Fig. 6. The saliency maps for the subjects with the top 5 improvements in the Shin2017A dataset. The first row presents the plots for the subject-independent baseline method and the second row lists the maps for FLEEG.



(a) Subject 10 (b) Subject 9 (c) Subject 25 (d) Subject 2 (e) Subject 26

Fig. 7. The saliency maps for the subjects with the bottom 5 improvements in the Shin2017A dataset. The first row presents the plots for the subject-independent baseline method and the second row lists the maps for FLEEG.

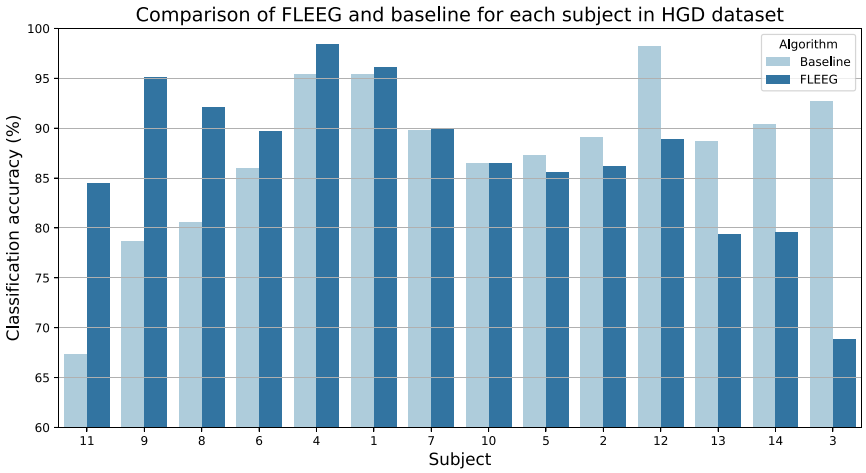
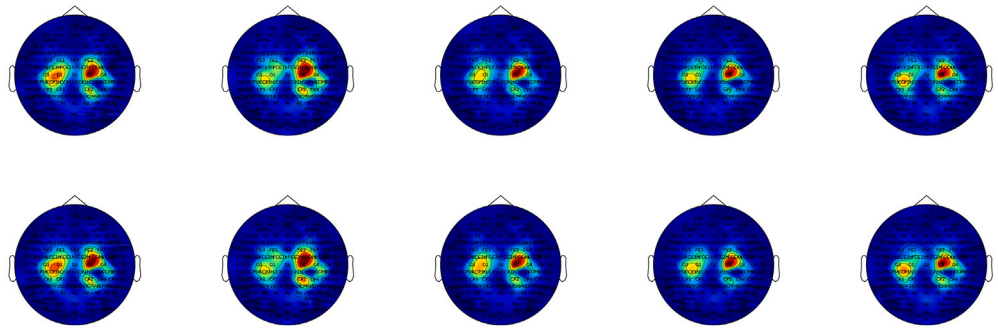
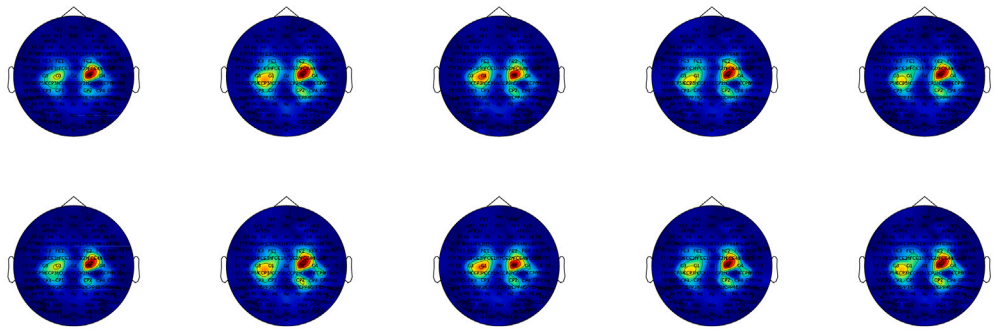


Fig. 8. The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the HGD MI EEG datasets.



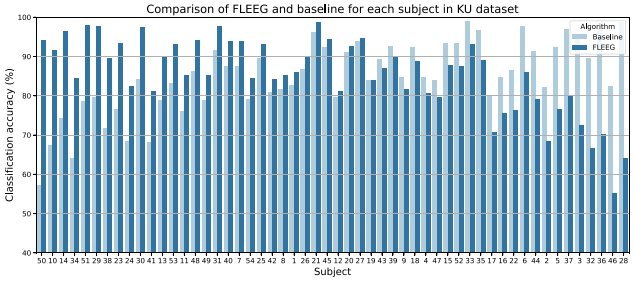
(a) Subject 11 (b) Subject 9 (c) Subject 8 (d) Subject 6 (e) Subject 4

Fig. 9. The saliency maps for the subjects with the top 5 improvements in the HGD dataset. The first row presents the plots for the subject-independent baseline method and the second row lists the maps for FLEEG.

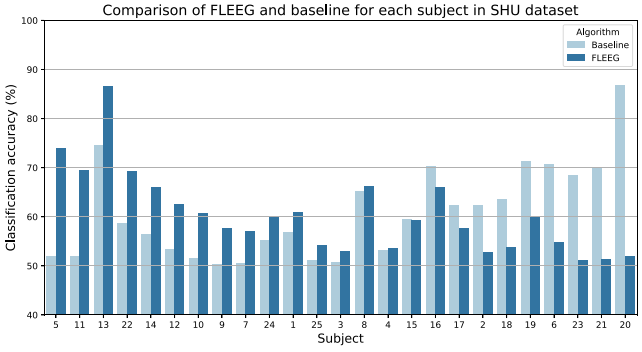


(a) Subject 2 (b) Subject 12 (c) Subject 13 (d) Subject 14 (e) Subject 3

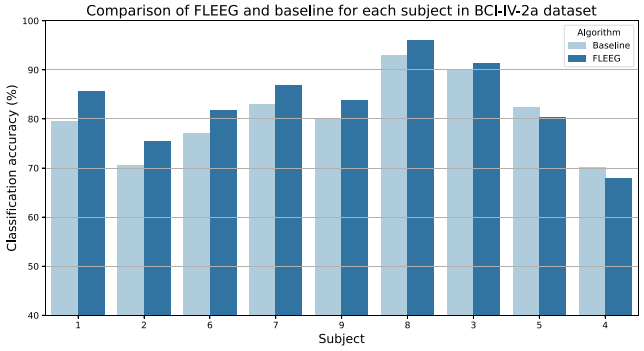
Fig. 10. The saliency maps for the subjects with the bottom 5 improvements in the HGD dataset. The first row presents the plots for the subject-independent baseline method and the second row lists the maps for FLEEG.



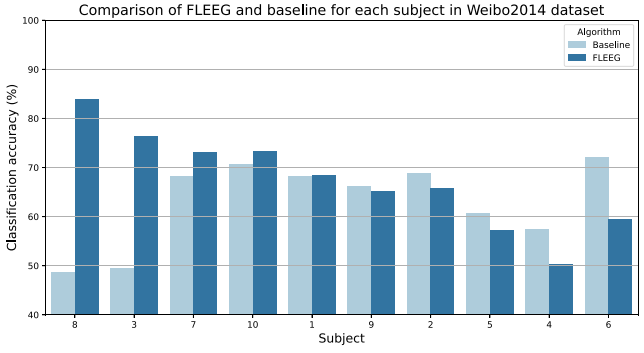
(a) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the KU dataset.



(b) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the SHU dataset.

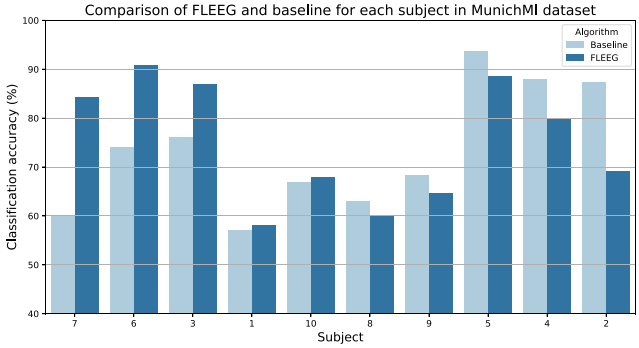


(c) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the BCI-IV-2a dataset.

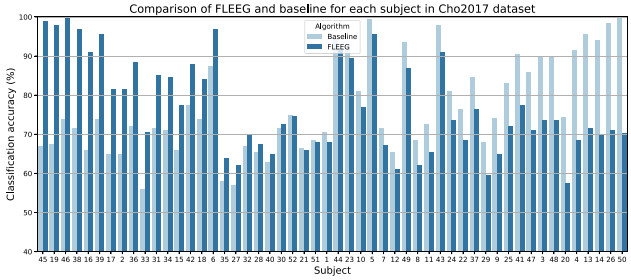


(d) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the Weibo2014 dataset.

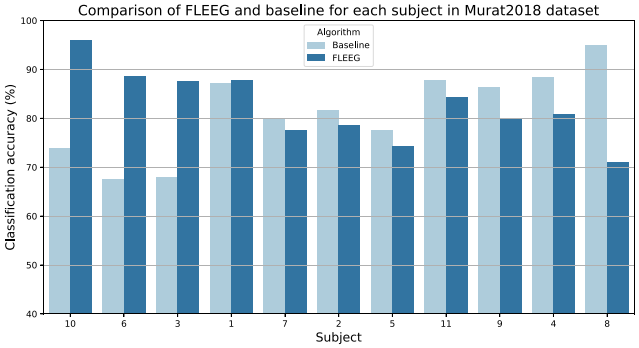
Fig. 11. The accuracy results of each subject with subject-independent baseline method and FLEEG for the KU, SHU, BCI-IV-2a, Weibo2014, MunichMI, Cho2017, and Murat2018 datasets.



(e) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the MunichMI dataset.



(f) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the Cho2017 dataset.



(g) The accuracy comparison of each subject with the subject-independent baseline method and FLEEG on the Murat2018 dataset.

Fig. 11. (*continued*).

Table 6
The network structure of the local modules for each dataset.

Dataset	Input size	Conv-temp		Conv-spatial			Conv-pool			Conv-pool			Output size
		num. of ker.	ker. size	num. of ker.	ker. size	pool. ker. size	num. of ker.	ker. size	pool. ker. size	num. of ker.	ker. size	pool. ker. size	
KU	[1, 62, 1000]	25	(1, 10)	25	(62, 1)	(1, 3)	50	(1, 10)	(1, 3)	100	(1, 10)	(1, 3)	[100, 1, 32]
SHU	[1, 32, 1000]	25	(1, 10)	25	(32, 1)	(1, 3)	50	(1, 10)	(1, 3)	100	(1, 10)	(1, 3)	[100, 1, 32]
Shin2017A	[1, 30, 2000]	25	(1, 8)	25	(30, 1)	(1, 5)	50	(1, 8)	(1, 4)	100	(1, 8)	(1, 3)	[100, 1, 30]
BCI-IV-2a	[1, 22, 1000]	25	(1, 10)	25	(22, 1)	(1, 3)	50	(1, 10)	(1, 3)	100	(1, 10)	(1, 3)	[100, 1, 32]
Weibo2014	[1, 60, 800]	25	(1, 8)	25	(60, 1)	(1, 2)	50	(1, 8)	(1, 3)	100	(1, 8)	(1, 3)	[100, 1, 30]
MunichMI	[1, 128, 3500]	25	(1, 10)	25	(128, 1)	(1, 4)	50	(1, 10)	(1, 4)	100	(1, 10)	(1, 3)	[100, 1, 32]
HGD	[1, 128, 2000]	25	(1, 20)	25	(128, 1)	(1, 6)	50	(1, 20)	(1, 3)	100	(1, 10)	(1, 3)	[100, 1, 31]
Cho2017	[1, 64, 1536]	25	(1, 22)	25	(64, 1)	(1, 4)	50	(1, 22)	(1, 3)	100	(1, 22)	(1, 3)	[100, 1, 32]
Murat2018	[1, 22, 200]	25	(1, 6)	25	(22, 1)	(1, 1)	50	(1, 6)	(1, 2)	100	(1, 6)	(1, 3)	[100, 1, 30]

Table 7
The network structure of the global module in the proposed algorithm.

Dataset	Input size	Conv-pool			Conv-pool		Softmax	Output size
		num. of ker.	ker. size	pool. ker. size	num. of ker.	num. of ker.		
KU	[100, 1, 32]							
SHU	[100, 1, 32]							
Shin2017A	[100, 1, 30]							
BCI-IV-2a	[100, 1, 32]							
Weibo2014	[100, 1, 30]	200	(1, 10)	(1, 3)	2	(1, 7)		[2, 1, 1]
MunichMI	[100, 1, 32]							
HGD	[100, 1, 31]							
Cho2017	[100, 1, 32]							
Murat2018	[100, 1, 30]							

6. Conclusions and future work

In this work, we proposed a new learning paradigm for BCI to train the high-performance EEG decoding model with multiple datasets. We designed a hierarchical personalized federated-learning-based framework FLEEG to solve the device-heterogeneity issue among multiple EEG datasets, enabling knowledge sharing between datasets. The proposed framework has been evaluated with nine real MI datasets and obtained promising results with reasonable interpretations. This framework overcomes the challenge of insufficient data for model training in BCI. Thus, small datasets can train better models by making use of the knowledge from other datasets with the help of FLEEG.

In the future, it is interesting to apply the proposed framework to more complex situations where the datasets have various protocols and tasks or improve the local module with more powerful feature encoders. Furthermore, how to measure the quality of datasets in each client and control their effects on the overall performance also deserves further investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used the in the work are public datasets.

Acknowledgments

This research/project is supported by the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund, Singapore (No. A20G8b0102); and the National Research Foundation, Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019).

References

Bakas, S., Ludwig, S., Barmpas, K., Bahri, M., Panagakis, Y., Laskaris, N., et al. (2022). Team cogitat at neurips 2021: Benchmarks for EEG transfer learning competition. arXiv preprint arXiv:2202.03267.

Bayram, H. C., & Rekik, I. (2021). A federated multigraph integration approach for connectional brain template learning. In *International workshop on multimodal learning for clinical decision support* (pp. 36–47). Springer.

Bica, I., & van der Schaar, M. (2022). Transfer learning on heterogeneous feature spaces for treatment effects estimation. *Advances in Neural Information Processing Systems*, *35*, 37184–37198.

Cho, H., Ahn, M., Ahn, S., Kwon, M., & Jun, S. C. (2017). EEG datasets for motor imagery brain–computer interface. *GigaScience*, *6*(7), gix034.

Cui, J., Yuan, L., Li, R., Wang, Z., Yang, D., & Jiang, T. (2023). Benchmarking EEG-based cross-dataset driver drowsiness recognition with deep transfer learning. *EMBC 2023*.

Ding, Y., Robinson, N., Tong, C., Zeng, Q., & Guan, C. (2023). Lggnet: learning from local-global-graph representations for brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*.

Feng, S., Li, B., Yu, H., Liu, Y., & Yang, Q. (2022). Semi-supervised federated heterogeneous transfer learning. *Knowledge-Based Systems*, *252*, Article 109384.

Gao, D., Yao, X., & Yang, Q. (2022). A survey on heterogeneous federated learning. arXiv preprint arXiv:2210.04505.

Goebel, R., Yu, H., Faltings, B., Fan, L., & Xiong, Z. (Eds.), (2023). *Trustworthy federated learning, vol. 13448* (p. 158). Springer, Cham.

Grosse-Wentrup, M., Liefhold, C., Gramann, K., & Buss, M. (2009). Beamforming in noninvasive brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, *56*(4), 1209–1219.

Gu, X., Cai, W., Gao, M., Jiang, Y., Ning, X., & Qian, P. (2022). Multi-source domain transfer discriminative dictionary learning modeling for electroencephalogram-based emotion recognition. *IEEE Transactions on Computational Social Systems*, *9*(6), 1604–1612.

Hang, W., Li, J., Liang, S., Wu, Y., Lei, B., Qin, J., et al. (2023). FedEEG: Federated EEG decoding via inter-subject structure matching. In *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing* (pp. 1–5). IEEE.

Hu, R., Zhou, X., Xu, Z., Liao, Z., Wu, H., Qu, H., et al. (2021). Cross-subject federated transfer learning with quanvolutional layer for motor imagery classification. In *2021 China automation congress* (pp. 5736–5741). IEEE.

Ju, C., Gao, D., Mane, R., Tan, B., Liu, Y., & Guan, C. (2020). Federated transfer learning for EEG signal classification. In *2020 42nd annual international conference of the ieee engineering in medicine & biology society* (pp. 3040–3045). IEEE.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1–2), 1–210.

Kaya, M., Binli, M. K., Ozbay, E., Yanar, H., & Mishchenko, Y. (2018). A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific Data*, *5*(1), 1–16.

Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, *15*, Article 653659.

Kuang, F., Shu, L., Hua, H., Wu, S., Zhang, L., Xu, X., et al. (2021). Cross-subject and cross-device wearable EEG emotion recognition using frontal EEG under virtual reality scenes. In *2021 IEEE international conference on bioinformatics and biomedicine* (pp. 3630–3637). IEEE.

Lee, T.-S., Goh, S. J. A., Quek, S. Y., Phillips, R., Guan, C., Cheung, Y. B., et al. (2013). A brain-computer interface based cognitive training system for healthy elderly: A randomized control pilot study for usability and preliminary efficacy. *PLoS One*, *8*(11), Article e79419.

Lee, M.-H., Kwon, O.-Y., Kim, Y.-J., Kim, H.-K., Lee, Y.-E., Williamson, J., et al. (2019). EEG dataset and openbmi toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*, *8*(5), giz002.

Li, A., Zhang, L., Wang, J., Han, F., & Li, X.-Y. (2021). Privacy-preserving efficient federated-learning model debugging. *IEEE Transactions on Parallel and Distributed Systems*, *33*(10), 2291–2303.

Liang, P. P., Liu, T., Ziyin, L., Allen, N. B., Auerbach, R. P., Brent, D., et al. (2020). Think locally, act globally: Federated learning with local and global representations. arXiv preprint arXiv:2001.01523.

Ma, J., Yang, B., Qiu, W., Li, Y., Gao, S., & Xia, X. (2022). A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface. *Scientific Data*, *9*(1), 531.

Mane, R., Chouhan, T., & Guan, C. (2020). BCI for stroke rehabilitation: Motor and beyond. *Journal of Neural Engineering*, *17*(4), Article 041001.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273–1282). PMLR.

Nijholt, A., Contreras-Vidal, J. L., Jeunet, C., & Väljamäe, A. (2022). Brain-computer interfaces for non-clinical (home, sports, art, entertainment, education, well-being) applications. *Frontiers in Computer Science*, *4*, Article 860619.

Peng, L., Wang, N., Dvornek, N., Zhu, X., & Li, X. (2022). Fedni: Federated graph learning with network inpainting for population-based disease prediction. *IEEE Transactions on Medical Imaging*.

Pfurtscheller, G., & Neuper, C. (1997). Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*, *239*(2–3), 65–68.

Saeed, A., Grangier, D., Pietquin, O., & Zeghidour, N. (2021). Learning from heterogeneous EEG signals with differentiable channel reordering. In *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing* (pp. 1255–1259). IEEE.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420.

Shin, J., von Lühmann, A., Blankertz, B., Kim, D.-W., Jeong, J., Hwang, H.-J., et al. (2016). Open access dataset for EEG+ NIRS single-trial classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(10), 1735–1745.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: visualising image classification models and saliency maps. In *Workshop at international conference on learning representations*.

Tan, A. Z., Yu, H., Cui, L., & Yang, Q. (2022). Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Frontiers in Neuroscience*, 55.

Wan, Z., Yang, R., Huang, M., Zeng, N., & Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing*, *421*, 1–14.

Wei, X., Ortega, P., & Faisal, A. A. (2021). Inter-subject deep transfer learning for motor imagery EEG decoding. In *2021 10th international IEEE/EMBS conference on neural engineering* (pp. 21–24). IEEE.

Xu, L., Xu, M., Ke, Y., An, X., Liu, S., & Ming, D. (2020). Cross-dataset variability problem in EEG decoding with deep learning. *Frontiers in Human Neuroscience*, *14*, 103.

Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., & Yu, H. (Eds.), (2020). *Federated learning* (p. 189). Springer, Cham.

Yi, W., Qiu, S., Wang, K., Qi, H., Zhang, L., Zhou, P., et al. (2014). Evaluation of EEG oscillatory patterns and cognitive process during simple and compound limb motor imagery. *PLoS One*, *9*(12), Article e114853.

Zhang, K., Xu, G., Zheng, X., Li, H., Zhang, S., Yu, Y., et al. (2020). Application of transfer learning in EEG decoding based on brain-computer interfaces: a review. *Sensors*, *20*(21), 6321.