UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
mixEEG: Enhancing EEG Federated Learning for Cross-subject EEG Classification with Tailored mixup

Permalink
https://escholarship.org/uc/item/0t123879

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Liu, Xuan-Hao
Lu, Bao-Liang
Zheng, Wei-Long

Publication Date
2025

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

*mix*EEG: Enhancing EEG Federated Learning for Cross-subject EEG Classification with Tailored *mixup*

Xuan-Hao Liu (haogram_sjtu@sjtu.edu.cn)

Bao-Liang Lu (bllu@sjtu.edu.cn)

Wei-Long Zheng* (weilong@sjtu.edu.cn)
School of Computer Science, Shanghai Jiao Tong University, Shanghai, China

Abstract

The cross-subject electroencephalography (EEG) classification exhibits great challenges due to the diversity of cognitive processes and physiological structures between different subjects. Modern EEG models are based on neural networks, demanding a large amount of data to achieve high performance and generalizability. However, privacy concerns associated with EEG pose significant limitations to data sharing between different hospitals and institutions, resulting in the lack of large dataset for most EEG tasks. Federated learning (FL) enables multiple decentralized clients to collaboratively train a global model without direct communication of raw data, thus preserving privacy. For the first time, we investigate the cross-subject EEG classification in the FL setting. In this paper, we propose a simple yet effective framework termed ***mix*EEG**. Specifically, we tailor the vanilla *mixup* considering the unique properties of the EEG modality. *mix*EEG shares the unlabeled averaged data of the unseen subject rather than simply sharing raw data under the domain adaptation setting, thus better preserving privacy and offering an averaged label as pseudo-label. Extensive experiments are conducted on an epilepsy detection and an emotion recognition dataset. The experimental result demonstrates that our *mix*EEG enhances the transferability of global model for cross-subject EEG classification consistently across different datasets and model architectures. Code is published at: https://github.com/XuanhaoLiu/mixEEG.

Keywords: EEG; Federated Learning; mixup; Domain Generalization; Domain Adaptation

Introduction

Brain-Computer Interfaces (BCIs) based on electroencephalography (EEG) signals have been significantly facilitating the health care of people in a wide range of areas, encompassing cognitive research (Tan et al., 2024; Finley et al., 2024; X.-H. Liu, Liu, et al., 2024), epilepsy detection (Shoeb & Guttag, 2010; Tasci et al., 2023), and emotion recognition (Aggarwal & Chugh, 2022; Jiang, Liu, Zheng, & Lu, 2023, 2024). A common but challenging practical requirement of EEG BCI is training an EEG BCI that achieves good performance on unseen subjects, which is also known as cross-subject EEG classification. Due to the diversity of cognitive processes and physiological structures, electroencephalogram (EEG) data present significant individual differences. Modern EEG BCIs are largely based on neural networks (NN) and it is widely acknowledged that NN follow the *Scaling Law* (Aghajanyan et al., 2023; Cherti et al., 2023), i.e., the performance of NN improves when scaling up the amount and quality of training data. However, due to the expensive collection costs and privacy concerns (X.-H. Liu, Song, He, Lu, & Zheng, 2024), there is no public EEG dataset that is large enough to train well generalizable EEG BCIs. The majority of EEG data around the world is stored in various medical hospitals or research institutes in the format of small datasets. How can we leverage these dispersed data resources to train more generalizable EEG BCIs?

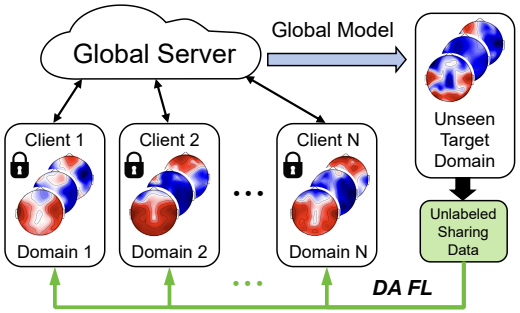

Figure 1: The problem setting of domain generalization federated learning (DG FL) and domain adaptation federated learning (DA FL), which aims to learn a global model from multiple decentralized source domains. The unlabeled sharing data is sent to client only under the DA FL setting.

Federated learning (FL) allows multiple decentralized clients to collaboratively train a global model without direct communication of raw data (R. Liu et al., 2024). For instance, during each communication round of the FedAvg (McMahan, Moore, Ramage, Hampson, & y Arcas, 2017), the clients update its local model exclusively with their own data, and upload the parameters of their local model to the global server, where the global model is then updated through the aggregation of parameters from these local models. Although there are many studies adopted FL in EEG-based BCI (Ju et al., 2020; Baghersalimi, Teijeiro, Atienza, & Aminifar, 2021; X. Chen, Niu, Zhao, & Qin, 2024; Chan, Zheng, Xu, Wang, & Heng, 2024), these studies predominantly focused on improving the performance of the global model on existing subjects of each client, overlooking the practical requirement of cross-subject EEG classification on unseen subjects.

Basically, there are two cross-subject settings: domain generalization (DG) and domain adaptation (DA). While DG requires no data from unseen subjects, DA allows the trainer to access the unlabeled data from unseen subjects. For the

*Corresponding author.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

first time, we investigate these two settings in FL, i.e., the DG FL setting and DA FL setting. Illustrated in Figure 1, these two settings require training a global model on multi-source domains while preserving the data privacy of each source domain. With the strict prohibition of data communication among source domains, the DG FL and DA FL problems present heightened challenges compared to traditional DG and DA settings, in which the global server can directly access to all source domains data.

Inspired by *mixup* (H. Zhang, Cisse, Dauphin, & Lopez-Paz, 2018), a simple yet surprisingly effective data augmentation technique, we propose a novel FL framework called ***mix*EEG** to tackle the two proposed problem settings. Specifically, we tailor the vanilla *mixup* for EEG modality to unleash the potential of *mixup* for developing better cross-subject EEG BCI. Solely through mixing up the local EEG data of each client, our *mix*EEG already enhances the generalizability of global models under the DG FL setting. By adopting the unlabeled data from the target domain, our *mix*EEG achieves even better performance under the DA FL setting.

In summary, the contributions of this paper are as follows:

- **New problem:** We for the first time investigate two new problem settings for cross-subject EEG FL: the DG FL setting and the DA FL setting.
- **New framework:** We introduce a new FL framework called *mix*EEG to improve the transferability of global models trained by FL methods.
- **New mixup:** Beyond the vanilla *mixup*, which linearly interpolates raw data, we specially design and investigate two new *mixup* approaches for EEG data: the *Channel Mixup* and the *Frequency Mixup*.
- **Extensive Experiments:** We conduct extensive experiments on two public datasets and two common network architectures to comprehensively investigate different *mix*EEG with diverse tailored *mixup*.

Related Work

Cross-subject EEG Classification

The inter-subject variability of EEG signals has hindered the promotion of EEG BCIs for a long period of time. Previous DG methods predominantly concentrated on traditional supervised learning or unsupervised learning. The mixture-of-experts are adopted to learn different brain regions representations (X.-H. Liu, Jiang, Zheng, & Lu, 2024). Contrastive learning enables networks to learn the difference among features, which is applied in both a graph-based multi-task self-supervised learning (GMSS) (Li et al., 2022) and a prototype contrastive domain generalization (PCDG) (Cai & Pan, 2023). However, these methods cannot be directly adopted for training a global model with multiple decentralized clients, limiting its access to large amount of data. Our *mix*EEG framework enables BCI developer to train a global model with the collaboration of multiple decentralized clients while preserving privacy.

EEG Federated Learning

Considering the privacy preservation of EEG, many previous works introduced FL in EEG classification tasks. R. Liu et al. proposed FLEEG to to surmount the challengings of the heterogeneity of EEG devices among different clients (hospitals and institutions) (R. Liu et al., 2024). FedEEG (Hang et al., 2023) used an inter-subject structure matching-based FL framework to extract the discriminative features for improving the performance of each clients. Mongardi and Pinoli explored the potential of federated learning for the task of emotion recognition from BCI data, focusing on performance with respect to centralized approaches (Mongardi & Pinoli, 2024). Despite there are lots of research on the EEG FL, the cross-subject issues are largely unexplored. Hence, we propose *mix*EEG to enhance the EEG FL for cross-subject EEG classification.

Mixup

Mixup (H. Zhang et al., 2018) is a simple yet surprisingly effective data augmentation technique, which linearly interpolates two data's inputs and labels to generate a *mixup* data. By adopting *mixup* strategy, centralized learning is able to train a more robust model with the same training dataset. Recently, lots of paper utilized *mixup* for fedreated learning in many research fields, like medical image segmentation (Wicaksana et al., 2022), vertical FL (Cheng et al., 2024), and image classification (Yoon, Shin, Hwang, & Yang, 2021).

Previous EEG classification work simply employed the vanilla mixup (Zhou et al., 2024; Yao, Liu, Lu, Wu, & Li, 2024). To tailor the vanilla *mixup* for diverse data modality, many works like mixing up data in the latent space (Verma et al., 2019), detecting the saliency feature for images (Uddin, Monira, Shin, Chung, & Bae, 2021), and *mixup* for natural language (H. Chen, Han, Yang, & Poria, 2022) were proposed. Taking the properties of EEG modality with multiple channels and frequency bands into consideration, we naturally propose *Channel Mixup* and *Frequency Mixup* for EEG modality.

Method

Problem Statement

Let $\mathcal{D}_s = \{X_s, Y_s\}$ and $\mathcal{D}_t = \{X_t, Y_t\}$ denote the source domain dataset and the target domain dataset. There are two normal cross-subject settings, domain generalization (DG) and domain adaptation (DA). Different from DG setting which requires training DL models only using the source dataset $\mathcal{D}_s$, DA setting allows trainers to have additional access to the unlabeled target data X_t.

Under the federated learning setting, there are K clients whose datasets multiple sub-dataset $\mathcal{D}_s^k$, where $k \in 1, ..., K$. For the purpose of privacy preservation, the direct data communication between clients is strictly forbidden, *i.e.*, each client train their local model only using their own sub-dataset $\mathcal{D}_s^k$. The goal of DG FL is to train a global model that can generalize to $\mathcal{D}_t$.

In contrast to DG FL, DA FL setting allows clients to access the unlabeled data X_t from target domain, making the cross-subject isssue easier but damaging the privacy of target domain $\mathcal{D}_t$. Hence, it is necessary to consider the trade-off of the privacy preservation and the data sharing. We define a sharing ratio of $r \in [0,1]$, denoting the ratio of the number of the shared data to the target domain data. When $r = 0$, the DA FL setting is equal to the DG FL setting. Larger r represents more data communication but less privacy preservation.

Baseline: FedAvg

Federated Averaging (FedAvg) is the most commonly used algorithm framework in FL (McMahan et al., 2017). For every communication round $t = 0, ..., T-1$, a part of clients train their own local models with their own sub-dataset $\mathcal{D}_s^k$. We denote the fraction of the number of clients participating in the update to the total number of clients as ϕ. Under the DA FL setting, each client also has the access to the unlabeled shared data from target domain dataset X_t. Afterwards, client k sends the model weight $\boldsymbol{w}_t^k$ to the global server, which update the parameters of global model by simply averaging the client parameters received by:

$$\boldsymbol{w}_t = \frac{1}{\phi K} \sum_{k=1}^{\phi K} \boldsymbol{w}_t^k. \tag{1}$$

The updated global model $\boldsymbol{w}_t$ is sent back to clients for the next round local training. The process will undergoes T times or until the global model converges.

Tailored *mixup*

Mixup (H. Zhang et al., 2018) is a simple data augmentation technique using a linear interpolation between two data-label pairs $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ to produce a mixup data-label pair

$$\widetilde{\boldsymbol{x}} = \lambda \boldsymbol{x}_i + (1-\lambda)\boldsymbol{x}_j, \quad \widetilde{y} = \lambda y_i + (1-\lambda) y_j, \tag{2}$$

where $\lambda \in [0,1]$ is a hyperparameter. Through this elegant augmentation, model gains better robustness and generalization (L. Zhang & Deng, 2021; Carratino, Cissé, Jenatton, & Vert, 2022). In this paper, we explore and validate three *mixup* methods on DE features as shown in Figure 2. It is worth noting that these tailored *mixup* might not always useful as we are the first to explore them.

1) *Linear Mixup*: The same as the original mixup.

2) *Channel Mixup*: Channel mixup could help the model learn localized spatial representations robust to inter-subject variability in channel activations. There are C channels for EEG data, we separate the whole C channels set into two non-overlapping subsets C_1 and C_2. The mixup data is:

$$\widetilde{\boldsymbol{x}} \in \mathbb{R}^{C \times F} = \boldsymbol{x}_i \odot M_r(C_1) + \boldsymbol{x}_j \odot M_r(C_2), \tag{3}$$

where $M_r() \in \{0,1\}^{C \times F}$ denotes a binary mask. The rows of $M_r(C_1)$ corresponding to the channels in C_1 are 1, and other rows are 0.

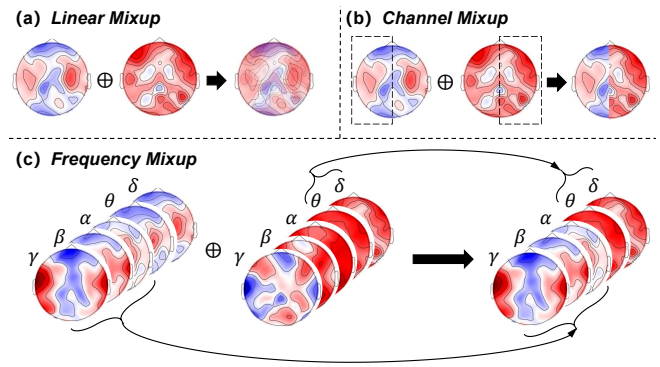

Figure 2: Three EEG mixup methods: (a) Linear Mixup, (b) Channel Mixup: an example that combines the left side and the right side of two EEG signals, (c) Frequency Mixup: an example that that combines the high frequency bands and the low frequency bands of two EEG signals.

3) *Frequency Mixup*: Frequency mixup may enable capturing discriminative features across multiple spectral bands, which is known to be important for EEG analysis tasks. There are F frequency bands for EEG data, we separate the whole F frequency bands set into two non-overlapping subsets F_1 and F_2. The mixup data is:

$$\widetilde{\boldsymbol{x}} \in \mathbb{R}^{C \times F} = \boldsymbol{x}_i \odot M_c(F_1) + \boldsymbol{x}_j \odot M_c(F_2), \tag{4}$$

where $M_c() \in \{0,1\}^{C \times F}$ denotes a binary mask. The columns of $M_c(F_1)$ corresponding to the frequency bands in F_1 are 1, and other columns are 0.

*mix*EEG

For FedAvg algorithm framework, the local clients updates the client model via stochastic gradient descent (SGD) for E epochs:

$$\boldsymbol{w}_{t+1,i+1}^k \leftarrow \boldsymbol{w}_{t+1,i}^k - \eta \nabla \mathcal{L}(f(\boldsymbol{x}_i^k; \boldsymbol{w}_{t+1,i}), y_i^k), i = 0, ..., E-1. \tag{5}$$

Here, η is learning rate, $\mathcal{L}$ is loss function, and $f(\boldsymbol{x}, \boldsymbol{w})$ is the model output of the input $\boldsymbol{x}$ given model weight $\boldsymbol{w}$.

Algorithm 1 *mix*EEG under the DA FL setting

Input: $\mathcal{D}_s^k = \boldsymbol{X}_s^k, \boldsymbol{Y}_s^k$ for $k = 1, ..., N$, $\mathcal{D}_t$, sharing ratio $r \in [0,1]$, when $r = 0$, this algorithm is FedAvg.
Initialize $\boldsymbol{w}_0$ for global model.
$\boldsymbol{X}_g \leftarrow GetSharingData(\mathcal{D}_t, r)$
for $t = 0, ..., T-1$ **do**
  for $k = 1, ..., N$ **do**
    $\boldsymbol{w}_{t+1}^k \leftarrow LocalUpdate(k, \boldsymbol{w}_t; \boldsymbol{X}_g)$
  end for
  $\boldsymbol{w}_t \leftarrow \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{w}_t^k$
end for

Algorithm 2 Local Update for *mix*EEG

Input: $k, \boldsymbol{w}_t; \boldsymbol{X}_g$, when $\boldsymbol{X}_g = \varnothing$, $\mathcal{L}_2 = 0$
$\boldsymbol{w} \leftarrow \boldsymbol{w}_t$
for $e = 0, ..., E-1$ **do**
  for $batch(\boldsymbol{X}, \boldsymbol{Y})$ **do**
    Select a sharing data $\boldsymbol{x}_g$ from $\boldsymbol{X}_g$
    $\boldsymbol{X}_m \leftarrow$ mixup $\boldsymbol{X}$ with $\boldsymbol{x}_g$
    $\mathcal{L}_1 \leftarrow \mathcal{L}(f(\boldsymbol{X}_m; \boldsymbol{w}), \boldsymbol{Y})$
    $\mathcal{L}_2 \leftarrow \sum_{c=0}^{n_class} \mathcal{L}(f(\boldsymbol{X}_m; \boldsymbol{w}), \mathsf{One}(c))$
    $\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta_{t+1} \nabla(\lambda \mathcal{L}_1 + (1-\lambda)\mathcal{L}_2)$
  end for
end for
return $\boldsymbol{w}$

The SGD algorithm for updating local models usually leads these models to overfit their own subset, damaging the global model's generalizability and making the global model hard to do cross-subject EEG classification. To this end, we propose *mix*EEG, a simple but effective framework for enhancing the generalizability of the global EEG model learned under the FL setting.

For the DG FL setting, *mix*EEG simply replacing the local updating strategy of local clients from SGD with *mixup*, i.e. using *mixup* technique to update the local model $\boldsymbol{w}_t^k$. We validate several different *mixup* strategies to fully investigate which type of *mixup* is suitable for EEG modality.

For the DA FL setting, local clients are able to see the $r \cdot |\mathcal{D}_t|$ unlabeled data. Here are two questions: **Q1**: what is the shared data? **Q2**: how do local clients leverage these unlabeled data?

For question **Q1**, we argue that directly sharing the subset of X_t definitely harm the privacy security of the target domain. So instead, we propose to generate shared data by averaging the target domain data X_t. Specifcally, we define a hyperparameter $s \in \mathbb{Z}^+$, indicating the number of data used for aggregating a single averaged data. We randomly shuffle the X_t, and linearly interpolating a number of s data from X_t to generate a single shared data by Eq 6. This process is repeated $r \cdot |\mathcal{D}_t|$ times to get enough shared data.

$$\widetilde{\boldsymbol{x}_t} = \frac{1}{s} \sum_{i=1}^{s} \boldsymbol{x}_t^i. \tag{6}$$

For question **Q2**, by assuming that the appearing probability of each emotion in the target dataset is equal, we can generate a pseudo label for each shared data: $\tilde{y}_t = \frac{1}{c}[1, 1, ..., 1] \in \mathbb{R}^c$, where c is the number of emotion classes. After generating the shared data and correpsonding labels, *mix*EEG can leverage these data from target dataset to improve the transfer ability. Unlike the DG FL issue, where we mixup data within a single batch, we mixup each data-label pairs $(\boldsymbol{x}_s^i, y_s^i)$ from $\mathcal{D}_s^k$ with the shared data. Decoding $\mathsf{One}()$ as the onehot encoding, the corresponding mixup label is:

$$\widetilde{y} = \lambda \mathsf{One}(y_s^i) + (1-\lambda)\frac{1}{c}[1, 1, ..., 1]. \tag{7}$$

Experiments

EEG Datasets

We exploit two datasets: an epilepsy detection dataset (Shoeb & Guttag, 2010) and an emotion recognition dataset (Zheng & Lu, 2015). Differential Entropy (DE) features (Duan, Zhu, & Lu, 2013) are extracted for classification.

Epilepsy Detection (ED): CHB-MIT dataset is an epilepsy dataset required from 23 patients. We cropped and resampled the CHB-MIT dataset to build an ED dataset with four types of EEG: ictal, preictal, postictal, and interictal phase EEG.

Emotion Recognition (ER): SEED dataset records the 62-channels EEG data corresponding to 3 types of emotions from 15 subjects: positive, neutral, and negative.

Experimental Setup

Each subject is regarded as a decentralized source domain, and the results under leave-one-subject-out (LOSO) settings are reported. Specifically, we select $n-1$ subjects as source domains, train a global model using FL methods on the source domains (the direct data communication among source domains) and test the performance on the last remaining subject, which is regarded as target domain.

The number of update round of the global server $T = 50$, for each client, the number of local update round $E = 5$. We use SGD as the optimizer, with its learning rate $\eta = 0.01$. The fraction ϕ is set to 0.2 for ER dataset, and 0.3 for ED dataset. The sharing ratio r is 0.1 and aggregating hyperparameter $s = 10$ for DA FL setting.

EEG Models

There are many methods for the classification of EEG data. We choose the following two type of models: **1) MLP:** Multi-Layer Perception, here we use an MLP with 2 hidden layers of 128 and 64 neurons. **2) CNN:** Convolutional Neural Network, here we use a 3 layer CNN with kernal sizes of {(3,3), (3,3), and (7,1)}. The activate function is the GELU function and the dropout probability is 0.5 for all models.

EEG Mixup Appoaches

Basically, we evaluate three *mixup* ways, but there are more implementation details to evaluate: **1) *Linear Mixup***: In the origin paper, $\lambda \sim \text{Beta}(\alpha, \alpha), \alpha \in (0, +\infty)$, we test two cases: $\alpha = 0.2$ and $\alpha = 5$. **2) *Channel Mixup***: We test two cases: (1) dividing the EEG channels into the left scalp and the right scalp; (2) dividing the EEG channels equally into two sets randomly. We set $\lambda = 0.5$ for both cases. **3) *Frequency Mixup***: For the DE features with five frequency bands (δ, θ, α, β, and γ), we test two cases: (1) dividing the EEG frequency bands into the $\{\alpha, \beta, \gamma\}$ and $\{\delta, \theta\}$; (2) dividing the EEG frequency bands into the $\{\delta, \alpha, \gamma\}$ and $\{\theta, \beta\}$. We set $\lambda = 0.6$ for both cases.

Results

DG FL Setting

Under DG FL settings, all clients can only access to their own data. We compare our *mix*EEG with the baseline FedAvg

Table 1: **The DG FL setting**: The accuracy, F1 score and Cohen's Kappa score (%) under the DG FL setting. L, C, and F indicate Linear, Channel, and Frequency mixup strategies, respectively.

	Methods	Epilepsy Detection: CHB-MIT, 4-class			Emotion Recognition: SEED, 3-class		
		Accuracy	F1 Score	Cohen's Kappa	Accuracy	F1 Score	Cohen's Kappa
MLP	FedAvg	43.676 ± 4.367	41.763 ± 4.349	27.376 ± 5.823	77.053 ±7.824	76.826 ± 8.008	65.514 ± 11.768
	L α = 0.2	47.618 ± 3.925	**46.895 ± 3.532**	30.158 ± 5.233	70.505 ± 4.987	69.946 ± 5.252	55.625 ± 7.551
	L α = 5	**47.679 ± 4.064**	46.103 ± 3.303	**30.238 ± 5.419**	77.387 ± 8.216	76.856 ± 8.536	65.939 ± 12.409
	C *Binary*	47.742 ± 3.863	45.766 ± 3.365	30.322 ± 5.151	86.082 ± 4.607	85.936 ± 4.753	79.062 ± 6.924
	C *Random*	46.984 ± 3.907	44.888 ± 3.139	29.312 ± 5.209	**86.306 ± 6.158**	**86.148 ± 6.246**	**79.419 ± 9.232**
	F {α, β, γ}	47.359 ± 3.773	44.941 ± 3.353	29.812 ± 5.031	83.079 ± 3.562	82.910 ± 3.594	74.524 ± 5.383
	F {δ, α, γ}	45.863 ± 4.359	44.432 ± 3.612	27.818 ± 5.812	82.224 ± 4.200	82.037 ± 4.249	73.240 ± 6.276
CNN	FedAvg	45.142 ± 4.420	44.000 ± 4.732	26.857 ± 5.894	83.943 ± 7.437	83.843 ± 7.443	75.876 ± 11.143
	L α = 0.2	**48.606 ± 4.266**	**48.047 ± 4.017**	**31.474 ± 5.688**	**88.814 ± 4.928**	**88.759 ± 4.953**	**83.205 ± 7.396**
	L α = 5	46.698 ± 3.275	44.967 ± 2.530	28.931 ± 4.367	87.254 ± 5.789	87.118 ± 4.706	80.850 ± 8.694
	C *Binary*	47.127 ± 4.147	45.587 ± 4.104	29.503 ± 5.530	84.486 ± 6.368	84.399 ± 6.384	76.718 ± 9.542
	C *Random*	46.577 ± 3.782	43.526 ± 3.925	28.771 ± 5.042	85.989 ± 7.578	85.916 ± 7.598	79.007 ± 11.301
	F {α, β, γ}	46.246 ± 3.467	44.000 ± 3.358	28.328 ± 4.622	86.890 ± 5.445	86.818 ± 5.549	80.304 ± 8.180
	F {δ, α, γ}	45.338 ± 3.251	43.599 ± 2.942	27.118 ± 4.335	87.170 ± 5.045	87.075 ± 5.124	80.732 ± 7.571

method. The LOSO results are presented in Table 1. It can be observed that simply updating the client models with local mixup have already enhanced the transferability of global model, as most results of *mix*EEG are higher than the results of the baseline FedAvg. *mix*EEG improves the cross-subject EEG classification from 77.0% to 86.3% for ER task, and from 43.6% to 47.6% for ED task.

It can be seen that for MLP architecture, *Linear Mixup* is the best for ED task, while *Channel Mixup* with Binary Split is the best for ER task. However, for CNN architecture, *Linear Mixup* is always better across two datasets.

DA FL Setting

Under DA FL settings, there are $r \cdot |\mathcal{D}_t|$ unlabeled data from target domain sharing to local clients. The results are presented in Table 2, it can been observed that *mix*EEG enhances the performance from 46% to 49% for the ED task and from 90% to 93% for the ER task.

In this setting, the *Channel Mixup* method with Binary Split outperforms other *mixup* strategies for MLP architecture across two datasets. While for CNN architecture, the *Linear Mixup* exhibits the best improvement on these two tasks. However, we notice that the performance of *Frequency Mixup* doesn't work well under the DA FL setting, except for MLP architecture on the ED dataset, *Frequency Mixup* always harms the generalizabitliy for other cases. Taking the performance of the DG FL setting into consideration, we can conclude that *Frequency Mixup* might not be suitable for EEG modality.

Ablation Study

***Linear Mixup* Hyperparameter α** Through the experiments of the DG/DA FL settings, it can be found that the performance of *Linear Mixup* are influenced by the hyperparameter α. Thus, we conduct an extra experiment to evaluate the impact of α. We choose 14 values for α and exhibits the performance of CNN model on the ED dataset, as shown in Fig 3. It can be seen that in this case, the best performance is achieved when α is around 0.3. When α is larger than 1.0, the performances are similar.

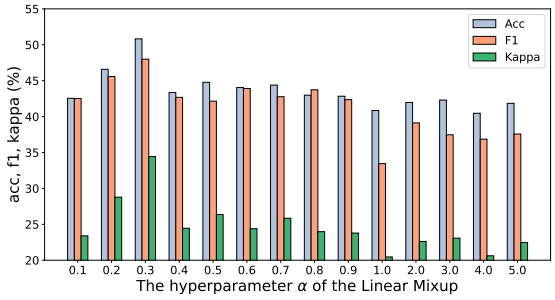

Figure 3: The accuracy, F1 score and Cohen's Kappa score of *Linear Mixup* with various α on the Epilepsy Detection dataset. The network architecture is MLP.

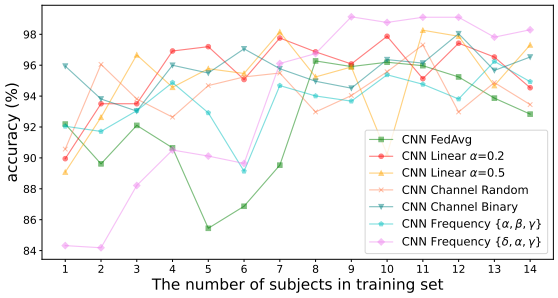

Figure 4: The accuracy of the global model trained with different training set sizes on the SEED dataset. We choose the No.15 subject as the target domain, and adopt other 14 subjects as the training set. The network architecture is CNN.

Training Set Size To investigate the *Scaling Law* in EEG BCIs, we further train the global model with different numbers of source domain clients. As depicted in Fig 4, for all *mixup* strategies, the performance fluctuates since the the EEG data from different subjects vary significantly. But despite the fluctuation, the performances increase when the

Table 2: **The DA FL setting**: The accuracy, F1 score and Cohen's Kappa score (%) under the DA FL setting, where we set $r = 0.1$ in our experiment. L, C, and F indicate Linear, Channel, and Frequency *mixup* strategies, respectively.

		Epilepsy Detection: CHB-MIT, 4-class			Emotion Recognition: SEED, 3-class		
	Methods	**Accuracy**	**F1 Score**	**Cohen's Kappa**	**Accuracy**	**F1 Score**	**Cohen's Kappa**
MLP	FedAvg	46.806 ± 4.730	45.948 ± 4.989	30.742 ± 6.307	90.832 ± 6.181	90.769 ± 6.266	86.222 ± 9.290
	L α = 0.2	48.572 ± 4.437	47.065 ± 4.331	31.429 ± 5.919	91.126 ± 5.376	91.007 ± 5.521	86.642 ± 8.099
	L α = 5.0	47.543 ± 3.945	45.040 ± 4.008	30.058 ± 5.260	90.933 ± 5.808	90.906 ± 5.814	86.379 ± 8.714
	C *Binary*	**49.181 ± 4.558**	**47.936 ± 4.208**	**32.242 ± 6.077**	**93.077 ± 3.690**	**93.046 ± 3.702**	**89.594 ± 5.543**
	C *Random*	48.625 ± 4.103	47.649 ± 3.733	31.500 ± 5.471	89.770 ± 4.386	89.616 ± 4.535	84.624 ± 6.568
	F {α, β, γ}	48.305 ± 4.351	47.237 ± 4.053	31.074 ± 5.802	88.602 ± 4.656	88.466 ± 4.730	82.865 ± 6.996
	F {δ, α, γ}	47.533 ± 3.979	45.944 ± 3.614	30.045 ± 5.306	89.255 ± 5.929	89.174 ± 5.969	83.878 ± 8.868
CNN	FedAvg	48.431 ± 3.978	**47.802 ± 3.496**	31.242 ± 5.304	90.013 ± 4.722	89.997 ± 4.699	84.998 ± 7.086
	L α = 0.2	**48.484 ± 3.773**	47.661 ± 3.473	**31.313 ± 5.031**	90.671 ± 5.113	90.654 ± 5.097	85.966 ± 7.689
	L α = 5	47.686 ± 4.068	45.914 ± 3.357	30.248 ± 5.425	**92.795 ± 4.469**	**92.733 ± 4.552**	**89.160 ± 6.734**
	C *Binary*	47.427 ± 4.277	46.233 ± 4.143	29.903 ± 5.702	88.784 ± 5.675	88.753 ± 5.701	83.142 ± 8.529
	C *Random*	47.654 ± 3.497	46.588 ± 3.230	30.206 ± 4.662	89.796 ± 5.265	89.705 ± 5.346	84.675 ± 4.706
	F {α, β, γ}	47.529 ± 3.690	46.519 ± 3.145	30.038 ± 4.920	88.391 ± 6.627	88.272 ± 6.779	82.549 ± 9.952
	F {δ, α, γ}	47.330 ± 3.907	46.507 ± 3.654	29.774 ± 5.209	88.366 ± 8.192	88.261 ± 8.253	82.491 ± 12.334

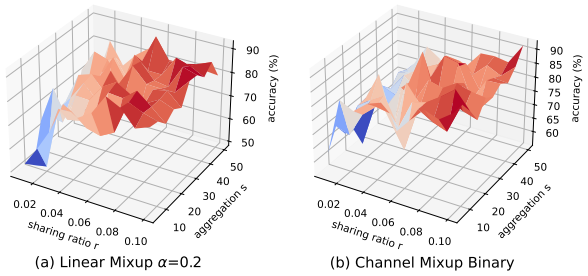

Figure 5: The 3D heatmap of the LOSO results on the SEED dataset with different sharing ratio r and aggregating number s. The network architecture is MLP.

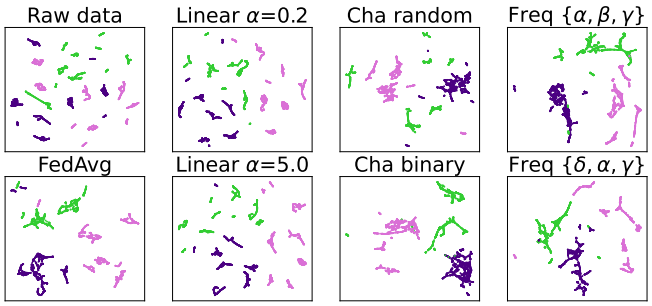

Figure 6: The UMAP visualization of learned feature on the SEED dataset of each global model using different training strategy under the DA FL setting. Purple, Green, and Pink stands for the negative, neutral, and positive emotion.

training set get larger, demonstrating that larger training data helps enhancing the generalizabitliy.

Sharing ratio and Aggregating number We conduct additional experiments to validate the effectiveness of the sharing data by adjusting the sharing ratio r and the aggregating hyperparameter s. The range of r is from 0.01 to 0.1 and the range of s is from 5 to 50 (step is 5). Since this experiment is quite large (consists of 100 sets of hyperparameters), we only run the global model for 10 epochs, i.e., $T = 10$ in this case. Displayed in Fig 5, it can be observed that the performances increase when sharing ratio r increases, indicating that more sharing data helps cross-subject EEG classification. We can also see that though the performances fluctuates, larger aggregation s achieves better results.

Representation Visualization

To better present the difference between each *mixup* strategies and explore the reason why some *mixup* strategies are better on learning transferable features, we adopt UMAP technique (McInnes, Healy, & Melville, 2018) to visualize the outputs of the global model's last layer, *i.e.*, the learned features of the target domain's data.

Demonstrated in Fig 6, where we plot the learned feature of the MLP architechture on the SEED dataset. It can be seen that FedAvg's feature has some chaotic distribution on the left up side. *Linear Mixup* tends to learn more dispersed features, thus may increase the generalizability in some cases. In this case (MLP on ER task), the *Channel Mixup* with Binary is the best, it can be seen that the learned feature are more distinguishable since the features of the same class are aggregated more tightly. *Frequency Mixup* also enables the global model to learn more tightly aggregated features, but the learning on the boundaries between classes is relatively vague, leading to suboptimal generalizabitliy.

Conclusion

In this paper, we investigate two new problem settings in EEG-based BCI, the DG FL setting and the DA FL setting. We propose a simple but effective framework *mix*EEG, and investigate multiple tailored *mixup* strategies on two datasets. We found that by sharing averaged data, *mix*EEG can leverage these unlabeled data to further improve the generalizabitliy of global model. Extensive experiments validate the efficacy of *mix*EEG and indicate that *Linear Mixup* (CNN perfers) and *Channel Mixup* (MLP perfers) are useful for enhancing generalizability.

Acknowledgement

This work was supported in part by grants from National Natural Science Foundation of China (Grant No. 62376158), STI 2030-Major Projects+2022ZD0208500, Shanghai Jiao Tong University 2030 Initiative, Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZD ZX), Medical-Engineering Interdisciplinary Research Foundation of Shanghai Jiao Tong University "Jiao Tong Star" Program (YG2023ZD25, YG2024ZD25 and YG2024QNA03), Shanghai Pujiang Program (Grant No. 22PJ1408600), Shanghai Pilot Program for Basic Research - Shanghai Jiao Tong University (No. 21TQ1400203), and Shanghai Jiao Tong University SCS-Shanghai EmoRays Technology Co., Ltd Joint Laboratory of Affective Brain-Computer Interfaces.

References

Aggarwal, S., & Chugh, N. (2022). Review of machine learning techniques for EEG based brain computer interface. *Archives of Computational Methods in Engineering*, *29*(5), 3001–3020.

Aghajanyan, A., Yu, L., Conneau, A., Hsu, W.-N., Hambardzumyan, K., Zhang, S., … Zettlemoyer, L. (2023). Scaling laws for generative mixed-modal language models. In *International Conference on Machine Learning* (pp. 265–279).

Baghersalimi, S., Teijeiro, T., Atienza, D., & Aminifar, A. (2021). Personalized real-time federated learning for epileptic seizure detection. *IEEE Journal of Biomedical and Health Informatics*, *26*(2), 898–909.

Cai, H., & Pan, J. (2023). Two-phase prototypical contrastive domain generalization for cross-subject EEG-based emotion recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–5).

Carratino, L., Cissé, M., Jenatton, R., & Vert, J.-P. (2022). On mixup regularization. *Journal of Machine Learning Research*, *23*(325), 1–31.

Chan, C., Zheng, Q., Xu, C., Wang, Q., & Heng, P. A. (2024). Adaptive federated learning for EEG emotion recognition. In *International Joint Conference on Neural Networks* (pp. 1–8).

Chen, H., Han, W., Yang, D., & Poria, S. (2022). Doublemix: Simple interpolation-based data augmentation for text classification. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 4622–4632).

Chen, X., Niu, Y., Zhao, Y., & Qin, X. (2024). An efficient group federated learning framework for large-scale EEG-based driver drowsiness detection. *International Journal of Neural Systems*, *34*(1).

Cheng, Y., Zhang, L., Wang, J., Chu, X., Huang, D., & Xu, L. (2024). Fedmix: Boosting with data mixture for vertical federated learning. In *IEEE 40th International Conference on Data Engineering* (pp. 3379–3392).

Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., … Jitsev, J. (2023). Reproducible scaling laws for contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2818–2829).

Duan, R.-N., Zhu, J.-Y., & Lu, B.-L. (2013). Differential entropy feature for EEG-based emotion classification. In *6th international IEEE/EMBS Conference on Neural Engineering* (pp. 81–84).

Finley, A. J., Angus, D. J., Knight, E. L., van Reekum, C. M., Lachman, M. E., Davidson, R. J., & Schaefer, S. M. (2024). Resting EEG periodic and aperiodic components predict cognitive decline over 10 years. *Journal of Neuroscience*, *44*(13).

Hang, W., Li, J., Liang, S., Wu, Y., Lei, B., Qin, J., … Choi, K.-S. (2023). FedEEG: Federated EEG decoding via inter-subject structure matching. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–5).

Jiang, W.-B., Liu, X.-H., Zheng, W.-L., & Lu, B.-L. (2023). Multimodal adaptive emotion transformer with flexible modality inputs on a novel dataset with continuous labels. In *proceedings of the 31st acm international conference on multimedia* (pp. 5975–5984).

Jiang, W.-B., Liu, X.-H., Zheng, W.-L., & Lu, B.-L. (2024). Seed-vii: A multimodal dataset of six basic emotions with continuous labels for emotion recognition. *IEEE Transactions on Affective Computing*.

Ju, C., Gao, D., Mane, R., Tan, B., Liu, Y., & Guan, C. (2020). Federated transfer learning for EEG signal classification. In *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (pp. 3040–3045).

Li, Y., Chen, J., Li, F., Fu, B., Wu, H., Ji, Y., … Zheng, W. (2022). Gmss: Graph-based multi-task self-supervised learning for EEG emotion recognition. *IEEE Transactions on Affective Computing*, 1-1.

Liu, R., Chen, Y., Li, A., Ding, Y., Yu, H., & Guan, C. (2024). Aggregating intrinsic information to enhance bci performance through federated learning. *Neural Networks*, *172*, 106100.

Liu, X.-H., Jiang, W.-B., Zheng, W.-L., & Lu, B.-L. (2024). Moge: Mixture of graph experts for cross-subject emotion recognition via decomposing eeg. In *2024 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 3515–3520).

Liu, X.-H., Liu, Y.-K., Wang, Y., Ren, K., Shi, H., Wang, Z., … Zheng, W.-L. (2024). EEG2video: Towards decoding dynamic visual perception from EEG signals. In *The thirty-eighth annual conference on neural information processing systems (neurips)*. Retrieved from `https://openreview.net/forum?id=RfsfRn9OFd`

Liu, X.-H., Song, X., He, D., Lu, B.-L., & Zheng, W.-L. (2024). Professor x: Manipulating eeg bci with invisible and robust backdoor attack. *arXiv preprint arXiv:2409.20158*.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension

reduction. *arXiv preprint arXiv:1802.03426*.

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).

Mongardi, S., & Pinoli, P. (2024). Exploring federated learning for emotion recognition on brain-computer interfaces. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization* (pp. 622–626).

Shoeb, A. H., & Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 975–982).

Tan, E., Troller-Renfree, S. V., Morales, S., Buzzell, G. A., McSweeney, M., Antúnez, M., & Fox, N. A. (2024). Theta activity and cognitive functioning: Integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research. *Developmental Cognitive Neuroscience*, *67*, 101404.

Tasci, I., Tasci, B., Barua, P. D., Dogan, S., Tuncer, T., Palmer, E. E., … Acharya, U. R. (2023). Epilepsy detection in 121 patient populations using hypercube pattern from EEG signals. *Information Fusion*, *96*, 252–268.

Uddin, A. S., Monira, M. S., Shin, W., Chung, T., & Bae, S.-H. (2021). Saliencymix: A saliency guided data augmentation strategy for better regularization. In *International Conference on Learning Representations.*

Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., & Bengio, Y. (2019). Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning* (pp. 6438–6447).

Wicaksana, J., Yan, Z., Zhang, D., Huang, X., Wu, H., Yang, X., & Cheng, K.-T. (2022). Fedmix: Mixed supervised federated learning for medical image segmentation. *IEEE Transactions on Medical Imaging*, *42*(7), 1955–1968.

Yao, S., Liu, L., Lu, J., Wu, D., & Li, Y. (2024). Advancing semi-supervised EEG emotion recognition through feature extraction with mixup and large language models. In *2024 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 2772–2779).

Yoon, T., Shin, S., Hwang, S. J., & Yang, E. (2021). Fedmix: Approximation of mixup under mean augmented federated learning. In *9th International Conference on Learning Representations.*

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond empirical risk minimization. In *International conference on learning representations.*

Zhang, L., & Deng, Z. (2021). How does mixup help with robustness and generalization? In *The Ninth International Conference on Learning Representations.*

Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175.

Zhou, R., Ye, W., Zhang, Z., Luo, Y., Zhang, L., Li, L., … Liang, Z. (2024). EEGmatch: Learning with incomplete labels for semisupervised eeg-based cross-subject emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*.