FedGLoRA: Grassmann-Manifold Federated Learning via Dual LoRA for Large EEG Models

Qianyu Chen[1], **Yihao Zhong**[2], **Runxuan Tang**[1], **Tianyi Zhang**[1,3], **Jing Liu**[4]
Ziyu Jia[5] and **Chenyu Liu**[1]
[1]Nanyang Technological University
[2]New York University
[3]Uber Technologies Inc.
[4]Amazon AWS
[5]Institute of Automation, Chinese Academy of Sciences
{chen1981,tang0609,chenyu003}@e.ntu.edu.sg, yz7654@nyu.edu, tyzhang621@gmail.com,
liumjin@amazon.com, jia.ziyu@outlook.com

Abstract

Large EEG Models (LEMs) are drawing increasing attention in EEG, as large-scale pretraining yields transferable representations that improve generalization. As EEG research moves to real-world deployment, objectives and paradigms diversify, yielding increasingly heterogeneous and unevenly scaled datasets across institutions. This evolution induces distribution drift and scale imbalance, making static pretrained LEMs brittle and requiring adaptation to newly collected data. Under the constraints of the high cost of retraining and the privacy sensitivity of newly collected data, federated learning offers a practical route to update LEMs via decentralized parameter updates. Yet standard federated fine-tuning is unstable for pretrained LEMs, suffering backbone drift, misaligned updates under non-IID heterogeneity, and scale imbalance. We propose FedGLoRA: we freeze the backbone and communicate only LoRA adapters via a dual-branch design, with a shared global branch and a private on-device branch. FedGLoRA aggregates global updates on a Grassmann subspace with a Grassmann-proximal constraint, and uses bidirectional distillation to mitigate imbalance. Experiments on motor imagery and emotion recognition across three different LEMs show consistent gains and improved stability over strong baselines.

1 Introduction

Electroencephalography (EEG) is a non-invasive modality for recording brain activity and supports cognitive neuroscience, neurological disorder analysis, and brain–computer interfaces (BCIs) [Zhou *et al.*, 2025]. However, real-world EEG datasets are often small and subject-specific, making robust generalization across individuals and recording conditions challenging [He and Wu, 2019; Chen and Yu, 2026]. Recently, Large EEG Models (LEMs) have addressed this limitation by self-supervised pretraining on large, diverse EEG corpora, learning transferable spatiotemporal representations that improve performance across downstream tasks and domains [Jiang *et al.*, 2024; Wang *et al.*, 2024; Wang *et al.*, 2025a; Liu *et al.*, 2025; Tang *et al.*, 2026].

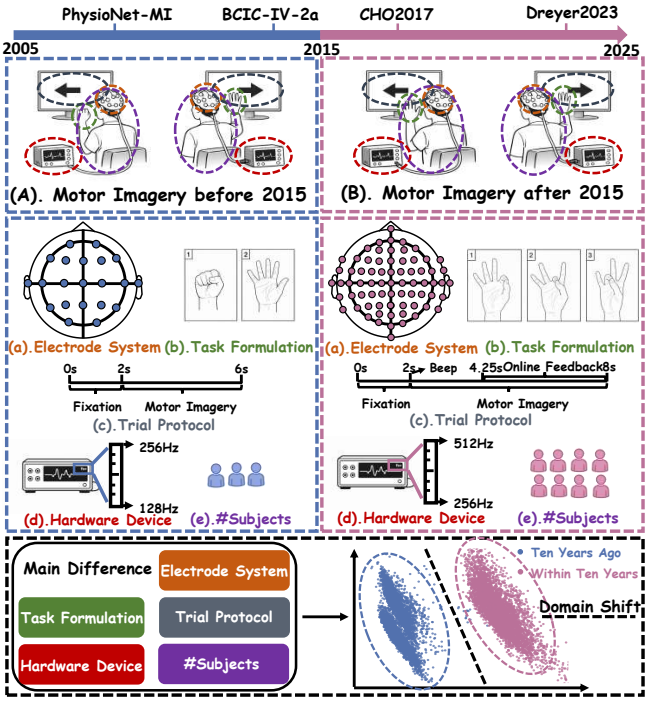

Figure 1: Evolution of motor-imagery EEG benchmarks and the resulting domain shift. We contrast representative datasets before and after 2015 and summarize five systematic changes: (a) electrode system, (b) task formulation, (c) trial protocol, (d) hardware device, and (e) #subjects. These co-evolving factors reshape the data distribution over time, inducing pronounced cross-dataset domain shift.

However, as LEMs move toward real-world deployment, we observe a phenomenon: **the distribution and scale of EEG datasets co-evolve with the research commu-**

nity. This evolution is driven by the evolving EEG research system, manifested as systematic shifts in research objectives, experimental paradigms, data-collection conditions, and devices [Xu *et al.*, 2020]. As EEG research progresses from controlled, small-scale laboratory validation to real-world deployment, the field broadens from relatively homogeneous paradigms to diverse in-the-wild scenarios, prioritizing robustness and generalization while exacerbating cross-paradigm distribution shift [Jayaram and Barachant, 2018]. Meanwhile, as more institutions participate and adoption grows, data collection scales unevenly across sites, resulting in pronounced scale imbalance [Luo *et al.*, 2021].

Formally, task evolution is evident in benchmark task formulations for MI and emotion recognition. In MI benchmarks, task formulations have evolved from simple left–right binary classification to finer-grained multi-class protocols [Zhang *et al.*, 2025]. Moreover, the labels are no longer tied solely to hand laterality, whereas in many paradigms they denote distinct unilateral action semantics and execution dynamics, such as finger-movement imagery or rhythmic grasping. Consequently, MI datasets can encode different action meanings, inducing systematic distribution shifts [Tangermann *et al.*, 2012; Cho *et al.*, 2017; Schalk *et al.*, 2004; Dreyer *et al.*, 2023]. In emotion recognition, protocols have moved from coarse valence categories to finer affective states, and elicitation paradigms have diversified, with video and game experiments inducing distinct EEG dynamics [Katsigiannis and Ramzan, 2017; Zheng *et al.*, 2018; Chen *et al.*, 2023; Alakus *et al.*, 2020]. These trends show that EEG benchmark distributions and scale are nonstationary, so LEMs pretrained on historical data can lose effectiveness as distributions drift [Gama *et al.*, 2014]. Therefore, sustaining the utility of LEMs requires adaptation on an enlarging corpus of newly collected EEG datasets [Bian *et al.*, 2025]. However, repeatedly training LEMs is resource-intensive, motivating the central question we study:

How can LEMs be updated efficiently on an enlarging corpus of newly collected EEG datasets under persistent domain shifts?

To solve this problem, LEMs must encounter heterogeneous, multi-site data distributions and align and integrate cross-domain updates into its representation space [Rieke *et al.*, 2020; Fu *et al.*, 2024]. However, relying solely on local fine-tuning is insufficient: it lacks multi-source diversity, readily overfits site-specific artifacts, and can induce representational drift that weakens cross-site transfer [Wang *et al.*, 2020]. Centralized joint fine-tuning could aggregate multi-source information, but it is often infeasible for multi-center EEG which is subject to strict privacy and regulatory constraints [Mongardi and Pinoli, 2024]. Therefore, federated learning (FL) offers a practical alternative: each site optimizes on private data and shares only model updates for server-side aggregation, enabling collaborative continual fine-tuning without exposing data [Guo *et al.*, 2021].

While FL has achieved promising results in many domains, directly applying standard FL to federated fine-tuning of LEMs is often ineffective, due to EEG's distinctive properties [Xu *et al.*, 2020; Cooray *et al.*, 2025].

(1) Destructive backbone drift. Direct full-parameter aggregation is a widely used fine-tuning strategy, but under cross-site domain shift it can be severely harmful for LEMs. Under cross-site heterogeneity, conflicting client gradients may drive uncontrolled backbone shifts that erase pretrained knowledge, causing negative transfer and degraded generalization [Huo *et al.*, 2026; Wang *et al.*, 2025b].

(2) Pronounced multi-source heterogeneity. Federated EEG exhibits strong cross-site heterogeneity, leading to a highly non-IID regime [Baykara *et al.*, 2025]. As a result, local updates are often misaligned, inducing negative transfer that harms site-wise and overall generalization.

(3) Severe scale imbalance across sites. Federated EEG often shows extreme data imbalance: due to differences in hardware devices, trial protocols, and recruitment cost, some sites contribute only a few subjects. This lets high-volume sites dominate aggregation, biasing the global model toward majority patterns and underfitting site-specific shifts, degrading performance at minority centers.

Motivated by these challenges, we propose **FedGLoRA**, a federated fine-tuning framework for LEMs that enables cross-site joint adaptation under privacy constraints. FedGLoRA continually integrates knowledge from diverse EEG domains while preserving site-specific behavior. For **Challenge 1**, we freeze the pretrained backbone and communicate only LoRA updates, using Dual LoRA with a shared global branch for transfer and private LoRA on device for local shifts. For **Challenge 2**, we aggregate global updates in a shared low-dimensional Grassmann subspace and apply a Grassmann proximal regularizer to reduce misalignment and optimization conflicts. For **Challenge 3**, we use bidirectional knowledge distillation to transfer shared knowledge to local models while feeding back residual site-specific cues to the global model, improving robustness under non-independent and identically distributed data and imbalance. Across three different LEMs on multi-center motor imagery and emotion recognition, FedGLoRA consistently outperforms federated baselines and yields more stable multi-task gains.

Our main contributions include:

- **First Systematic Study of Federated Fine-tuning for LEMs.** We present the first systematic investigation of federated fine-tuning for LEMs on multi-site EEG, identifying key cross-site deployment challenges and outlining a practical privacy-preserving adaptation route.
- **FedGLoRA: A New FL Paradigm.** We propose FedGLoRA, a new FL paradigm that aggregates LoRA updates on Grassmann subspaces to improve robustness under scale imbalance and non-IID heterogeneity.
- **Improved Performance on Diverse EEG Tasks.** Experiments on motor imagery and emotion recognition achieve improved performance, with FedGLoRA outperforming strong federated baselines.

2 Related Work

2.1 Large EEG Models (LEMs)

LEMs for brain decoding are typically self-supervised pretrained on large EEG corpora to learn transferable spatiotemporal representations [Wu *et al.*, 2024], which improves downstream performance [Jiang *et al.*, 2024; Wang *et al.*, 2024; Wang *et al.*, 2025a]. However, deployment faces substantial distribution shift from evolving acquisition conditions, so pretrained LEMs often require adaptation to new populations and settings [Gama *et al.*, 2014]. Although centralized multi-domain fine-tuning can reduce this drift, it is rarely feasible because raw EEG is privacy-sensitive [Liu *et al.*, 2024a]. We therefore propose a federated framework that enables cross-site fine-tuning without sharing raw data, improving generalization under privacy constraints.

2.2 Federated Learning for Foundation Models

Pretraining followed by downstream fine-tuning is the dominant foundation-model paradigm, but effective adaptation often needs large, diverse data [Zhou *et al.*, 2025]. Because downstream datasets are siloed across institutions and cannot be centralized under privacy constraints, federated learning enables joint fine-tuning while keeping raw data local [McMahan *et al.*, 2017; Zhou *et al.*, 2025]. Although recent work studies federated adaptation of large language models [Qi *et al.*, 2024; Yang *et al.*, 2025], EEG poses distinct challenges as structured non-IID shifts from sites cause update misalignment, further exacerbated by cross-site scale imbalance [Zhang *et al.*, 2022]. This motivates EEG federated designs that couple shared transfer with site-specific adaptation.

2.3 Federated Learning for EEG

EEG varies across sites, so reliable performance needs diverse cohorts [Liu *et al.*, 2024a]. Privacy and compliance rules prevent centralizing multicenter EEG, so federated learning enables joint training without sharing raw signals [Wang *et al.*, 2020]. Studies on motor imagery and emotion recognition identify two bottlenecks as non-IID heterogeneity and scale imbalance [Rieke *et al.*, 2020]. FedBS and FLEEG mitigate them via normalization, sharpness optimization, and hierarchical personalization [Liu *et al.*, 2024b; Jia *et al.*, 2024]. Yet most work targets task models, leaving federated adaptation of LEMs underexplored [Kuruppu *et al.*, 2026]. We study federated fine-tuning of LEMs and propose a framework for heterogeneity [Bian *et al.*, 2025].

3 Preliminaries

Federated Learning For LEMs. Given a multi-center EEG network with N sites, each site i owns a private dataset $D^{(i)} = (X^{(i)}, Y^{(i)}, t^{(i)})$, where $X^{(i)} \in \mathbb{R}^{n_i \times T_i \times C_i}$ denotes EEG segments with center-dependent channel layouts and acquisition settings, $Y^{(i)}$ denotes supervision signals whose label spaces may vary across tasks, and $t^{(i)}$ is a paradigm indicator. Due to privacy and regulatory constraints, raw EEG cannot be pooled across sites. Let $f(X \mid t; \Theta)$ be a pretrained LEM with parameters Θ. Federated learning for LEMs aims to perform cross-site joint fine-tuning by collaboratively optimizing a set of trainable parameters ϕ on EEG data:

$$\min_{\phi} \mathcal{L}(\phi) = \sum_{i=1}^{N} \alpha_i \, \mathbb{E}_{(X,Y,t)\sim D^{(i)}} \big[\ell\big(f(X \mid t; \Theta, \phi),\, Y\big)\big], \tag{1}$$

where α_i is the aggregation weight and $\ell(\cdot,\cdot)$ is the task loss. This formulation provides a unified view of federated fine-tuning for LEMs; in practice, however, the multi-center EEG setting induces pronounced heterogeneity and scale imbalance across sites, which motivates the designs in our method.

4 Method

4.1 Framework Overview

In this subsection, we present our federated learning framework as shown in Figure 2, which iterates four stages. **(1) Site Upload:** prioritize privacy and locality, sharing only non-sensitive updates. **(2) Server Update:** robustly aggregate heterogeneous clients to recover shared structure and mitigate imbalance. **(3) Site Update:** integrate global knowledge while stabilizing training and preserving personalization. **(4) Inference:** deploy a site-composed model combining shared and local adaptations on downstreams. We next detail the design and methods in each stage.

4.2 Dual Low-rank Adaptation

Intuition: In the site-upload step, motivated by the central tension in full-parameter fine-tuning and highly heterogeneous cross-site EEG, models must capture both transferable shared representations and site-specific personalization.

Therefore, we introduce a dual-LoRA structure that explicitly disentangles a communicable global LoRA from a locally updated personalized LoRA. Formally, for client k, we parameterize any LoRA-injected weight matrix $W^{(k)} \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ as the sum of the frozen pretrained weight W_0 and the low-rank increments from a global and a personalized LoRA branch as:

$$W^{(k)} = W_0 + s\Big(\Delta W_g^{(k)} + \Delta W_p^{(k)}\Big), \tag{2}$$

where W_0 denotes the shared pretrained backbone parameters and is kept frozen during training, $\Delta W_g^{(k)} = B_g^{(k)} A_g^{(k)}$ and $\Delta W_p^{(k)} = B_p^{(k)} A_p^{(k)}$, and s is the LoRA scaling factor. Here, $\Delta W_g^{(k)}$ is the communicable global update used for server-side aggregation, while $\Delta W_p^{(k)}$ is the client-private personalized update kept local to preserve site-specific adaptation.

Based on the Dual LoRA setting, we partition the parameters in federated training into three parts: the frozen pretrained backbone parameters θ_0, the global LoRA parameters on site k, denoted by $\phi_k = (A_g^{(k)}, B_g^{(k)})$, and the personalized LoRA parameters $\psi_k = (A_p^{(k)}, B_p^{(k)})$. At communication round t, the server broadcasts the shared global LoRA parameters ϕ^t to all sites. After receiving ϕ^t, site k initializes (ϕ_k, ψ_k) from (ϕ^t, ψ_k^t) and jointly optimizes them on D_k:

$$\big(\phi_k^{t+1}, \psi_k^{t+1}\big) = \arg\min_{\phi_k, \psi_k} \mathbb{E}_{(x,y)\sim D_k} \Big[\ell\big(f(x; \theta_0, \phi_k, \psi_k), y\big)\Big]. \tag{3}$$

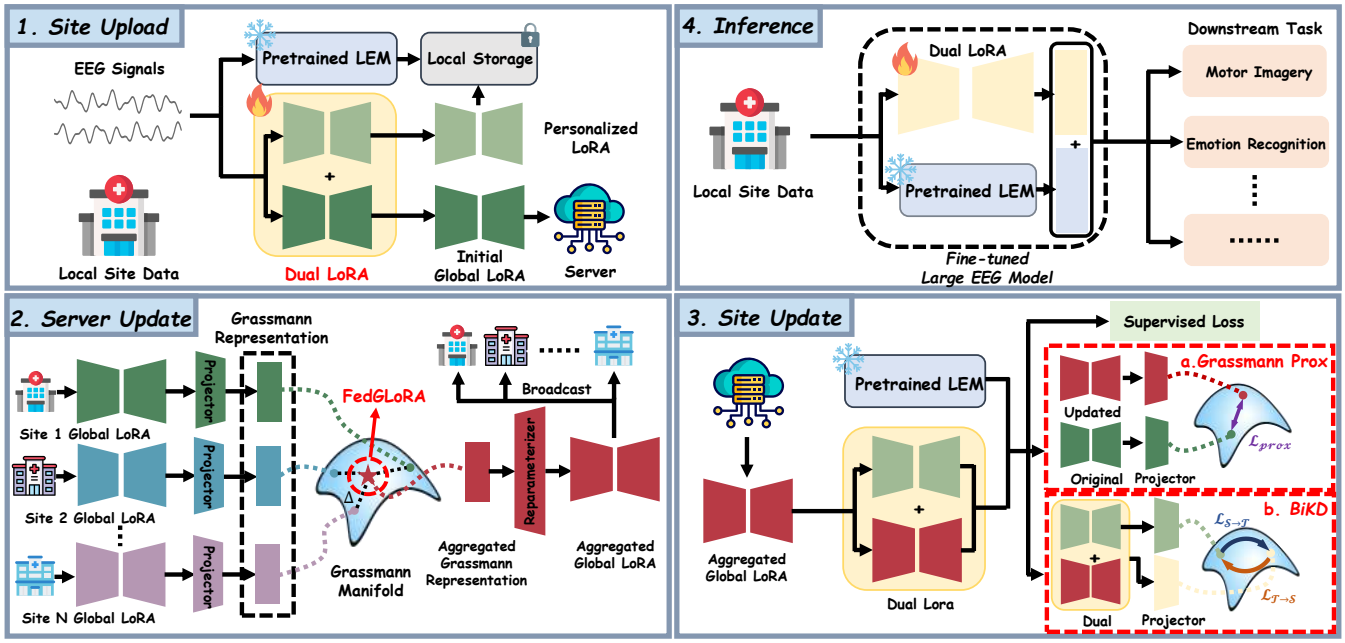

Figure 2: Framework overview of FedGLoRA.

where $\ell(\cdot)$ denotes the supervised loss function. Notably, ϕ^t denotes the global parameters broadcast by the server at round t, while ϕ_k^{t+1} is the site-updated global branch after local optimization on site k and aggregated to obtain ϕ^{t+1} for the next round.

4.3 FedGLoRA

Intuition: On the server, global LoRA updates from clients must be aggregated effectively, but FedAvg is ill-suited for LoRA as site-specific factors can be misaligned, causing cancellation and rank inflation, and the result is not guaranteed to remain a rank-r LoRA update.

Based on these considerations, we propose FedGLoRA, which operates at the subspace level, aligning and aggregating the low-rank subspaces induced by sites on the Grassmann manifold, thereby preserving the structure while improving the expressiveness of the global representation.

Grassmann Manifold. We first briefly introduce the Grassmann manifold. Given n and r, the Grassmann manifold $\mathcal{G}(n,r)$ is defined as the set of all r-dimensional linear subspaces of $\mathbb{R}^n$:

$$\mathcal{G}(n,r) = \{\mathcal{U} \subset \mathbb{R}^n : \dim(\mathcal{U}) = r\}. \tag{4}$$

Numerically, a subspace point $\mathcal{U} \in \mathcal{G}(n,r)$ can be represented by any orthonormal basis $U \in \mathbb{R}^{n\times r}$ such that $\mathcal{U} = \mathrm{col}(U)$. Since $\mathrm{col}(U) = \mathrm{col}(UQ)$ for any $Q \in \mathcal{O}(r)$, the Grassmann manifold characterizes only the subspace direction and is invariant to basis rotations.

These properties naturally align with the intrinsic non-uniqueness of LoRA factorization, since for any invertible matrix $R \in \mathbb{R}^{r\times r}$,

$$\Delta W = BA = (BR)(R^{-1}A). \tag{5}$$

Therefore, during aggregation, we lift each site update to the low-rank subspace and perform alignment on the Grassmann manifold to eliminate basis misalignment caused by LoRA reparameterization.

Grassmann Subspace Representation of LoRA. We then explain how to lift each site's Global LoRA update to a subspace representation on the Grassmann manifold. For models with multiple LoRA-injected layers, the following decomposition and aggregation are applied independently to each layer. For site k, we denote its Global LoRA update in the current round as $\Delta W_g^{(k)} \in \mathbb{R}^{d_{\text{out}}\times d_{\text{in}}}$. Rather than aggregating directly in the Euclidean coordinates, we extract the basis-invariant subspace structure. Concretely, we compute a rank-r orthogonal factorization of $\Delta W_g^{(k)}$ as:

$$\Delta W_g^{(k)} \;=\; U_k\, C_k\, V_k^\top, \tag{6}$$

where $U_k \in \mathbb{R}^{d_{\text{out}}\times r}$ and $V_k \in \mathbb{R}^{d_{\text{in}}\times r}$ satisfy $U_k^\top U_k = I_r$ and $V_k^\top V_k = I_r$. This induces rank-r output and input subspaces, $\mathcal{U}_k = \mathrm{col}(U_k) \in \mathcal{G}(d_{\text{out}}, r)$ and $\mathcal{V}_k = \mathrm{col}(V_k) \in \mathcal{G}(d_{\text{in}}, r)$, respectively. From this, we derive the coefficient matrix C_k:

$$C_k \;=\; U_k^\top\, \Delta W_g^{(k)}\, V_k \in \mathbb{R}^{r\times r}. \tag{7}$$

The coefficient matrix characterizes the magnitude and coupling of the update under the chosen subspace coordinates. Note that (U_k, V_k) determine the subspaces but not a unique basis. Therefore, $(\mathcal{U}_k, \mathcal{V}_k)$ provides a geometric representation that is invariant to basis rotations and serves as the object for subsequent alignment and aggregation.

Grassmann Subspace Distance. To quantify subspace deviation, we use the projection distance. For two output-side bases U_m and U_n, we define

$$d_{\text{proj}}^2(U_m, U_n) \;=\; \left\|U_m U_m^\top - U_n U_n^\top\right\|_F^2. \tag{8}$$

The input-side distance is defined analogously. For a candidate global subspace (U, V) and client k, the alignment discrepancy is

$$\Delta_k(U,V) \;=\; d^2_{\mathrm{proj}}(U,U_k) \;+\; d^2_{\mathrm{proj}}(V,V_k). \tag{9}$$

Aggregation Algorithm. We now describe federated aggregation in FedGLoRA. In each round, site k lifts its global LoRA increment $\Delta W_g^{(k)}$ to a Grassmann subspace representation (U_k, V_k, C_k). The server then estimates a shared rank-r global LoRA (U, V, C) that aligns client updates at the subspace level to reduce optimization conflicts and negative transfer, while faithfully reconstructing their low-rank increments within the shared subspace. To this end, we formulate aggregation as a weighted joint estimation problem on the Grassmann manifold:

$$\min_{U,\,V,\,C} \sum_{k=1}^{N} w_k \Big[\Delta_k(U,V) \;+\; \big\|U\,C\,V^\top \;-\; U_k C_k V_k^\top\big\|_F^2\Big], \tag{10}$$

where $w_k = 1/N$ is a uniform site weight in our implementation. The first term aligns the subspaces (U, V) with (U_k, V_k) on the Grassmann manifold, removing LoRA reparameterization ambiguity and reducing misalignment-induced cancellation. The second term preserves the dominant energy of client increments within the shared subspace, yielding a rank-r global update by construction.

4.4 Grassmann Prox and BiKD

Corresponding to the Site Update stage, after the server broadcasts the aggregated global LoRA, each client performs local optimization to update its global and personalized LoRA. In this stage, our goal is to strengthen transferable global representations while balancing site-specific personalization. To this end, we introduce Grassmann Manifold Prox and global-local bidirectional distillation.

Grassmann Prox

Under strong non-IID heterogeneity, local updates can drift to disparate directions, which degrades subspace alignment and destabilizes training.

Inspired by FedProx, we regularize client optimization by constraining the global LoRA subspaces to stay close to the server initialization. Specifically, at round t, client k receives reference bases (U_m, V_m) from the server; after local updates induce new bases (U_n, V_n), we define

$$\mathcal{L}^{(k)}_{\mathrm{prox}} = d^2_{\mathrm{proj}}(U_n,U_m) + d^2_{\mathrm{proj}}(V_n,V_m). \tag{11}$$

This Grassmann proximal regularization limits subspace drift across rounds, improving convergence consistency.

Bidirectional Knowledge Distillation (BiKD)

During local training, each client maintains a global and a personalized LoRA. The key challenge is balancing cross-site transfer and local personalization: direct aggregation can let global knowledge dominate personalization or allow site-specific bias to contaminate the shared representation, degrading generalization and local adaptation.

Therefore, we introduce bidirectional knowledge distillation (BiKD) to learn transferable global semantics while retaining personalized information each round. For each LoRA-injected layer, write the global LoRA increment as $\Delta W_g = U_g C_g V_g^\top$ and use its output-side basis U_g to define $P_g = U_g U_g^\top$. Let $\widetilde{U}_p$ be the output-side basis of the personalized update; orthonormalizing $(I - P_g)\widetilde{U}_p$ gives U_p and $P_p = U_p U_p^\top$. Thus, P_g and P_p act on output-side hidden activations, capturing site-shared and site-specific structure.

Global Semantic Alignment. First, during local updating at client k, we treat the Global-LoRA model as teacher f_G and the model with both global and personalized LoRA as student f_P. Let $h_G(x), h_P(x) \in \mathbb{R}^{d_{\mathrm{out}}}$ be their output-side hidden activations at this layer; f_G and f_P still denote the full predictors. We distill only on $\mathcal{U}^g$, aligning shared semantics while leaving the private subspace unconstrained as:

$$\mathcal{L}_{T\to S} = \mathbb{E}_{x\sim D_k}\left[\left\| \frac{P_g h_P(x)}{\|P_g h_P(x)\|_2} - \frac{P_g h_G(x)}{\|P_g h_G(x)\|_2} \right\|_2^2\right], \tag{12}$$

where P_g projects hidden representations onto the shared Grassmann subspace. This directional loss is robust to magnitude variation and suppresses drift.

Personalized Semantic Feedback. One-way distillation can over-transfer the student toward the global mode and weaken personalized LoRA on local private structures. We therefore feed the student's residual in $\mathcal{U}^p$ back to the teacher, allowing recurring useful private patterns to influence the global branch. Let $z^p_m(x) = g_{\mathrm{cls}}(P_p h_m(x))$ and $p^p_m(x;\tau) = \mathrm{softmax}(z^p_m(x)/\tau)$ for $m \in \{P, G\}$ and $\tau > 0$. We align the teacher to the detached student private-channel distribution via

$$\mathcal{L}_{S\to T} = \tau^2 \mathbb{E}_{x\sim D_k}\Big[\mathrm{KL}\big(\,\mathrm{sg}[p^p_P(x;\tau)] \,\|\, p^p_G(x;\tau)\big)\Big]. \tag{13}$$

Here, $\mathrm{sg}[\cdot]$ denotes stop-gradient; the projection is applied to hidden representations, while the softmax converts private-channel logits into valid predictive distributions.

Overall Local Training Objective. At round t, site k optimizes its Global LoRA and Personalized LoRA by minimizing the task loss augmented with Grassmann Prox and BiKD:

$$\begin{aligned}\mathcal{L}^{(k)}_{\mathrm{local}} = \mathbb{E}_{(x,y)\sim D_k}\big[\ell\big(f_P(x),y\big)\big] \;+\; \lambda_1 \mathcal{L}^{(k)}_{\mathrm{prox}} \\ +\; \lambda_2 \mathcal{L}_{T\to S} \;+\; \lambda_3 \mathcal{L}_{S\to T},\end{aligned} \tag{14}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$. $\mathcal{L}_{T\to S}$ and $\mathcal{L}_{S\to T}$ are defined in Eqs. (12) and (13), respectively, and $\mathcal{L}^{(k)}_{\mathrm{prox}}$ is given in Eq. (11).

5 Experiments and Results

5.1 Experimental Setup

Datasets. To comprehensively assess our method across diverse EEG application domains, we evaluate on two downstream tasks: motor imagery and emotion recognition. Specifically, for motor imagery, we use four widely adopted benchmarks including BCIC-IV-2a [Tangermann *et al.*, 2012], Cho2017 [Cho *et al.*, 2017], PhysioNet-MI [Schalk *et al.*, 2004], and Dreyer2023 [Dreyer *et al.*, 2023].

Model	Setting	Motor Imagery					Emotion Recognition				
		BCIC IV-2a	CHO 2017	PhysioNet -MI	Dreyer 2023	Avg Acc	DREAM ER	SEED-IV	FACED	GAME EMO	Avg Acc
EEGPT	Local Full-Finetune	39.50	56.21	48.13	58.72	50.64	56.99	33.32	48.07	58.39	49.20
	Local LoRA	37.83	54.92	46.31	58.26	49.33	56.09	33.02	47.96	56.54	48.40
	FedAvg	40.01	53.96	51.77	56.50	50.56	54.88	33.18	48.23	55.82	48.03
	FedProx	37.83	54.21	49.85	56.71	49.65	54.39	33.13	48.39	56.15	48.02
	SCAFFOLD	31.98	49.75	44.09	55.49	45.33	54.02	32.42	46.60	54.08	46.78
	FedKD	41.04	54.54	50.15	58.96	51.17	55.47	33.07	48.35	55.48	48.09
	FedRoD	42.90	56.04	53.94	61.24	53.53	56.17	33.16	48.40	56.82	48.64
	FDLoRA	43.99	57.38	52.58	61.76	53.93	57.03	33.26	48.82	58.05	49.29
	FedGLoRA(ours)	**49.71**	**59.54**	**57.83**	**63.27**	**57.59**	**57.74**	**33.68**	**49.37**	**61.69**	**50.62**
LaBraM	Local Full-Finetune	50.80	61.29	52.73	55.10	54.98	54.12	41.67	43.89	54.08	48.44
	Local LoRA	48.04	59.42	51.57	55.12	53.53	53.88	41.53	43.72	52.29	47.86
	FedAvg	50.10	58.25	53.64	54.18	54.04	53.99	41.05	43.06	53.02	47.78
	FedProx	45.47	58.67	52.22	53.51	52.47	52.92	41.20	42.42	53.30	47.46
	SCAFFOLD	41.10	55.21	50.35	52.30	49.74	51.95	40.22	41.60	50.11	45.97
	FedKD	49.00	58.21	53.43	55.25	53.98	53.69	41.44	42.82	53.80	47.94
	FedRoD	49.39	61.33	54.90	56.24	55.47	54.07	42.16	43.64	55.31	48.80
	FDLoRA	52.73	60.54	55.25	55.71	56.06	54.78	42.11	43.97	56.15	49.25
	FedGLoRA(ours)	**57.61**	**64.33**	**58.33**	**58.41**	**59.67**	**56.11**	**43.27**	**45.41**	**59.51**	**51.08**
CBraMod	Local Full-Finetune	45.73	62.63	58.99	63.86	57.80	59.66	40.67	51.12	60.12	52.89
	Local LoRA	44.64	61.17	56.82	63.39	56.50	59.93	40.16	50.72	56.88	51.92
	FedAvg	46.95	58.67	57.53	63.55	56.67	57.79	39.61	50.42	56.43	51.06
	FedProx	45.22	59.71	58.69	62.60	56.55	58.81	39.41	50.09	56.71	51.25
	SCAFFOLD	41.17	55.50	54.24	61.31	53.05	57.56	39.08	49.31	54.70	50.16
	FedKD	46.69	57.88	57.78	61.83	56.04	58.52	40.01	50.60	57.21	51.58
	FedRoD	49.39	61.29	59.49	63.12	58.33	60.01	40.64	50.93	58.67	52.56
	FDLoRA	53.63	62.00	61.26	64.75	60.41	59.63	41.05	51.43	59.45	52.89
	FedGLoRA(ours)	**55.04**	**66.54**	**66.61**	**65.24**	**63.36**	**61.62**	**42.45**	**53.25**	**61.02**	**54.58**

Table 1: Final local test accuracy (%) across downstream datasets. For each method, results are evaluated on each site's held-out local test set. Avg Acc is the mean over datasets within each task group. Within each model and dataset column, Top-1/Top-2/Top-3 results across *all settings* (including local) are highlighted with progressively darker shading.

For emotion recognition, we use four representative datasets: DREAMER [Katsigiannis and Ramzan, 2017], SEED-IV [Zheng *et al.*, 2018], FACED [Chen *et al.*, 2023], and GAMEEMO [Alakus *et al.*, 2020]. For each dataset, we adopt a cross-subject protocol. We randomly split subjects into disjoint train/validation/test sets with an 8:1:1 ratio, assigning all trials from each subject to a single split.

Large EEG Models (LEMs). We consider three representative LEMs: LaBraM [Jiang *et al.*, 2024], EEGPT [Wang *et al.*, 2024], and CBraMod [Wang *et al.*, 2025a]. For each model, we initialize from the publicly released pretrained checkpoint and follow the corresponding default fine-tuning configuration to ensure fair backbone comparison.

Implementation. Throughout all experiments, we apply a unified preprocessing pipeline to standardize EEG and fix the federated-learning hyperparameters to a single default setting across all models and datasets.

5.2 Cross-site generalization performance

To assess whether the cross-site domain shift hypothesized above manifests in practice, we conduct a cross-site generalization evaluation. Specifically, we train the model using data from a single site and then evaluate it on all remaining sites. As shown in Fig. 3, off-diagonal accuracies are consistently lower than the diagonal, indicating substantial performance drops under cross-site transfer for both motor imagery and emotion recognition. This pronounced gap suggests a strong distribution mismatch across sites. Therefore, we can conclude that EEG data in cross-hospital settings exhibit objectively significant site heterogeneity and domain shift. This

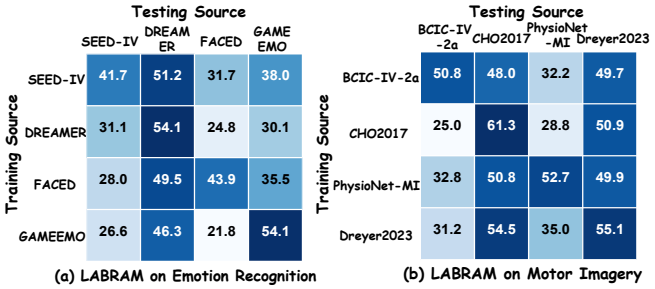

Figure 3: Cross-hospital generalization of local models. Rows denote training source and columns denote testing source. Entries report test accuracy (%), where diagonal values correspond to in-domain performance and other values quantify cross-site transfer.

observation provides key motivation for our subsequent federated learning setting: under privacy constraints, collaborative training must explicitly model and mitigate cross-site heterogeneity to improve transferability across hospitals.

5.3 Comparison with FL Baselines

As summarized in Table 1, we compare FedGLoRA with FedAvg [McMahan *et al.*, 2017], FedProx [Li *et al.*, 2020], SCAFFOLD [Karimireddy *et al.*, 2020], FedKD [Wu *et al.*, 2022], FedRoD [Chen and Chao, 2021], and FDLoRA [Qi *et al.*, 2024] on three LEMs. Reported numbers are final local test accuracies aggregated at the dataset level. FedGLoRA is consistently strong across baselines, achieving the best performance and improving the within-task-group average accu-

Model	Setting	Motor Imagery				Emotion Recognition			
		BCIC IV-2a	CHO 2017	PhysioNet -MI	Dreyer 2023	DREAM ER	SEED-IV	FACED	GAME EMO
EEGPT	Full Finetune	0.395	0.562	0.481	0.587	0.570	0.333	0.481	0.584
	$r = 2$	0.365	0.547	**0.469**	**0.584**	0.560	0.330	0.478	0.567
	$r = 4$	0.378	0.548	0.460	0.581	**0.562**	0.330	0.478	**0.570**
	$r = 8$	**0.379**	0.549	0.463	0.583	0.561	0.330	**0.480**	0.565
	$r = 16$	0.369	**0.552**	0.467	0.580	**0.562**	**0.331**	**0.480**	0.567
LaBraM	Full Finetune	0.508	0.613	0.527	0.551	0.541	0.417	0.439	0.541
	$r = 2$	0.471	0.587	**0.520**	0.550	0.537	0.402	0.436	0.527
	$r = 4$	0.473	0.591	0.518	0.550	0.537	0.403	0.436	0.527
	$r = 8$	**0.480**	**0.594**	0.516	0.551	**0.539**	0.403	**0.437**	0.523
	$r = 16$	**0.480**	0.588	0.518	**0.552**	**0.539**	**0.404**	0.434	**0.544**
CBraMod	Full Finetune	0.457	0.626	0.590	0.639	0.597	0.407	0.511	0.601
	$r = 2$	0.443	0.610	0.572	0.633	0.594	0.400	0.508	**0.576**
	$r = 4$	0.443	**0.612**	**0.575**	**0.634**	0.594	0.401	0.508	0.569
	$r = 8$	**0.449**	**0.612**	0.568	0.630	**0.601**	**0.403**	0.507	0.569
	$r = 16$	0.446	0.611	0.569	0.630	0.598	0.400	**0.509**	0.572

Table 2: Fine-tuning accuracy of three LEMs with different LoRA ranks r compared to full fine-tuning across motor imagery and emotion recognition datasets. Higher is better; bold denotes the best LoRA result within each model and dataset.

racy for both motor imagery and emotion recognition. Notably, it also outperforms single-site full fine-tuning, indicating that cross-site collaboration provides more effective information than isolated local updates. FDLoRA achieves the second-highest performance overall. FedAvg offers limited gains, while FedProx and SCAFFOLD improve performance in some cases but are not consistently effective across models and task groups. These results support our motivation: under cross-site non-IID and imbalance, naive aggregation suffers from update misalignment and drift, whereas FedGLoRA's subspace alignment yields transferable global updates.

5.4 Effectiveness of LoRA for Fine-tuning LEMs

To demonstrate the effectiveness of LoRA for fine-tuning LEMs, we compare LoRA with full fine-tuning on three LEMs across eight datasets. We further vary the LoRA rank $r \in \{2, 4, 8, 16\}$ to assess accuracy under different parameter budgets. As shown in Table 2, LoRA achieves performance comparable to full fine-tuning on most datasets. These results confirm that LoRA provides an accuracy-preserving adaptation strategy with substantially fewer trainable parameters. This supports FedGLoRA: small site updates learn useful EEG cues without moving the full backbone during training. We thus use $r = 8$ to balance accuracy and efficiency.

5.5 Ablation Study

As shown in Table 3, we conduct ablation studies of FedGLoRA. Dual LoRA is essential: using only the personalized branch causes a large accuracy drop, while using only the global branch improves performance but remains inferior, confirming complementary shared transfer and local adaptation. Removing FedGLoRA yields the largest degradation, highlighting the importance of subspace alignment under non-IID update misalignment. Grassmann-Prox improves performance by limiting subspace drift. Bidirectional distillation further boosts accuracy: removing BiKD degrades results, and either one-way variant is weaker than the bidirectional design, indicating that both global-to-local alignment and local-to-global feedback are needed. Overall, FedGLoRA drives the primary accuracy gains, while Grassmann-Prox and BiKD provide additional improvements.

(a) Motor Imagery (MI) with LaBraM.

Method	BCIC-IV 2a	CHO2017	PhysioNet-MI	DRE2023	Avg Acc.
Full	**57.61**	**64.33**	**58.33**	**58.41**	**59.67**
Only Personalized LoRA	48.04	59.42	51.57	55.12	53.53
Only Global LoRA	53.11	60.38	55.20	56.70	56.35
w.o. FedGLoRA	48.75	59.21	54.24	55.58	54.45
w.o. Grassmann Prox	55.36	63.46	56.82	58.11	58.44
w.o. BiKD	54.14	60.25	55.15	57.19	56.68
only Global-Local	55.17	60.46	55.71	57.49	57.21
only Local-Global	54.40	61.21	54.60	57.77	56.99

(b) Emotion Recognition with CBraMod.

Method	DREAMER	SEED-IV	FACED	GAMEEMO	Avg Acc.
Full	**61.62**	**42.45**	**53.25**	**61.02**	**54.58**
Only Personalized LoRA	53.88	41.53	43.72	52.29	47.86
Only Global LoRA	58.41	40.95	51.62	57.55	52.13
w.o. FedGLoRA	57.98	40.16	50.91	55.03	51.02
w.o. Grassmann Prox	60.26	41.99	52.84	59.73	53.71
w.o. BiKD	58.87	41.32	51.54	58.05	52.45
only Global-Local	59.30	41.63	52.41	59.12	53.11
only Local-Global	59.18	41.47	52.89	58.67	53.05

Table 3: Ablation study on (a) Motor Imagery (MI) and (b) Emotion Recognition. Metrics are accuracy (%).

6 Conclusion

In this work, we observe that EEG dataset co-evolve with the research community, inducing cross-site drift and fragmented deployments that require LEM adaptation. We formulate federated fine-tuning around two key challenges: multi-source heterogeneity and scale imbalance. To address them, we propose FedGLoRA, a Dual LoRA framework that separates global transfer from local personalization, aligns global updates on basis-invariant Grassmann subspaces, and introduces Grassmann Prox and BiKD. Across motor imagery and emotion recognition with three LEMs, FedGLoRA consistently outperforms strong federated baselines, and ablations confirm the complementary effects of its components. FedGLoRA offers a route for privacy-preserving cross-site LEM adaptation.

Acknowledgments

This work is supported by the Youth Science Fund Project of National Natural Science Foundation of China (Grant No.62306317), and the Open Research Fund of the Key Laboratory of Social Computing and Cognitive Intelligence (Dalian University of Technology), Ministry of Education (Grant No.SCCI2025YB01), and Sponsored by Beijing Nova Program (Grant No.20250484804), and Young Elite Scientists Sponsorship Program of the Beijing High Innovation Plan (Grant No.20250912).

References

[Alakus *et al.*, 2020] Talha Burak Alakus, Murat Gonen, and Ibrahim Turkoglu. Database for an emotion recognition system based on eeg signals and various computer games–gameemo. *Biomedical Signal Processing and Control*, 60:101951, 2020.

[Baykara *et al.*, 2025] Cem Ata Baykara, Saurav Raj Pandey, Ali Burak Ünal, Harlin Lee, and Mete Akgün. Federated learning for epileptic seizure prediction across heterogeneous eeg datasets. *arXiv preprint arXiv:2508.08159*, 2025.

[Bian *et al.*, 2025] Jieming Bian, Yuanzhe Peng, Lei Wang, Yin Huang, and Jie Xu. A survey on parameter-efficient fine-tuning for foundation models in federated learning. *arXiv preprint arXiv:2504.21099*, 2025.

[Chen and Chao, 2021] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning for image classification. *arXiv preprint arXiv:2107.00778*, 2021.

[Chen and Yu, 2026] Qianyu Chen and Shujian Yu. Continual learning for fmri-based brain disorder diagnosis via functional connectivity matrices generative replay. *arXiv preprint arXiv:2604.14259*, 2026.

[Chen *et al.*, 2023] Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023.

[Cho *et al.*, 2017] Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 2017.

[Cooray *et al.*, 2025] Lakshan Cooray, Janaka Sendanayake, Pramuditha Vithanaarachchi, and YHPP Priyadarshana. Deep federated learning: a systematic review of methods, applications, and challenges. *Frontiers in Computer Science*, 7:1617597, 2025.

[Dreyer *et al.*, 2023] Pauline Dreyer, Aline Roc, Léa Pillette, Sébastien Rimbert, and Fabien Lotte. A large eeg database with users' profile information for motor imagery brain-computer interface research. *Scientific Data*, 10(1):580, 2023.

[Fu *et al.*, 2024] Baole Fu, Wenhao Chu, Chunrui Gu, and Yinhua Liu. Cross-modal guiding neural network for multimodal emotion recognition from eeg and eye movement signals. *IEEE Journal of Biomedical and Health Informatics*, 28(10):5865–5876, 2024.

[Gama *et al.*, 2014] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37, 2014.

[Guo *et al.*, 2021] Pengfei Guo, Puyang Wang, Jinyuan Zhou, Shanshan Jiang, and Vishal M Patel. Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2423–2432, 2021.

[He and Wu, 2019] He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019.

[Huo *et al.*, 2026] Yujia Huo, Jianchun Liu, Hongli Xu, Zhenguo Ma, Shilong Wang, and Liusheng Huang. Mitigating catastrophic forgetting with adaptive transformer block expansion in federated fine-tuning. *IEEE Transactions on Mobile Computing*, 2026.

[Jayaram and Barachant, 2018] Vinay Jayaram and Alexandre Barachant. Moabb: trustworthy algorithm benchmarking for bcis. *Journal of neural engineering*, 15(6):066011, 2018.

[Jia *et al.*, 2024] Tianwang Jia, Lubin Meng, Siyang Li, Jiajing Liu, and Dongrui Wu. Federated motor imagery classification for privacy-preserving brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:3442–3451, 2024.

[Jiang *et al.*, 2024] Wei-Bang Jiang, Liming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *International Conference on Learning Representations*, volume 2024, pages 16405–16426, 2024.

[Karimireddy *et al.*, 2020] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

[Katsigiannis and Ramzan, 2017] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017.

[Kuruppu *et al.*, 2026] Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, and Yogatheesan Varatharajah. Eeg foundation models: a critical review of current progress and future directions. *Journal of neural engineering*, 23(2):021001, 2026.

[Li *et al.*, 2020] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith.

Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[Liu *et al.*, 2024a] Chenyu Liu, Xinliang Zhou, Zhengri Zhu, Liming Zhai, Ziyu Jia, and Yang Liu. Vbh-gnn: Variational bayesian heterogeneous graph neural networks for cross-subject emotion recognition. In *ICLR*, 2024.

[Liu *et al.*, 2024b] Rui Liu, Yuanyuan Chen, Anran Li, Yi Ding, Han Yu, and Cuntai Guan. Aggregating intrinsic information to enhance bci performance through federated learning. *Neural Networks*, 172:106100, 2024.

[Liu *et al.*, 2025] Chenyu Liu, Yuqiu Deng, Tianyu Liu, Jinan Zhou, Xinliang Zhou, Ziyu Jia, and Yi Ding. Echo: Toward contextual seq2seq paradigms in large eeg models. *arXiv preprint arXiv:2509.22556*, 2025.

[Luo *et al.*, 2021] Mi Luo, Fei Chen, Dapeng Hu, Yifan Zhang, Jian Liang, and Jiashi Feng. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data. *Advances in Neural Information Processing Systems*, 34:5972–5984, 2021.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. Pmlr, 2017.

[Mongardi and Pinoli, 2024] Sofia Mongardi and Pietro Pinoli. Exploring federated learning for emotion recognition on brain-computer interfaces. In *Adjunct proceedings of the 32nd ACM conference on user modeling, adaptation and personalization*, pages 622–626, 2024.

[Qi *et al.*, 2024] Jiaxing Qi, Zhongzhi Luan, Shaohan Huang, Carol Fung, Hailong Yang, and Depei Qian. Fdlora: Personalized federated learning of large language model via dual lora tuning. *arXiv preprint arXiv:2406.07925*, 2024.

[Rieke *et al.*, 2020] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *NPJ digital medicine*, 3(1):119, 2020.

[Schalk *et al.*, 2004] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

[Tang *et al.*, 2026] Yuting Tang, Weibang Jiang, Shanglin Li, Yong Li, Chenyu Liu, Xinliang Zhou, Yi Ding, and Cuntai Guan. Eeg-dlite: Dataset distillation for efficient large eeg model training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 17733–17741, 2026.

[Tangermann *et al.*, 2012] Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot R Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

[Wang *et al.*, 2020] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33:7611–7623, 2020.

[Wang *et al.*, 2024] Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. *Advances in Neural Information Processing Systems*, 37:39249–39280, 2024.

[Wang *et al.*, 2025a] Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. In *International conference on learning representations*, volume 2025, pages 75310–75346, 2025.

[Wang *et al.*, 2025b] Zheng Wang, Zihui Wang, Xiaoliang Fan, and Cheng Wang. Federated learning with domain shift eraser. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4978–4987, 2025.

[Wu *et al.*, 2022] Chuhan Wu, Fangzhao Wu, Lingjuan Lyu, Yongfeng Huang, and Xing Xie. Communication-efficient federated learning via knowledge distillation. *Nature communications*, 13(1):2032, 2022.

[Wu *et al.*, 2024] Di Wu, Siyuan Li, Jie Yang, and Mohamad Sawan. Neuro-bert: Rethinking masked autoencoding for self-supervised neurological pretraining. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[Xu *et al.*, 2020] Lichao Xu, Minpeng Xu, Yufeng Ke, Xingwei An, Shuang Liu, and Dong Ming. Cross-dataset variability problem in eeg decoding with deep learning. *Frontiers in human neuroscience*, 14:103, 2020.

[Yang *et al.*, 2025] Yiyuan Yang, Guodong Long, Qinghua Lu, Liming Zhu, Jing Jiang, and Chengqi Zhang. Federated low-rank adaptation for foundation models: A survey. *arXiv preprint arXiv:2505.13502*, 2025.

[Zhang *et al.*, 2022] Jie Zhang, Zhiqi Li, Bo Li, Jianghe Xu, Shuang Wu, Shouhong Ding, and Chao Wu. Federated learning with label distribution skew via logits calibration. In *International Conference on Machine Learning*, pages 26311–26329. PMLR, 2022.

[Zhang *et al.*, 2025] Jiayang Zhang, Kang Li, Banghua Yang, and Zhengrun Zhao. Cross-dataset motor imagery decoding—a transfer learning assisted graph convolutional network approach. *Biomedical Signal Processing and Control*, 102:107213, 2025.

[Zheng *et al.*, 2018] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49(3):1110–1122, 2018.

[Zhou *et al.*, 2025] Xinliang Zhou, Chenyu Liu, Zhisheng Chen, Kun Wang, Yi Ding, Ziyu Jia, and Qingsong Wen. Brain foundation models: A survey on advancements in neural signal processing and brain discovery. *IEEE Signal Processing Magazine*, 42(5):22–35, 2025.