EVOBRAIN: CONTINUAL LEARNING OF EEG FOUNDATION MODELS ACROSS HETEROGENEOUS BCI TASKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalography (EEG) is the predominant modality for non-invasive Brain-Computer Interfaces (BCIs) due to its high temporal resolution. However, conventional decoding approaches rely on fragmented, task-specific architectures, limiting scalability across heterogeneous BCI tasks. EEG foundation models pre-trained on large-scale corpora offer a promising solution for universal brain decoding, yet current post-training strategies remain dominated by task-isolated fine-tuning. This static paradigm restricts knowledge transfer across tasks and incurs computational and storage costs that grow linearly with the number of downstream tasks. In this paper, we formulate downstream adaptation of EEG foundation models as a cross-task continual learning problem and propose EvoBrain, a dynamic task-aware post-training framework that enables a unified foundation model to continually generalize across diverse EEG tasks. EvoBrain addresses the plasticity-stability trade-off through two complementary components. Neuro-Spectral Task Normalization (NSN) facilitates adaptation to distributional and neuro-spectral shifts by aligning task statistics and recalibrating spectral responses. Response-Affinity Distillation (RAD), together with time-dependent replay, preserves historical task knowledge while promoting selective transfer between spectrally compatible tasks, thereby mitigating catastrophic forgetting. Extensive evaluations on six distinct BCI tasks demonstrate that our approach consistently surpasses state-of-the-art methods across diverse foundation backbones, effectively balancing the plasticity-stability trade-off.

1 INTRODUCTION

Brain-computer interfaces (BCIs) provide a direct pathway for translating neural activity into control signals for external devices Wu et al. (2016); Schalk et al. (2004). As the predominant non-invasive BCI technique, electroencephalography (EEG) records cortical electrical activity with high temporal resolution, enabling its widespread deployment across practical and clinical applications Cowie et al. (2001); Jeong (2004); Jenke et al. (2014). Early studies typically relied on hand-crafted features combined with machine learning algorithms to analyze EEG signals Lotte et al. (2007). With the advances in deep learning, a variety of models have been proposed to automatically decode EEG patterns and identify brain states across diverse experimental paradigms, including but not limited to sleep staging Zhou et al. (2025a), emotion recognition Ding et al. (2022), and motor imagery Altaheri et al. (2023). These models are typically task-specific with different architectures, and are trained and evaluated on the same datasets, which inherently constrains their generalization capability and scalability across heterogeneous EEG tasks. To overcome this bottleneck, recent advancements have introduced EEG foundation models Kostas et al. (2021); Yang et al. (2023); Zhang et al. (2023); Jiang et al. (2024a); Wang et al. (2024a); Shi et al. (2024); Jiang et al. (2024b); Wang et al. (2024b; 2025); Zhou et al. (2025b); Ding et al. (2025); Chen et al. (2025a); Wang et al. (2026), which are pretrained in a self-supervised manner on large-scale unlabeled EEG data to learn universal, task-agnostic neural representations. In this manner, EEG foundation models can utilize a unified backbone architecture that requires only task-specific classifiers to be attached, and can then be effectively adapted to various downstream tasks via post-training (i.e., fine-tuning), achieving performance comparable to or even surpassing that of traditional task-specific models.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

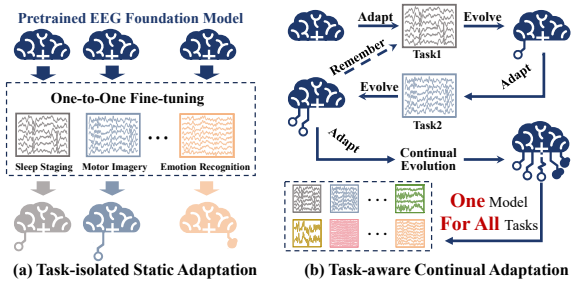

Figure 1: (a) The traditional task-isolated static post-training paradigm; (b) the task-aware continual adaptation strategy for a universal EEG decoding model.

Despite the success of EEG foundation models, current research disproportionately prioritizes pre-training, leaving post-training adaptation largely underexplored. As illustrated in Fig. 1(a), even with a unified backbone, existing methods predominantly resort to task-isolated full fine-tuning, a practice that imposes two fundamental limitations: **First, task-isolated adaptation inherently hinders cross-task knowledge transfer.** Unlike large language models (LLMs) that exhibit emergent abilities through scaling or cross-task learning Wei et al. (2022); Schaeffer et al. (2023), EEG foundation models under current post-training paradigms remain confined to narrow task boundaries. This isolation prevents the model from synthesizing higher-order, invariant neural patterns, thereby severely constraining its generalization potential beyond individual tasks. **Second, this static adaptation strategy struggles to accommodate the dynamic expansion requirements of EEG foundation models.** Adapting to a new EEG task necessitates a complete re-training from the pre-trained initialization and the storage of a separate set of parameters. Consequently, computational overhead and storage demands scale linearly with the number of tasks, creating prohibitive bottlenecks for scalable, multi-task deployment. In summary, current static and task-isolated post-training strategies fundamentally hinder the realization of universal brain decoding models. There is an urgent need for a dynamic and task-aware post-training strategy that enables a single, unified EEG foundation model to generalize across diverse EEG tasks, ultimately achieving the one-for-all objective. Critically, this dynamic adaptation naturally formulates a Continual Learning (CL) problem Wang et al. (2024c), wherein the foundation model incrementally acquires knowledge from a sequence of heterogeneous EEG tasks without catastrophic forgetting. To succeed, the framework must effectively maintain an appropriate trade-off between learning plasticity (i.e., the capacity to adapt to new tasks) and memory stability (i.e., the ability to retain knowledge on prior tasks).

However, enabling a pretrained EEG foundation model to continually adapt to sequentially arriving different BCI tasks while preserving previously acquired capabilities remains highly challenging. Two coupled obstacles are central to this setting. First, learning plasticity is limited by the multi-level heterogeneity of EEG task streams. Different BCI paradigms may differ in acquisition settings, label semantics, signal morphology, and task-specific neuro-spectral patterns. Therefore, effective adaptation must go beyond generic distribution alignment to adapt task-specific temporal statistics and recalibrate task-relevant spectral responses. Second, memory stability is challenged by continual updates that may distort the response organization learned from previous tasks. In heterogeneous EEG streams, transferable structures are often only partially shared, and indiscriminate alignment across unrelated paradigms can induce negative transfer. Thus, continual adaptation must preserve old-task response relations while enabling transfer between tasks with compatible neural signatures.

To address these challenges, **we propose EvoBrain, a cross-task continual learning framework for adapting pretrained EEG foundation models toward one-for-all brain decoding.** EvoBrain is built upon two complementary principles. First, Neuro-Spectral Task Normalization (NSN) enhances learning plasticity by aligning incoming tasks with accumulated historical statistics while recalibrating task-specific spectral responses. Second, Response-Affinity Distillation (RAD), together with time-dependent replay, improves memory stability by preserving old-task response organization and promoting selective knowledge transfer between spectrally compatible tasks. In this way, EvoBrain enables a single EEG foundation backbone to evolve across heterogeneous BCI tasks without repeatedly restarting task-isolated fine-tuning. The framework is also plug-and-play, allowing integration with both Transformer-based and Mamba-based EEG foundation models without requiring backbone-specific architectural modifications. The key contributions of this paper are summarized as follows:

- **To the best of our knowledge, this work pioneers the exploration of cross-task continual learning in EEG-related analysis.** We propose a novel framework, termed EvoBrain, to overcome the critical limitation of static, task-isolated adaptation in pretrained EEG

foundation models. The framework enables a single unified model to dynamically generalize across heterogeneous EEG tasks, thereby advancing toward the long-sought one-for-all objective for universal brain decoding.

- **We design Neuro-Spectral Task Normalization (NSN) and Response-Affinity Distillation (RAD) to address the plasticity-stability trade-off in cross-task continual EEG decoding.** NSN enhances adaptation to heterogeneous incoming tasks by jointly stabilizing temporal statistics and recalibrating task-specific spectral responses, while RAD preserves old-task response organization and promotes affinity-aware knowledge transfer across spectrally compatible tasks.
- **By integrating our framework with four mainstream EEG foundation models of diverse architectures**, we demonstrate its architecture-agnostic efficacy, consistently improving adaptation plasticity and memory stability across six EEG tasks.

2 METHODOLOGY

2.1 PROBLEM SETUP AND PRELIMINARIES

This work aims to enable a pretrained EEG foundation model to continually adapt to a sequence of heterogeneous EEG decoding tasks while preserving its performance on previously encountered ones. We formalize this setting as cross-task continual learning. Formally, let $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_\mathcal{N}\}$ denote a stream of $\mathcal{N}$ distinct EEG tasks (e.g., sleep staging, motor imagery, emotion recognition). Each task $\mathcal{T}_i$ is associated with a unique data distribution $\mathcal{D}_i$, where $\mathcal{D}_i \neq \mathcal{D}_j$ and $1 \leq i \neq j \leq \mathcal{N}$, reflecting substantial domain shifts in signal statistics and label spaces across tasks. Let $\mathcal{M}$ denote the continual learning model, where $\mathcal{M}_0$ represents the initial pretrained EEG foundation model and $\mathcal{M}_i$ denotes the updated model after adaptation to the i-th task $\mathcal{T}_i$. Our core objective is to simultaneously optimize two competing criteria, as formally described below:

(1) **Learning plasticity:** The continual learning model $\mathcal{M}_i$ should be able to adapt effectively to a new task $\mathcal{T}_i$.

$$\min_{\theta_\mathcal{M}} (\mathbb{E}_{(x,y)\sim\mathcal{D}_i} \mathcal{L}(\mathcal{M}_i(x), y)) \tag{1}$$

where $\mathcal{M}$ is parameterized by $\theta_\mathcal{M}$ and $\mathcal{L}$ denotes a task-specific loss that measures the prediction error of the model $\mathcal{M}_i$ on the given task $\mathcal{T}_i$.

(2) **Memory stability:** The updated model $\mathcal{M}_i$ should retain performance on all previously learned tasks $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_j\}$, i.e., for all $j < i$.

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_j} [\mathcal{L}(\mathcal{M}_j(x), y)] \approx \mathbb{E}_{(x,y)\sim\mathcal{D}_j} [\mathcal{L}(\mathcal{M}_i(x), y)] \tag{2}$$

The fundamental challenge lies in mitigating catastrophic forgetting, the tendency of $\mathcal{M}_i$ to overfit to $\mathcal{D}_i$ and degrade performance on $\mathcal{D}_j$ ($j < i$), while effectively acquiring new knowledge from $\mathcal{D}_i$. This necessitates a dynamic trade-off between plasticity for adaptation and stability for memory retention. To this end, we formulate the continual learning objective under task $\mathcal{T}_i$ as follows:

$$\min_{\mathcal{M}_i} \underbrace{\mathbb{E}_{(x,y)\sim\mathcal{D}_i} \mathcal{L}(\mathcal{M}_i(x), y)}_{\text{New Task adaptation}} + \underbrace{\mathcal{R}\big(\mathcal{M}_i; \{\mathcal{M}_j\}_{j=1}^{i-1}, \{\mathcal{D}_j\}_{j=1}^{i-1}\big)}_{\text{Historical knowledge retention}} \tag{3}$$

where $\mathcal{R}(\cdot)$ is a generic regularization term that penalizes deviations from the knowledge encoded in previously learned models $\{\mathcal{M}_j\}_{j=1}^{i-1}$ with respect to their corresponding data distributions $\{\mathcal{D}_j\}_{j=1}^{i-1}$. Our goal is to learn a universal EEG decoder $\mathcal{M}_\mathcal{N}$ that generalizes across all learned $\mathcal{N}$ tasks without requiring retraining, thus realizing the one model for all tasks vision.

2.2 OVERVIEW

EvoBrain is a continual adaptation framework that evolves a pretrained EEG foundation model across a stream of heterogeneous BCI tasks, as illustrated in Fig. 2. Starting from $\mathcal{M}_0$, the backbone is sequentially updated from $\mathcal{M}_{i-1}$ to $\mathcal{M}_i$ as each task $\mathcal{T}_i$ arrives, with a lightweight task-specific classifier head attached for each task. EvoBrain combines Neuro-Spectral Task Normalization (NSN) to adapt task-dependent EEG distributions with replay-based Response-Affinity Distillation (RAD) to preserve prior knowledge and facilitate cross-task transfer. At inference, only

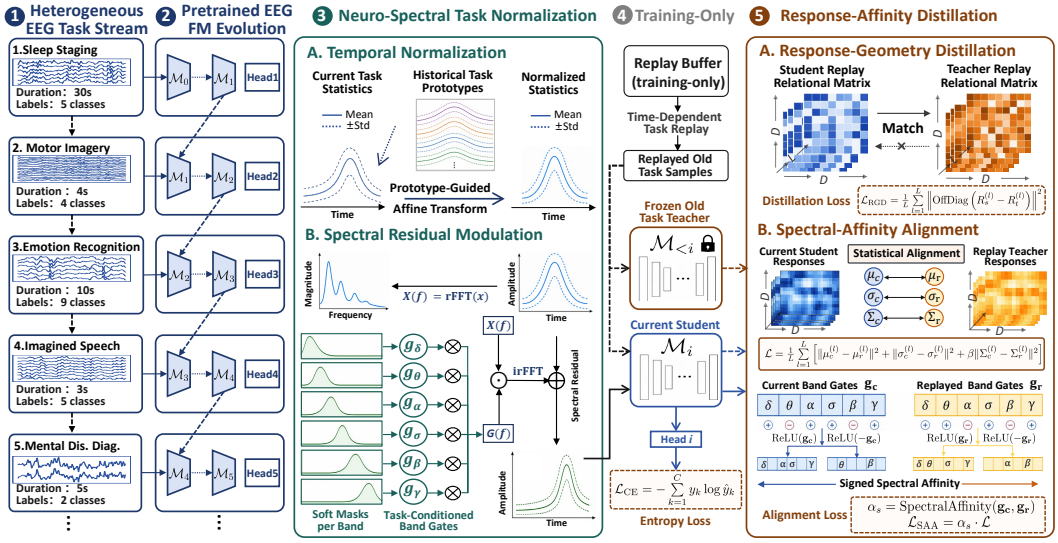

Figure 2: **Overview of EvoBrain for continual adaptation of EEG foundation models across heterogeneous BCI tasks.** EvoBrain progressively adapts a pretrained backbone through Neuro-Spectral Task Normalization (NSN) and Response-Affinity Distillation (RAD), which jointly mitigate cross-task shifts, preserve previous-task knowledge, and promote transferable representations. Only the adapted backbone and task-specific heads are retained for inference.

the adapted backbone and corresponding task head are retained. Notably, EvoBrain is architecture-agnostic and can be readily integrated into mainstream EEG foundation models without complex architectural modifications.

2.3 Time-Dependent Task Replay

EvoBrain employs a lightweight time-dependent replay strategy with an index-based catalog, enabling historical sample replay without additional storage overhead.. Since earlier tasks undergo more subsequent updates and are typically more vulnerable to forgetting, we assign larger replay probabilities to older tasks instead of sampling all stored tasks uniformly. Let $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{i-1}\}$ be the historical tasks before learning $\mathcal{T}_i$. For each task $\mathcal{T}_t$, its temporal depth and replay weight are:

$$d_t = i - t + 1, \qquad w_t = \left(1 - e^{-0.8 \cdot d_t}\right) \tag{4}$$

For each current-task minibatch, we sample a historical task according to w_t and uniformly draw a same-sized replay minibatch from its training catalog. The replay samples are used to construct the response-preservation and response-alignment objectives using the corresponding frozen historical teacher model $\mathcal{M}_{j<i}$. Both the replay catalog and historical teachers are used only during training; inference retains the final continually adapted backbone, the learned NSN parameters, and the task-specific classifiers.

2.4 Neuro-Spectral Task Normalization

Heterogeneous BCI tasks exhibit substantial shifts in both signal distributions and neuro-spectral response patterns, challenging the plasticity and stability of continual adaptation. To address these task-dependent variations, we propose *Neuro-Spectral Task Normalization* (NSN), which combines Prototype-guided Temporal Normalization (PTN) to stabilize cross-task statistics with task-conditioned Spectral Residual Modulation (SRM) to recalibrate task-relevant frequency responses.

Prototype-guided Temporal Normalization: Given an input tensor $\mathbf{X} \in \mathbb{R}^{\mathbf{B}\times\mathbf{C}\times\mathbf{D}\times d}$ from the current task $\mathcal{T}_i$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ denote the batch size, number of channels, and duration of EEG segments, respectively, while d denotes the number of temporal samples within each one-second EEG patch. For each task $\mathcal{T}_i$, we maintain a historical task prototype $\mathbf{p}_i \in \mathbb{R}^d$, which is updated as an exponential moving average of the task-wise data mean of $\mathbf{X}$ over the batch, channel, and

duration dimensions during the first training epoch. When a new input arrives, we compute the current-task statistics (μ_i, σ_i) from $\mathbf{X}$ and derive the historical reference statistics (μ_g, σ_g) from the set of previous task prototypes $\mathcal{P}_{<i} = \{\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}\}$. Here, (μ_g, σ_g) summarize the historical feature distribution accumulated before learning $\mathcal{T}_i$.

Instead of forcing the current task $\mathcal{T}_i$ to rigidly match the historical statistics of $\mathcal{T}_{<i}$, we utilize learnable task-specific moment-gating parameters $\eta_\mu^i, \eta_\sigma^i \in \mathbb{R}$ to dynamically rescale and shift the i-th task data. We define gating coefficients as follows:

$$\lambda_\mu^i = \sigma(\eta_\mu^i), \qquad \lambda_\sigma^i = \sigma(\eta_\sigma^i), \tag{5}$$

where $\sigma(\cdot)$ denotes the sigmoid function. The target moments are then obtained by

$$\mu_i^{\mathrm{tar}} = \lambda_\mu^i \mu_g + (1 - \lambda_\mu^i)\mu_i, \qquad \sigma_i^{\mathrm{tar}} = \lambda_\sigma^i \sigma_g + (1 - \lambda_\sigma^i)\sigma_i. \tag{6}$$

The affine transformation parameters are defined as

$$\gamma_i = \frac{\sigma_i^{\mathrm{tar}}}{\sigma_i}, \qquad \beta_i = \mu_i^{\mathrm{tar}} - \gamma_i \mu_i, \tag{7}$$

and the current-task input $\mathbf{X}$ is normalized as

$$\mathbf{X}_{\mathrm{tn}} = \gamma_i \mathbf{X} + \beta_i. \tag{8}$$

Task-conditioned Spectral Residual Modulation: After temporal normalization, we further introduce spectral residual modulation to explicitly model task-specific spectral response patterns. Given the temporally normalized representation $\mathbf{X}_{\mathrm{tn}}$, we first compute the real-valued Fourier spectrum along the within-patch temporal dimension using rFFT, with the resulting frequency bins determined by the 200-Hz sampling rate.

$$\widehat{\mathbf{X}}(f) = \mathrm{rFFT}(\mathbf{X_{tn}}). \tag{9}$$

Let $\mathcal{B}$ denote the standard EEG bands: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), σ (12–16 Hz), β (16–30 Hz), and γ (30–45 Hz). For each band $b \in \mathcal{B}$ with lower and upper boundaries (l_b, u_b), we construct a soft band mask $M_b(f)$ with a smooth transition width $\tau = 1$ Hz:

$$M_b(f) = \begin{cases} 1, & l_b \le f \le u_b, \\ \frac{1}{2} - \frac{1}{2}\cos\left(\pi \frac{f - (l_b - \tau)}{\tau}\right), & l_b - \tau \le f < l_b, \\ \frac{1}{2} + \frac{1}{2}\cos\left(\pi \frac{f - u_b}{\tau}\right), & u_b < f \le u_b + \tau, \\ 0, & \text{otherwise}. \end{cases} \tag{10}$$

This construction assigns full weights to frequencies inside each EEG band and smoothly attenuates frequencies near band boundaries. To handle overlapping frequency regions, the masks are normalized across bands as follows:

$$W_b(f) = \frac{M_b(f)}{\sum_{b' \in \mathcal{B}} M_{b'}(f) + \epsilon} \tag{11}$$

where ϵ is a small positive constant for numerical stability. For the current task, we assign a learnable parameter a_b to each band $b \in \mathcal{B}$ and obtain the corresponding band gate by

$$g_b = \tanh(a_b), \qquad b \in \mathcal{B}, \tag{12}$$

where $g_b \in [-1, 1]$ controls whether the response of band b is enhanced or suppressed. The band-level gates are projected to frequency bins to form the modulation function $G(f)$, which generates the residual spectrum as

$$G(f) = \sum_{b \in \mathcal{B}} g_b W_b(f), \qquad \Delta\widehat{\mathbf{X}}(f) = G(f) \odot \widehat{\mathbf{X}}(f). \tag{13}$$

The modulated spectrum is transformed back to the feature domain and added residually to the normalized features:

$$\Delta\mathbf{X} = \mathrm{irFFT}\left(\Delta\widehat{\mathbf{X}}(f)\right), \qquad \mathbf{X}_{\mathrm{nsn}} = \mathbf{X}_{\mathrm{tn}} + \alpha_m \Delta\mathbf{X}, \tag{14}$$

where α_m controls the modulation strength. By preserving the temporally normalized representation, this residual formulation enables task-adaptive enhancement or suppression of relevant spectral responses. Meanwhile, the learned band gate g_b serves as a compact task-specific spectral signature for subsequent cross-task affinity estimation.

2.5 RESPONSE-AFFINITY DISTILLATION

After NSN, sequential adaptation still requires preserving old-task response structure while transferring shared representations across heterogeneous tasks. To this end, we propose *Response-Affinity Distillation* (RAD), comprising Response-Geometry Distillation (RGD) for memory stability and Spectral-Affinity Alignment (SAA) for cross-task transfer. When learning a new task $\mathcal{T}_i$, EvoBrain samples replay data from a previous task $\mathcal{T}_j$ $(j < i)$ using the time-dependent replay strategy and loads the corresponding frozen checkpoint $\mathcal{M}_j$, saved immediately after learning $\mathcal{T}_j$, as the teacher. The replayed samples are fed into both the current student model $\mathcal{M}_i$ and the frozen teacher model $\mathcal{M}_{<i}$. For the ℓ-th encoder layer, we denote their intermediate responses as $\mathbf{H}_{s,r}^{(\ell)}$ and $\mathbf{H}_{t,r}^{(\ell)}$, respectively, where the subscripts s, t, and r indicate student, teacher, and replayed old-task samples. Similarly, the response of current-task samples in the student model is denoted as $\mathbf{H}_{s,c}^{(\ell)}$. We then apply global average pooling over the channel and duration dimensions to these intermediate responses, yielding scale-consistent representations in $\mathbb{R}^{\mathbf{B}\times d}$ across heterogeneous EEG tasks for subsequent relation and statistic computation.

Response-Geometry Distillation: RGD preserves old-task stability by matching the response geometry of replayed samples between the current student and the frozen old-task teacher. For the ℓ-th encoder layer, we first ℓ_2-normalize the pooled replay responses and compute their sample-wise relational matrices:

$$\mathbf{R}_{s,r}^{(\ell)} = \bar{\mathbf{H}}_{s,r}^{(\ell)} \bar{\mathbf{H}}_{s,r}^{(\ell)\top}, \qquad \mathbf{R}_{t,r}^{(\ell)} = \bar{\mathbf{H}}_{t,r}^{(\ell)} \bar{\mathbf{H}}_{t,r}^{(\ell)\top}, \tag{15}$$

where $\bar{\mathbf{H}}$ denotes the ℓ_2-normalized response representation. RGD matches only the off-diagonal entries of the two relational matrices:

$$\mathcal{L}_{\mathrm{RGD}} = \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{\mathbf{B}(\mathbf{B}-1)} \sum_{i=1}^{\mathbf{B}} \sum_{\substack{j=1 \\ j\neq i}}^{\mathbf{B}} \left(R_{s,r,ij}^{(\ell)} - R_{t,r,ij}^{(\ell)} \right)^2, \tag{16}$$

where L denotes the number of encoder layers, $\mathbf{B}$ denotes the batch size, and $R_{s,r,ij}^{(\ell)}$ and $R_{t,r,ij}^{(\ell)}$ denote the (i,j)-th entries of the student and teacher relational matrices at layer ℓ, respectively. The diagonal entries are excluded because they only measure self-similarities, whereas the off-diagonal entries characterize the pairwise response geometry among replayed samples. In this way, the proposed Response-Geometry Distillation preserves the old-task response geometry encoded by the frozen teacher, i.e., the past model immediately after adaptation to the replayed task, and retains relational knowledge to mitigate catastrophic forgetting during new-task adaptation.

Spectral-Affinity Alignment: RGD focuses on old-task stability, whereas SAA encourages transferable response alignment between current and previous tasks. For each layer, we compute the response statistics of current-task student responses $\mathbf{H}_{s,c}^{(\ell)}$ and replay-teacher responses $\mathbf{H}_{t,r}^{(\ell)}$, denoted as $(\boldsymbol{\mu}_c^{(\ell)}, \boldsymbol{\sigma}_c^{(\ell)}, \boldsymbol{\Sigma}_c^{(\ell)})$ and $(\boldsymbol{\mu}_r^{(\ell)}, \boldsymbol{\sigma}_r^{(\ell)}, \boldsymbol{\Sigma}_r^{(\ell)})$, respectively. The statistical alignment loss is defined as:

$$\mathcal{L}_{\mathrm{align}} = \frac{1}{L} \sum_{\ell=1}^{L} \Big(\|\boldsymbol{\mu}_c^{(\ell)} - \boldsymbol{\mu}_r^{(\ell)}\|_2^2 + \|\boldsymbol{\sigma}_c^{(\ell)} - \boldsymbol{\sigma}_r^{(\ell)}\|_2^2 + \rho \|\boldsymbol{\Sigma}_c^{(\ell)} - \boldsymbol{\Sigma}_r^{(\ell)}\|_F^2 \Big), \tag{17}$$

where ρ balances covariance alignment. To avoid indiscriminate alignment between unrelated EEG paradigms, we weight this loss by cross-task spectral affinity. Specifically, let $\mathbf{g}_c = [g_{c,b}]_{b\in\mathcal{B}}$ and $\mathbf{g}_r = [g_{r,b}]_{b\in\mathcal{B}}$ denote the NSN band-gate vectors of the current and replayed tasks, respectively. We estimate the signed spectral affinity α_s from positive and negative gate vectors as follows:

$$s_{\mathrm{pos}} = \cos\big(\mathrm{ReLU}(\mathbf{g}_c), \mathrm{ReLU}(\mathbf{g}_r)\big), \qquad s_{\mathrm{neg}} = \cos\big(\mathrm{ReLU}(-\mathbf{g}_c), \mathrm{ReLU}(-\mathbf{g}_r)\big), \tag{18}$$

The spectral-affinity response alignment loss is then

$$\mathcal{L}_{\mathrm{SAA}} = \alpha_s \cdot \mathcal{L}_{\mathrm{align}}, \qquad \alpha_s = \mathrm{ReLU}\left(\frac{s_{\mathrm{pos}} + s_{\mathrm{neg}}}{2} \right). \tag{19}$$

Thus, response statistics are aligned only when the current and replayed tasks exhibit similar task-conditioned spectral signatures, reducing the risk of negative transfer across heterogeneous BCI paradigms. The overall RAD objective is

$$\mathcal{L}_{\mathrm{RAD}} = \lambda_r \mathcal{L}_{\mathrm{RGD}} + (1-\lambda_r)\mathcal{L}_{\mathrm{SAA}} \tag{20}$$

where λ_r controls the trade-off between old-task response-geometry retention and spectral-affinity-guided cross-task response alignment.

2.6 Overall Loss Function

For each task $\mathcal{T}_i$, EvoBrain optimizes the task-specific head with the classification loss $\mathcal{L}_{\mathrm{CE}}$ and, for $i > 1$, incorporates $\mathcal{L}_{\mathrm{RAD}}$ to regularize continual adaptation. The overall objective is

$$\mathcal{L} = \begin{cases} \mathcal{L}_{\mathrm{CE}}, & i = 1, \\ \lambda_{\mathrm{ce}}\mathcal{L}_{\mathrm{CE}} + (1 - \lambda_{\mathrm{ce}})\mathcal{L}_{\mathrm{RAD}}, & i > 1. \end{cases} \tag{21}$$

where λ_{ce} balances new-task plasticity and old-task stability. Together, NSN and RAD enable stable continual adaptation by mitigating cross-task shifts and preserving transferable response structures. Further details are provided in Appendix B.

3 Main Results

3.1 Experimental Setup

Datasets and Baselines: We evaluate EvoBrain on six heterogeneous EEG datasets covering emotion recognition, motor imagery, sleep staging, imagined speech, and mental-disorder diagnosis: FACED Chen et al. (2023), PhysioNet-MI Schalk et al. (2004), BCIC-IV-2a Brunner et al. (2008), ISRUC Khalighi et al. (2016), BCIC2020-3 Jeong et al. (2022), and Mumtaz2016 Mumtaz et al. (2017). We integrate EvoBrain with four representative pretrained EEG foundation backbones—CBraMod Wang et al. (2024b), EEGMamba Wang et al. (2025), CSBrain Zhou et al. (2025b), and DeeperBrain Wang et al. (2026)—using their official implementations and pretrained weights. Detailed dataset statistics, preprocessing protocols, and backbone descriptions are provided in Appendix C.

Evaluation Metrics: Consistent with prior studies, we use B-ACC, WF1, and κ for multi-class tasks, and B-ACC, AUROC, and AUC-PR for binary tasks. Plasticity is measured by performance on $\mathcal{T}_i$ after adaptation, while memory stability is measured by the Relative Forgetting Rate:

$$\mathrm{RFR}_i = \frac{A_{i,i} - A_{i,\mathcal{N}}}{A_{i,i}}, \tag{22}$$

where $A_{i,i}$ and $A_{i,\mathcal{N}}$ denote the B-ACC on $\mathcal{T}_i$ immediately after learning and after all $\mathcal{N}$ tasks, respectively. Lower values indicate better retention, with negative values indicating positive transfer.

3.2 Overview Performance

We evaluate EvoBrain across four EEG foundation models in a heterogeneous cross-task continual learning setting, with three task arrival orders reported in Tab. 1. We denote the performance of the i-th task immediately after adaptation as $\mathcal{M}_i$, and its performance after completing the full task sequence as $\mathcal{M}_{\mathcal{N}}$. We analyze EvoBrain from three perspectives. **Cross-Task Learning Ability**: EvoBrain effectively adapts to heterogeneous EEG tasks while preserving previously acquired knowledge. NSN mitigates task-specific temporal and spectral shifts, while RAD maintains old-task response structures and facilitates transfer across spectrally compatible tasks. As a result, EvoBrain consistently achieves low RFR, with occasional negative forgetting indicating beneficial cross-task transfer. **Cross-Architecture Compatibility**: EvoBrain remains effective across four EEG foundation models despite their distinct architectures, pretraining corpora, and representation mechanisms. Across all backbone-order combinations, the average RFR remains below 4.0%, demonstrating that EvoBrain can serve as an architecture-agnostic continual adaptation framework without structural modification. **Cross-Order Robustness**: Across three task permutations, EvoBrain maintains stable final performance and consistently low forgetting for all backbones, indicating limited sensitivity to task order. This robustness is particularly important for practical BCI scenarios, where new tasks may arrive in unpredictable sequences.

3.3 Ablation Study

Tab. 2 reports the ablation results across four EEG foundation models. The full EvoBrain achieves the strongest or tied-strongest retained Average B-ACC with consistently low RFR, showing that its improvement comes from both task adaptation and knowledge retention. Removing NSN clearly degrades performance on most backbones, indicating the necessity of task-aware normalization for heterogeneous EEG tasks. Its two components are complementary: PTN stabilizes temporal statistics, while SRM recalibrates task-specific spectral responses. Removing RAD further reduces retained performance and increases forgetting, demonstrating that response-level regularization is essential

Table 1: **Overview performance of cross-task continual learning under different task input orders across EEG foundation models.** $\mathcal{M}_i$ denotes the performance immediately after adapting to the i-th task, whereas $\mathcal{M}_\mathcal{N}$ represents the performance after adapting to the final task.

Order 1		ISRUC, 5-classes [Sleep Staging]			PhysioNet-MI, 4-classes [Motor Imagery]			FACED, 9-classes [Emotion Recognition]			BCIC-2020-3, 5-classes [Imagined Speech]			Mumtaz, 2-classes [Mental Dis. Diag.]			BCIC-IV-2a, 4-classes [Motor Imagery]			Average B-ACC		
		$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR
CBraMod	B-ACC	0.751	0.735	2.0%	0.573	0.547	4.5%	0.499	0.480	3.5%	0.488	0.476	2.5%	0.902	0.883	2.1%	0.520	0.520	-			
	κ/AUC-PR	0.697	0.672	3.7%	0.430	0.396	7.9%	0.431	0.411	4.5%	0.360	0.345	4.1%	0.960	0.922	3.9%	0.360	0.360	-	0.622	0.607	2.4%
	WF1/AUROC	0.763	0.733	3.9%	0.569	0.546	4.0%	0.492	0.477	3.0%	0.488	0.476	2.4%	0.969	0.951	1.8%	0.514	0.514	-			
EEGMamba	B-ACC	0.754	0.748	0.8%	0.529	0.527	0.4%	0.439	0.450	-2.6%	0.493	0.484	1.9%	0.864	0.862	0.2%	0.490	0.490	-			
	κ/AUC-PR	0.698	0.685	1.8%	0.372	0.369	0.8%	0.368	0.378	-2.7%	0.367	0.355	3.2%	0.925	0.921	0.4%	0.319	0.319	-	0.595	0.594	0.1%
	WF1/AUROC	0.765	0.757	1.0%	0.524	0.515	1.6%	0.439	0.445	-1.4%	0.493	0.483	2.1%	0.921	0.915	0.6%	0.473	0.473	-			
CSBrain	B-ACC	0.771	0.761	1.2%	0.571	0.541	5.2%	0.499	0.486	2.6%	0.529	0.500	5.5%	0.911	0.907	0.4%	0.516	0.516	-			
	κ/AUC-PR	0.709	0.705	0.6%	0.428	0.389	9.3%	0.436	0.418	4.2%	0.412	0.375	8.9%	0.980	0.979	0.1%	0.355	0.355	-	0.633	0.619	2.3%
	WF1/AUROC	0.775	0.773	0.2%	0.573	0.540	5.9%	0.508	0.482	5.1%	0.529	0.498	5.9%	0.978	0.977	0.1%	0.506	0.506	-			
DeeperBrain	B-ACC	0.762	0.715	6.2%	0.585	0.579	1.0%	0.537	0.530	1.2%	0.413	0.405	1.9%	0.913	0.918	-0.5%	0.577	0.577	-			
	κ/AUC-PR	0.731	0.666	9.0%	0.446	0.438	1.8%	0.472	0.462	2.1%	0.267	0.257	3.8%	0.978	0.980	-0.2%	0.436	0.436	-	0.631	0.621	1.6%
	WF1/AUROC	0.784	0.722	7.9%	0.585	0.575	1.7%	0.530	0.512	3.3%	0.414	0.406	1.9%	0.976	0.978	-0.2%	0.568	0.568	-			

Order 2		ISRUC, 5-classes [Sleep Staging]			BCIC-IV-2a, 4-classes [Motor Imagery]			Mumtaz, 2-classes [Mental Dis. Diag.]			PhysioNet-MI, 4-classes [Motor Imagery]			FACED, 9-classes [Emotion Recognition]			BCIC-2020-3, 5-classes [Imagined Speech]			Average B-ACC		
		$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR
CBraMod	B-ACC	0.751	0.741	1.2%	0.475	0.423	10.9%	0.871	0.860	1.2%	0.594	0.586	1.2%	0.492	0.504	-2.5%	0.485	0.485	-			
	κ/AUC-PR	0.697	0.685	1.8%	0.300	0.230	23.2%	0.966	0.889	7.9%	0.458	0.448	2.1%	0.427	0.441	-3.4%	0.357	0.357	-	0.611	0.600	1.8%
	WF1/AUROC	0.763	0.749	1.9%	0.444	0.362	18.4%	0.964	0.920	4.5%	0.596	0.587	1.6%	0.490	0.507	-3.5%	0.485	0.485	-			
EEGMamba	B-ACC	0.742	0.745	-0.4%	0.475	0.453	4.6%	0.865	0.813	6.1%	0.543	0.543	0.0%	0.439	0.444	-1.1%	0.469	0.469	-			
	κ/AUC-PR	0.681	0.689	-1.1%	0.300	0.271	9.7%	0.934	0.916	2.0%	0.390	0.390	0.0%	0.370	0.374	-1.3%	0.337	0.337	-	0.589	0.578	1.9%
	WF1/AUROC	0.747	0.754	-1.0%	0.476	0.446	6.2%	0.918	0.903	1.6%	0.546	0.540	1.1%	0.441	0.444	-0.7%	0.468	0.468	-			
CSBrain	B-ACC	0.746	0.719	3.7%	0.482	0.456	5.4%	0.903	0.902	0.2%	0.600	0.546	8.9%	0.526	0.507	3.5%	0.483	0.483	-			
	κ/AUC-PR	0.685	0.644	5.9%	0.309	0.274	11.2%	0.982	0.979	3.6%	0.466	0.395	15.2%	0.466	0.444	4.8%	0.353	0.353	-	0.623	0.602	3.4%
	WF1/AUROC	0.747	0.719	3.8%	0.466	0.448	3.9%	0.982	0.979	0.3%	0.604	0.552	8.6%	0.539	0.516	4.3%	0.483	0.483	-			
DeeperBrain	B-ACC	0.756	0.707	6.4%	0.580	0.563	2.8%	0.920	0.906	1.6%	0.585	0.573	2.0%	0.529	0.534	-1.0%	0.408	0.408	-			
	κ/AUC-PR	0.703	0.638	9.2%	0.440	0.418	5.0%	0.977	0.976	0.1%	0.446	0.431	3.5%	0.469	0.473	-0.1%	0.260	0.260	-	0.630	0.615	2.3%
	WF1/AUROC	0.770	0.719	6.7%	0.567	0.545	4.0%	0.976	0.977	-0.1%	0.587	0.573	2.3%	0.532	0.533	-0.1%	0.406	0.406	-			

Order 3		ISRUC, 5-classes [Sleep Staging]			PhysioNet-MI, 4-classes [Motor Imagery]			BCIC-IV-2a, 4-classes [Motor Imagery]			BCIC-2020-3, 5-classes [Imagined Speech]			FACED, 9-classes [Emotion Recognition]			Mumtaz, 2-classes [Mental Dis. Diag.]			Average B-ACC		
		$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR
CBraMod	B-ACC	0.755	0.705	6.6%	0.582	0.541	7.0%	0.470	0.512	-9.1%	0.509	0.485	4.7%	0.518	0.463	10.6%	0.904	0.904	-			
	κ/AUC-PR	0.683	0.625	8.4%	0.443	0.388	12.3%	0.293	0.350	-19.4%	0.387	0.357	7.8%	0.454	0.392	13.7%	0.969	0.969	-	0.623	0.602	3.4%
	WF1/AUROC	0.751	0.697	7.2%	0.581	0.536	7.6%	0.455	0.496	-9.2%	0.509	0.484	4.9%	0.522	0.470	9.9%	0.965	0.965	-			
EEGMamba	B-ACC	0.746	0.742	0.6%	0.541	0.534	1.2%	0.473	0.473	0.0%	0.457	0.441	3.5%	0.433	0.433	0.0%	0.859	0.859	-			
	κ/AUC-PR	0.679	0.674	0.7%	0.388	0.379	2.3%	0.297	0.297	0.0%	0.322	0.302	6.2%	0.363	0.364	-0.1%	0.936	0.936	-	0.585	0.581	0.8%
	WF1/AUROC	0.747	0.745	0.1%	0.542	0.526	2.9%	0.463	0.462	0.1%	0.457	0.439	4.0%	0.434	0.436	-0.4%	0.919	0.919	-			
CSBrain	B-ACC	0.752	0.750	0.2%	0.576	0.539	6.4%	0.487	0.477	2.0%	0.508	0.483	4.9%	0.473	0.448	5.3%	0.900	0.900	-			
	κ/AUC-PR	0.680	0.679	0.2%	0.434	0.385	11.4%	0.316	0.303	4.0%	0.385	0.353	8.2%	0.404	0.376	6.9%	0.972	0.972	-	0.616	0.599	2.7%
	WF1/AUROC	0.748	0.751	-0.4%	0.578	0.537	7.1%	0.485	0.475	2.1%	0.506	0.482	4.8%	0.472	0.448	5.0%	0.966	0.966	-			
DeeperBrain	B-ACC	0.770	0.704	8.5%	0.582	0.562	3.3%	0.568	0.543	4.4%	0.404	0.373	7.6%	0.473	0.470	0.5%	0.905	0.905	-			
	κ/AUC-PR	0.702	0.608	13.3%	0.442	0.417	5.8%	0.424	0.390	7.9%	0.255	0.217	15.0%	0.401	0.398	0.8%	0.978	0.978	-	0.617	0.593	3.9%
	WF1/AUROC	0.768	0.680	11.4%	0.583	0.563	3.5%	0.559	0.527	5.6%	0.404	0.373	7.8%	0.474	0.473	0.2%	0.975	0.975	-			

Table 2: **Ablation study of EvoBrain across diverse EEG foundation models.**

		ISRUC, 5-classes			PhysioNet-MI, 4-classes			FACED, 9-classes			BCIC-2020-3, 5-classes			Mumtaz, 2-classes			BCIC-IV-2a, 4-classes			Average B-ACC		
		B-ACC	κ	WF1	B-ACC	κ	WF1	B-ACC	κ	WF1	B-ACC	κ	WF1	B-ACC	AUC-PR	AUROC	B-ACC	κ	WF1	$\mathcal{M}_i$	$\mathcal{M}_\mathcal{N}$	RFR
CBraMod	EvoBrain	0.735	0.672	0.733	0.547	0.396	0.546	0.480	0.411	0.477	0.476	0.345	0.476	0.883	0.922	0.951	0.520	0.360	0.514	0.622	**0.607**	**2.4%**
	w/o NSN	0.743	0.697	0.760	0.516	0.354	0.519	0.393	0.309	0.379	0.359	0.198	0.350	0.877	0.957	0.961	0.455	0.273	0.429	0.590	0.557	5.6%
	▷ w/o PTN	0.754	0.700	0.761	0.533	0.377	0.533	0.408	0.329	0.399	0.357	0.197	0.355	0.901	0.951	0.961	0.457	0.277	0.429	0.583	0.569	2.5%
	▷ w/o SRM	0.704	0.645	0.716	0.515	0.354	0.516	0.440	0.363	0.433	0.444	0.305	0.443	0.809	0.800	0.879	0.522	0.362	0.512	0.613	0.572	6.7%
	w/o RAD	0.731	0.679	0.742	0.557	0.409	0.559	0.463	0.395	0.470	0.437	0.297	0.434	0.767	0.816	0.879	0.509	0.345	0.500	0.605	0.577	4.6%
	▷ w/o RGD	0.723	0.670	0.736	0.551	0.402	0.551	0.455	0.385	0.456	0.451	0.313	0.449	0.653	0.770	0.832	0.504	0.339	0.499	0.608	0.556	8.6%
	▷ w/o SAA	0.732	0.667	0.731	0.545	0.394	0.542	0.456	0.382	0.445	0.464	0.330	0.462	0.843	0.911	0.944	0.515	0.353	0.501	0.618	0.592	4.2%
	w/o NSN+RAD	0.732	0.695	0.760	0.464	0.285	0.465	0.353	0.271	0.344	0.337	0.172	0.330	0.879	0.933	0.944	0.458	0.278	0.431	0.589	0.537	8.8%
EEGMamba	EvoBrain	0.748	0.685	0.757	0.527	0.369	0.515	0.450	0.378	0.445	0.484	0.355	0.483	0.862	0.921	0.915	0.490	0.319	0.473	0.595	**0.594**	**0.1%**
	w/o NSN	0.714	0.633	0.716	0.538	0.384	0.524	0.450	0.376	0.446	0.448	0.310	0.447	0.857	0.919	0.901	0.475	0.300	0.470	0.587	0.580	1.2%
	▷ w/o PTN	0.735	0.667	0.734	0.527	0.369	0.518	0.442	0.370	0.441	0.487	0.358	0.486	0.858	0.921	0.911	0.483	0.311	0.464	0.590	0.588	0.2%
	▷ w/o SRM	0.755	0.689	0.757	0.526	0.368	0.517	0.447	0.374	0.440	0.493	0.367	0.493	0.859	0.929	0.917	0.484	0.311	0.465	0.596	0.594	0.3%
	w/o RAD	0.686	0.630	0.706	0.521	0.362	0.523	0.443	0.371	0.444	0.451	0.313	0.449	0.861	0.919	0.910	0.491	0.322	0.485	0.580	0.575	0.8%
	▷ w/o RGD	0.702	0.649	0.722	0.528	0.371	0.520	0.413	0.339	0.413	0.489	0.362	0.489	0.838	0.931	0.919	0.490	0.319	0.472	0.589	0.577	2.1%
	▷ w/o SAA	0.748	0.702	0.770	0.535	0.524	0.380	0.451	0.381	0.449	0.464	0.330	0.463	0.853	0.925	0.914	0.483	0.310	0.474	0.591	0.589	0.3%
	w/o NSN+RAD	0.695	0.632	0.712	0.527	0.369	0.528	0.455	0.384	0.456	0.489	0.362	0.489	0.844	0.925	0.905	0.484	0.311	0.471	0.590	0.582	1.3%
CSBrain	EvoBrain	0.761	0.705	0.773	0.541	0.389	0.540	0.486	0.418	0.482	0.500	0.375	0.498	0.911	0.979	0.977	0.516	0.355	0.506	0.633	**0.619**	2.3%
	w/o NSN	0.761	0.717	0.785	0.557	0.409	0.557	0.436	0.364	0.438	0.399	0.248	0.399	0.905	0.977	0.977	0.508	0.344	0.493	0.606	0.593	2.2%
	▷ w/o PTN	0.729	0.666	0.755	0.535	0.380	0.536	0.457	0.385	0.454	0.445	0.307	0.445	0.905	0.968	0.963	0.506	0.341	0.497	0.613	0.596	2.7%
	▷ w/o SRM	0.745	0.679	0.756	0.549	0.398	0.546	0.470	0.398	0.464	0.479	0.348	0.477	0.898	0.967	0.963	0.498	0.331	0.492	0.621	0.607	2.3%
	w/o RAD	0.747	0.684	0.757	0.512	0.349	0.514	0.460	0.386	0.460	0.463	0.328	0.463	0.891	0.971	0.964	0.509	0.346	0.502	0.631	0.597	5.4%
	▷ w/o RGD	0.722	0.660	0.746	0.522	0.362	0.518	0.476	0.405	0.467	0.473	0.342	0.468	0.908	0.968	0.960	0.467	0.289	0.464	0.622	0.595	4.4%
	▷ w/o SAA	0.749	0.686	0.763	0.531	0.374	0.529	0.482	0.413	0.480	0.485	0.357	0.485	0.891	0.963	0.959	0.502	0.336	0.496	0.619	0.607	**2.0%**
	w/o NSN+RAD	0.708	0.625	0.718	0.553	0.404	0.551	0.350	0.268	0.348	0.304	0.130	0.298	0.886	0.969	0.969	0.487	0.316	0.471	0.585	0.548	6.4%
DeeperBrain	EvoBrain	0.715	0.666	0.722	0.579	0.438	0.575	0.530	0.462	0.513	0.405	0.257	0.406	0.918	0.980	0.978	0.577	0.436	0.568	0.631	**0.621**	**1.6%**
	w/o NSN	0.602	0.507	0.586	0.543	0.391	0.545	0.521	0.453	0.513	0.453	0.317	0.452	0.872	0.977	0.977	0.564	0.419	0.558	0.627	0.593	5.6%
	▷ w/o PTN	0.637	0.544	0.612	0.544	0.392	0.544	0.513	0.441	0.492	0.419	0.273	0.416	0.902	0.976	0.975	0.563	0.418	0.551	0.622	0.596	4.1%
	▷ w/o SRM	0.691	0.637	0.705	0.557	0.409	0.556	0.514	0.447	0.508	0.376	0.220	0.369	0.908	0.978	0.977	0.563	0.417	0.555	0.617	0.601	2.6%
	w/o RAD	0.561	0.431	0.498	0.531	0.374	0.524	0.536	0.472	0.529	0.436	0.295	0.435	0.912	0.982	0.982	0.574	0.432	0.569	0.632	0.592	6.4%
	▷ w/o RGD	0.629	0.534	0.609	0.563	0.417	0.564	0.514	0.445	0.499	0.444	0.305	0.442	0.894	0.975	0.975	0.582	0.442	0.573	0.629	0.604	3.9%
	▷ w/o SAA	0.629	0.546	0.623	0.558	0.410	0.556	0.528	0.462	0.527	0.429	0.287	0.424	0.901	0.977	0.976	0.559	0.412	0.553	0.633	0.600	5.1%
	w/o NSN+RAD	0.535	0.399	0.468	0.532	0.376	0.532	0.523	0.459	0.518	0.419	0.273	0.417	0.897	0.979	0.977	0.556	0.409	0.548	0.630	0.577	8.5%

for preserving historical task competence. Specifically, RGD maintains old-task response geometry, while SAA enables selective transfer between spectrally compatible tasks and avoids indiscriminate alignment across unrelated paradigms. Despite EEGMamba's lower ablation sensitivity, EvoBrain maintains the best overall stability-plasticity trade-off.

3.4 Comparison with Other Methods

To rigorously evaluate EvoBrain, we compare EvoBrain with representative continual learning baselines under the Order 1 sequence, including MMD Gretton et al. (2006), EWC Kirkpatrick et al. (2017), LwF Li and Hoiem (2017), DER++ Buzzega et al. (2020), AGEM Chaudhry et al. (2018),

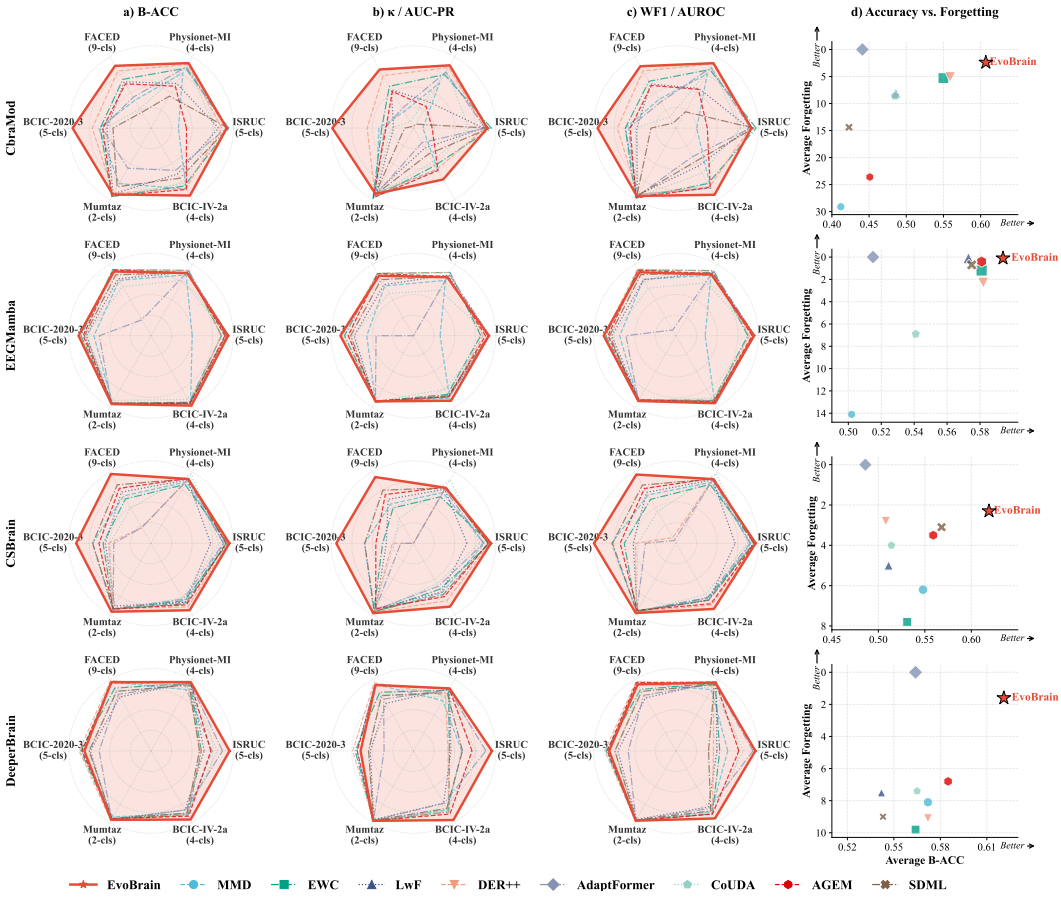

Figure 3: **Comparison of EvoBrain with competing methods across four EEG foundation models.** (a)–(c) report retained performance across six heterogeneous tasks, while (d) summarizes the stability–plasticity trade-off in terms of average B-ACC and forgetting.

SDML Wang et al. (2022), CoUDA Chen et al. (2025b), and AdaptFormer Chen et al. (2022), covering major paradigms such as regularization, replay, domain adaptation, and parameter-efficient adaptation. As shown in Fig. 3(a)–(c), EvoBrain achieves a larger and more balanced performance envelope across six heterogeneous EEG tasks under B-ACC, κ/AUC-PR, and WF1/AUROC, demonstrating stronger retained performance after continual adaptation. In contrast, competing methods show clear contractions on challenging tasks such as FACED and BCIC-2020-3, indicating that conventional regularization, replay, or adapter-based strategies struggle with large cross-task discrepancies. Fig. 3(d) further confirms that EvoBrain provides a better stability-plasticity trade-off, maintaining higher Average B-ACC with lower Average Forgetting. While some baselines are more competitive on EEGMamba and AdaptFormer performs reasonably on simpler tasks such as PhysioNet-MI and Mumtaz, they remain limited on complex tasks requiring high-order cognitive representations. These results suggest that EvoBrain's task-adaptive recalibration and response-level knowledge preservation are more effective for heterogeneous cross-task continual EEG learning.

4 Conclusion

In this paper, we presented EvoBrain, an architecture-agnostic continual learning framework that enables a single pretrained EEG foundation model to incrementally acquire heterogeneous BCI tasks while retaining prior decoding capabilities. By jointly addressing cross-task distribution shifts and catastrophic forgetting through NSN and RAD, EvoBrain achieves a favorable stability–plasticity trade-off across six EEG tasks and four foundation backbones. Our analyses further show that the evolving model learns transferable representations while preserving task-specific neuro-spectral characteristics. To our knowledge, this work provides the first systematic study of task-level continual learning for EEG foundation models, establishing continual adaptation as a practical post-training paradigm toward scalable one-for-all brain decoding.

5 AI USE STATEMENT

In this work, generative AI tools were used for language polishing and assistance in drafting code for figure generation. Generative AI was not used to formulate the core research ideas, research questions or scientific conclusions. All AI-assisted text was reviewed and revised by the authors, and all AI-assisted plotting code was manually inspected, tested, and verified for correctness. The authors take full responsibility for the final content of this work, including all text, analyses, claims, figures, and conclusions.

REFERENCES

Zhaohui Wu, Yongdi Zhou, Zhongzhi Shi, Changshui Zhang, Guanglin Li, Xiaoxiang Zheng, Nenggan Zheng, and Gang Pan. Cyborg intelligence: recent progress and future directions. *IEEE Intelligent Systems*, 31(6):44–50, 2016.

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.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

Jaeseung Jeong. Eeg dynamics in patients with alzheimer's disease. *Clinical neurophysiology*, 115 (7):1490–1505, 2004.

Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective computing*, 5(3):327–339, 2014.

Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.

Yangxuan Zhou, Sha Zhao, Jiquan Wang, Haiteng Jiang, Shijian Li, Benyan Luo, Tao Li, and Gang Pan. Personalized sleep staging leveraging source-free unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14529–14537, 2025a.

Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2022.

Hamdi Altaheri, Ghulam Muhammad, Mansour Alsulaiman, Syed Umar Amin, Ghadir Ali Altuwaijri, Wadood Abdul, Mohamed A Bencherif, and Mohammed Faisal. Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review. *Neural Computing and Applications*, 35(20):14681–14722, 2023.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36:78240–78260, 2023.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36:26304–26321, 2023.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024a.

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, 2024a.

Enze Shi, Kui Zhao, Qilong Yuan, Jiaqi Wang, Huawen Hu, Sigang Yu, and Shu Zhang. Fome: A foundation model for eeg using adaptive temporal-lateral attention scaling. *arXiv preprint arXiv:2409.12454*, 2024.

Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals. *arXiv preprint arXiv:2409.00101*, 2024b.

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. *arXiv preprint arXiv:2412.07236*, 2024b.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Shijian Li, and Gang Pan. Eegmamba: An eeg foundation model with mamba. *Neural Networks*, page 107816, 2025.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. Csbrain: A cross-scale spatiotemporal brain foundation model for eeg decoding. *arXiv preprint arXiv:2506.23075*, 2025b.

Yi Ding, Muyun Jiang, Weibang Jiang, Shuailei Zhang, Xinliang Zhou, Chenyu Liu, Shanglin Li, Yong Li, and Cuntai Guan. Brainpro: Towards large-scale brain state-aware eeg representation learning. *arXiv preprint arXiv:2509.22050*, 2025.

Zhisheng Chen, Yingwei Zhang, Qizhen Lan, Tianyu Liu, Huacan Wang, Yi Ding, Ziyu Jia, Ronghao Chen, Kun Wang, and Xinliang Zhou. Uni-ntfm: A unified foundation model for eeg signal representation learning. *arXiv preprint arXiv:2509.24222*, 2025a.

Jiquan Wang, Sha Zhao, Yangxuan Zhou, Yiming Kang, Shijian Li, and Gang Pan. Deeperbrain: A neuro-grounded eeg foundation model towards universal bci, 2026.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. Are emergent abilities of large language models a mirage? *Advances in neural information processing systems*, 36:55565–55581, 2023.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: Theory, method and application. *IEEE transactions on pattern analysis and machine intelligence*, 46(8):5362–5383, 2024c.

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.

Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. Bci competition 2008–graz data set a. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 16(1-6):34, 2008.

Sirvan Khalighi, Teresa Sousa, José Moutinho Santos, and Urbano Nunes. Isruc-sleep: A comprehensive public dataset for sleep researchers. *Computer Methods and Programs in Biomedicine*, 124:180–192, 2016. ISSN 0169-2607.

Ji-Hoon Jeong, Jeong-Hyun Cho, Young-Eun Lee, Seo-Hyun Lee, Gi-Hwan Shin, Young-Seok Kweon, José del R Millán, Klaus-Robert Müller, and Seong-Whan Lee. 2020 international brain–computer interface competition: A review. *Frontiers in human neuroscience*, 16:898300, 2022.

Wajid Mumtaz, Likun Xia, Syed Saad Azhar Ali, Mohd Azhar Mohd Yasin, Muhammad Hussain, and Aamir Saeed Malik. Electroencephalogram (eeg)-based computer-aided technique to diagnose major depressive disorder (mdd). *Biomedical Signal Processing and Control*, 31:108–115, 2017.

Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19, 2006.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33:15920–15930, 2020.

Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420*, 2018.

Zhenyi Wang, Li Shen, Tiehang Duan, Donglin Zhan, Le Fang, and Mingchen Gao. Learning to learn and remember super long multi-domain task sequence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7982–7992, 2022.

Bojian Chen, Xinmin Zhang, Changqing Shen, Qi Li, and Zhihuan Song. Couda: Continual unsupervised domain adaptation for industrial fault diagnosis under dynamic working conditions. *IEEE Transactions on Industrial Informatics*, 2025b.

Shoufa Chen, Chongjian Ge, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35:16664–16678, 2022.

Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R32, 2007.

Wei Wu, Zhe Chen, Xiaorong Gao, Yuanqing Li, Emery N Brown, and Shangkai Gao. Probabilistic common spatial patterns for multichannel eeg analysis. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):639–653, 2014.

Suresh Balakrishnama and Aravind Ganapathiraju. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18(1998):1–8, 1998.

Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE transactions on neural networks and learning systems*, 29(11):5619–5629, 2018.

Ping Wang, Aimin Jiang, Xiaofeng Liu, Jing Shang, and Li Zhang. Lstm-based eeg classification in motor imagery tasks. *IEEE transactions on neural systems and rehabilitation engineering*, 26 (11):2086–2095, 2018.

Ahmed Abdelhameed and Magdy Bayoumi. A deep learning approach for automatic seizure detection in children with epilepsy. *Frontiers in Computational Neuroscience*, 15:650050, 2021.

Yonghao Song, Xueyu Jia, Lie Yang, and Longhan Xie. Transformer-based spatial-temporal feature learning for eeg decoding. *arXiv preprint arXiv:2106.11170*, 2021.

Yang Du, Yongling Xu, Xiaoan Wang, Li Liu, and Pengcheng Ma. Eeg temporal–spatial transformer for person identification. *Scientific Reports*, 12(1):14378, 2022.

Huy Phan, Kaare Mikkelsen, Oliver Y Chén, Philipp Koch, Alfred Mertins, and Maarten De Vos. Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering*, 69(8):2456–2467, 2022.

Jiayao Sun, Jin Xie, and Huihui Zhou. Eeg classification with transformer-based models. In *2021 ieee 3rd global conference on life sciences and technologies (lifetech)*, pages 92–93. IEEE, 2021.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

Jin Xie, Jie Zhang, Jiayao Sun, Zheng Ma, Liuni Qin, Guanglin Li, Huihui Zhou, and Yang Zhan. A transformer-based approach combining deep learning network and spatial-temporal information for raw eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2126–2136, 2022.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021.

Junhong Lai, Jiyu Wei, Lin Yao, and Yueming Wang. A simple review of eeg foundation models: Datasets, advancements and future perspectives. *arXiv preprint arXiv:2504.20069*, 2025.

Gayal Kuruppu, Neeraj Wagh, and Yogatheesan Varatharajah. Eeg foundation models: A critical review of current progress and future directions. *arXiv preprint arXiv:2507.11783*, 2025.

Wei Xiong, Jiangtong Li, Jie Li, and Kun Zhu. Eeg-fm-bench: A comprehensive benchmark for the systematic evaluation of eeg foundation models. *arXiv preprint arXiv:2508.17742*, 2025.

Daoze Zhang, Zhizhang Yuan, Junru Chen, Kerui Chen, and Yang Yang. Brant-x: A unified physiological signal alignment framework. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4155–4166, 2024.

Yuqi Chen, Kan Ren, Kaitao Song, Yansen Wang, Yifan Wang, Dongsheng Li, and Lili Qiu. Eegformer: Towards transferable and interpretable large-scale eeg foundation model. *arXiv preprint arXiv:2401.10278*, 2024.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.

Jingying Ma, Feng Wu, Qika Lin, Yucheng Xing, Chenyu Liu, Ziyu Jia, and Mengling Feng. Codebrain: Bridging decoupled tokenizer and multi-scale architecture for eeg foundation model. *arXiv preprint arXiv:2506.09110*, 2025.

Ang Li, Zikai Wang, Liuyin Yang, Zhenyu Wang, Tianheng Xu, Honglin Hu, and Marc M Van Hulle. Comet: A contrastive-masked brain foundation model for universal eeg representation. *arXiv preprint arXiv:2509.00314*, 2025a.

Sicheng Dai, Kai Chen, Hongwang Xiao, Shan Yu, and Qiwei Ye. Mteeg: A multi-task learning framework for enhanced electroencephalography analysis using low-rank adaptation.

Weiheng Lu, Chunfeng Song, Jiamin Wu, Pengyu Zhu, Yuchen Zhou, Weijian Mai, Qihao Zheng, and Wanli Ouyang. Unimind: Unleashing the power of llms for unified multi-task brain decoding. *arXiv preprint arXiv:2506.18962*, 2025.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pages 3987–3995. PMLR, 2017.

Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016.

Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*, pages 233–248, 2018.

Tiehang Duan, Zhenyi Wang, Gianfranco Doretto, Fang Li, Cui Tao, and Donald Adjeroh. Replay with stochastic neural transformation for online continual eeg classification. In *2023 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 1874–1879. IEEE, 2023.

Amirhossein Shahbazinia, Flavio Ponzina, Jose A Miranda, Jonathan Dan, Giovanni Ansaloni, and David Atienza. Resource-efficient continual learning for personalized online seizure detection. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–7. IEEE, 2024.

Jiarui Jin, Zongnan Chen, Honghua Cai, and Jiahui Pan. Affective eeg-based person identification with continual learning. *IEEE Transactions on Instrumentation and Measurement*, 73:1–16, 2024.

Dan Li, Hye-Bin Shin, and Kang Yin. Personalized continual eeg decoding: Retaining and transferring knowledge. In *2025 13th International Conference on Brain-Computer Interface (BCI)*, pages 1–4. IEEE, 2025b.

Yangxuan Zhou, Sha Zhao, Jiquan Wang, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Brainuicl: An unsupervised individual continual learning framework for eeg applications. In *The Thirteenth International Conference on Learning Representations*, 2025c.

Yangxuan Zhou, Sha Zhao, Jiquan Wang, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Spiced: A synaptic homeostasis-inspired framework for unsupervised continual eeg decoding. *Advances in Neural Information Processing Systems*, 38:153014–153040, 2026.

Richard B Berry, Rita Brooks, Charlene Gamaldo, Susan M Harding, Robin M Lloyd, Stuart F Quan, Matthew T Troester, and Bradley V Vaughn. Aasm scoring manual updates for 2017 (version 2.4), 2017.

A Related Work

A.1 EEG Decoding

EEG employs scalp sensors to record the brain's spontaneous electrical activity. Early EEG decoding approaches predominantly relied on handcrafted features combined with traditional machine learning methods Lotte et al. (2007); Bashashati et al. (2007). These features included time-domain statistics, frequency-band power and spatial filters such as Common Spatial Patterns (CSP) Wu et al. (2014). The extracted features were then fed into shallow classifiers, typically Linear Discriminant Analysis (LDA) Balakrishnama and Ganapathiraju (1998) or Support Vector Machines (SVM) Hearst et al. (1998). These methods achieved moderate performance but were sensitive to inter-subject variability and required extensive domain expertise for feature engineering. The advent of deep learning LeCun et al. (2015) has shifted the paradigm toward end-to-end trainable architectures that automatically learn discriminative representations from raw or minimally preprocessed EEG signals. For example, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks are widely used to capture the spatio-temporal dynamics inherent in EEG signals and have demonstrated superior accuracy across diverse experimental paradigms Schirrmeister et al. (2017); Sakhavi et al. (2018); Wang et al. (2018); Abdelhameed and Bayoumi (2021). The Transformer architecture has also been adopted to learn spatio-temporal features for various EEG decoding applications Song et al. (2021); Du et al. (2022); Phan et al. (2022). To leverage the complementary advantages of CNNs and Transformers, several studies have proposed hybrid CNN–Transformer models for EEG classification tasks Sun et al. (2021); Song et al. (2022); Xie et al. (2022). While these methods perform well on specific tasks or paradigm settings, their inherent task-specific architectural design limits their ability to generalize across diverse EEG tasks, thereby restricting their applications in practice.

A.2 EEG Foundation Models

Inspired by the remarkable success of foundation models in natural language processing Achiam et al. (2023); Devlin et al. (2019), computer vision He et al. (2022); Kirillov et al. (2023), and multimodal perception Radford et al. (2021), a growing body of recent work has begun to explore their application to the decoding of brain signals. EEG foundation models typically follow a pretraining–post-training paradigm Lai et al. (2025); Kuruppu et al. (2025); Xiong et al. (2025): they are first pretrained on large-scale unlabeled EEG data and subsequently adapted to diverse downstream tasks through post-training strategies. Current research has primarily focused on improving the learning of general-purpose EEG representations during pretraining, while largely neglecting the design of effective post-training adaptation mechanisms. Existing adapting approaches can be broadly categorized into two types: (i) single-task fine-tuning Zhang et al. (2023); Jiang et al. (2024a); Wang et al. (2024a); Shi et al. (2024); Wang et al. (2024b; 2025); Zhou et al. (2025b); Ding et al. (2025); Zhang et al. (2024); Chen et al. (2024); Cui et al. (2024); Ma et al. (2025); Li et al. (2025a), which fine-tunes the entire model separately for each task, leading to linear growth in computational and storage costs with respect to the number of tasks; and (ii) multi-task fine-tuning Jiang et al. (2024b); Chen et al. (2025a); Dai et al.; Lu et al. (2025), which shares backbone parameters across tasks during joint optimization. Although the latter mitigates resource overhead, it remains a static adaptation scheme that requires re-fine-tuning whenever a new task is introduced and employs only naive multi-task learning without explicitly modeling cross-task knowledge transfer. Consequently, there exists a compelling need for a dynamic and task-aware adaptation mechanism to overcome the inherent limitations of current post-training strategies for EEG foundation models.

A.3 Continual EEG Decoding

Studies in continual learning can be broadly categorized into three major paradigms: regularization-based methods Kirkpatrick et al. (2017); Zenke et al. (2017), parameter isolation–based methods Rusu et al. (2016); Mallya and Lazebnik (2018), and rehearsal-based methods Rebuffi et al. (2017); Castro et al. (2018). Recently, research has increasingly focused on continual EEG decoding, a setting well-suited to real-world scenarios wherein static decoding models are unable to handle the continual and dynamic emergence of new subjects and tasks. Existing efforts, however, are predominantly confined to a single EEG paradigm, devising individual-level continual decoders to mitigate the continual domain shift caused by inter-subject variability Duan et al. (2023); Shahbazinia et al.

(2024); Jin et al. (2024); Li et al. (2025b); Zhou et al. (2025c; 2026). To the best of our knowledge, the exploration of EEG-based task-level continual learning remains entirely open. This presents a substantially greater challenge, as the macroscopic disparities between distinct tasks are far more pronounced than the microscopic differences among individuals. To address this, we propose EvoBrain, a dynamic and task-aware continual decoding framework. EvoBrain enables a single, unified EEG foundation model to incrementally acquire and consolidate cross-task knowledge, ultimately evolving into a general-purpose expert for EEG understanding.

B PROCEDURE

In our setting, the pretrained foundation model $\mathcal{M}_0$ sequentially adapts to all tasks in multiple fixed orders to evaluate its robustness to the input sequence. Each task's data is split into training, validation, and test sets. Specifically, for the i-th task, the model $\mathcal{M}_{i-1}$ is trained on the training set under identical conditions, and the best checkpoint is selected based on validation performance. The corresponding test performance is then reported. This selected model becomes $\mathcal{M}_i$ and serves as the starting point for subsequent continual learning. Following adaptation to a new task, we conduct a comprehensive evaluation focusing on two critical properties: the learning plasticity, reflected by the performance on the current task, and the memory stability, measured by the performance on all previously encountered tasks. The detailed algorithmic pipeline is illustrated in Algorithm 1.

Algorithm 1: Procedure of Cross-Task Continual Learning for Pretrained EEG Foundation Model

Input: Sequential EEG tasks $\{\mathcal{T}_i\}_{i=1}^{\mathcal{N}}$, pretrained EEG foundation model $\mathcal{E}$, paradigm storage $\mathcal{R}$
Output: EEG foundation model $\mathcal{E}$, set of task-specific classifiers $\{\mathcal{C}_i\}_{i=1}^{\mathcal{N}}$
Training Stage: $\mathcal{R} \leftarrow \emptyset$;
for $i \leftarrow 1$ *to* $\mathcal{N}$ **do**
  if $i = 1$ **then**
    Update the prototype of $\mathcal{T}_1$ during the first training epoch;
    Optimize $\mathcal{E}$ and $\mathcal{C}_1$ by minimizing Eq. 21;
    Register the training-sample indices of $\mathcal{T}_1$ in $\mathcal{R}$;
  else
    Sample a historical task and replay minibatch from $\mathcal{R}$ according to Eq. 4;
    Update the prototype of $\mathcal{T}_i$ during the first training epoch;
    Normalize input data via NSN by Eq. 14;
    Optimize $\mathcal{E}$ and $\mathcal{C}_i$ by Eq. 21;
    Register the training-sample indices of $\mathcal{T}_i$ in $\mathcal{R}$;
  end
end
Test Stage:
for $i \leftarrow 1$ *to* $\mathcal{N}$ **do**
  Select the task-specific classifier $\mathcal{C}_i$;
  if $i = 1$ **then**
    Predict with the raw-input anchor-task branch and $\mathcal{C}_1$;
  else
    Compute current moments from the evaluation minibatch;
    Apply task-indexed NSN using the fixed historical prototypes and learned gates;
    Predict with $\mathcal{C}_i$;
  end
end

C DETAILED DATASETS AND BASELINES

C.1 DATASET DETAILS

FACED Dataset is a large-scale, high-resolution affective computing dataset that captures nine distinct affective states: amusement, inspiration, joy, tenderness, anger, fear, disgust, sadness, and

Table 3: Overview of downstream BCI tasks and datasets.

BCI Tasks	Datasets	Rate	# Channels	Duration	# Samples	# Subjects	Label
I. Emotion Recognition	FACED Chen et al. (2023)	250Hz	32	10s	10,332	123	9-class
II. Motor Imagery Classification	PhysioNet-MI Schalk et al. (2004)	160Hz	64	4s	9,837	109	4-class
	BCIC-IV-2a Brunner et al. (2008)	250Hz	22	4s	5088	9	4-class
III. Sleep Staging	ISRUC Khalighi et al. (2016)	200Hz	6	30s	89,240	100	5-class
IV. Imagined Speech Classification	BCIC2020-3 Jeong et al. (2022)	256Hz	64	3s	6,000	15	5-class
V. Mental Disorder Diagnosis	Mumtaz2016 Mumtaz et al. (2017)	256Hz	19	5s	7,143	64	2-class

neutral. The dataset includes EEG recordings from 123 participants, acquired using a 32-channel system at a sampling rate of 250 Hz. The continuous signals were subsequently divided into non-overlapping 10-second epochs to facilitate affective state analysis.

PhysioNet-MI Dataset is a classical motor imagery dataset comprising recordings from 109 participants, acquired using a 64-channel system at an initial sampling rate of 160 Hz. The dataset encompasses four motor imagery tasks: left fist, right fist, both fists, and both feet. The original signals were segmented into 9,837 trials, each lasting 4 seconds.

BCIC-IV-2a Dataset is a motor imagery dataset comprising recordings from nine subjects performing four distinct imagery tasks: left hand, right hand, both feet, and tongue. Data were acquired over two separate sessions using a 22-channel Ag/AgCl electrode montage at a sampling rate of 250 Hz. Each session contains 288 trials, evenly distributed across the four classes. For preprocessing, we extract the 2–6 second interval from each trial and apply a band-pass filter with cutoff frequencies of 0.3 Hz and 45 Hz to attenuate noise and non-neural artifacts. This yields a total of 5,088 four-second EEG segments for analysis. Following a subject-independent split, we utilize subjects 1–5 for training, 6–7 for validation, and 8–9 for testing. Notably, although both BCIC-IV-2a and PhysioNet-MI involve four-class MI tasks, **they differ in both label space and data acquisition paradigms**.

ISRUC Dataset is a publicly available sleep dataset that includes three distinct subgroups. We specifically utilize Subgroup I of the ISRUC-Sleep dataset, which consists of 100 full-night polysomnography (PSG) recordings collected from adult participants. Electroencephalographic (EEG) signals were recorded from six standard channels (F3–A2, C3–A2, O1–A2, F4–A1, C4–A1, and O2–A1) at a sampling rate of 200 Hz. The recordings were partitioned into non-overlapping 30-second epochs, each of which was manually scored by certified sleep technicians into one of five sleep stages: Wake, N1, N2, N3, and REM, following the guidelines established by the American Academy of Sleep Medicine (AASM) Berry et al. (2017).

BCIC2020-3 Dataset is a dataset designed for imagined speech classification and includes EEG recordings from 15 participants who silently articulated one of five phrases—"hello," "help me," "stop," "thank you," or "yes"—without vocalization or movement. Data were acquired using a 64-channel system at a sampling rate of 256 Hz.

Mumtaz2016 Dataset is a dataset comprising recordings from 34 patients diagnosed with major depressive disorder (MDD) and 30 healthy controls (NCs). EEG signals were acquired from 19 electrodes positioned according to the international 10–20 system at a sampling rate of 256 Hz. The original protocol includes three recording sessions: eyes-open, eyes-closed, and a task-based session. In this work, only the eyes-open and eyes-closed resting-state sessions are utilized.

For all EEG-based downstream tasks, the signals were resampled to 200 Hz. Except for the BCIC-IV-2a dataset, all other datasets were processed and partitioned following the same protocol as CBraMod Wang et al. (2024b) to ensure consistent and comparable experimental conditions.

C.2 Baseline Details

CBraMod Wang et al. (2024b) is an EEG foundation model that employs a specialized criss-cross Transformer architecture to capture the distinct spatial and temporal dependencies inherent in EEG signals. Its backbone incorporates two parallel attention pathways: one dedicated to modeling inter-patch spatial dependencies and the other to capturing temporal dynamics. Furthermore, CBraMod introduces an asymmetric positional encoding scheme to effectively encode channel-wise positional information across heterogeneous EEG signals. It is pretrained on a large-scale dataset

using a patch-masked reconstruction strategy and achieves state-of-the-art performance across multiple downstream tasks after task-specific fine-tuning.

EEGMamba Wang et al. (2025) is an EEG foundation model that leverages a bidirectional Mamba-based architecture to effectively capture the complex spatiotemporal dependencies in EEG signals. Its backbone consists of stacked forward and backward Mamba blocks built upon the selective state space model, enabling efficient linear-time sequence modeling while capturing both forward and backward contextual dynamics. EEGMamba employs a dual-domain patch encoder that fuses time-frequency domain features and incorporates conditional positional encoding for adaptive positional awareness. The model is pretrained on a large-scale, heterogeneous EEG corpus using a patch-based masked reconstruction objective.

CSBrain Zhou et al. (2025b) is a cross-scale EEG foundation model that leverages a hierarchical architecture to capture diverse spatiotemporal neural patterns. Its backbone alternately stacks Cross-scale Spatiotemporal Tokenization (CST) to aggregate multi-scale features into scale-aware tokens, and Structured Sparse Attention (SSA) to model long-range cross-window and cross-region dependencies. By replacing dense attention with structured sparsity, CSBrain progressively integrates these dependencies to eliminate spurious correlations and enhance generalization across heterogeneous decoding tasks.

DeeperBrain Wang et al. (2026) is a neurophysiologically informed foundation model that embeds biophysical principles into both its architecture and pretraining objectives. It incorporates specific inductive biases to model volume conduction and adaptation dynamics through specialized spatiotemporal encodings. Furthermore, DeeperBrain employs a dual-objective strategy combining Masked EEG Reconstruction (MER) and Neurodynamics Statistics Prediction (NSP), which aligns learned representations with intrinsic neural mechanisms—such as functional connectivity and dynamic complexity—to ensure robust generalization across diverse decoding tasks.

D Hyperparameters and Implementation Details

Our model is trained on a single machine equipped with an Intel Core i9 10900K CPU and eight NVIDIA RTX 3080 GPUs. The architecture of task-specific classifiers is identical to that of CBraMod. Detailed training configurations are provided in Table 4.

Table 4: Hyperparameters of EvoBrain and detailed training configurations.

Category	Hyperparameter	Value
Training	Epochs	20
	Batch size	32
	Backbone learning rate	2×10^{-6}
	NSN parameter learning rate	1×10^{-3}
	Classifier learning rate	5×10^{-4}
	Weight decay	5×10^{-2}
	Dropout	0.1
	Gradient clipping norm	0.5
	Optimizer	AdamW, $\beta = (0.5, 0.99)$
	Label smoothing	0.1
NSN	Moment-gate initialization (η_μ, η_σ)	5.0
	Band-gate initialization (a_b)	0.0
	Spectral residual strength (α_m)	1.0
	Band-mask transition width (τ)	1 Hz
	Frequency bands $(\mathcal{B})$	$\delta, \theta, \alpha, \sigma, \beta, \gamma$
RAD	Classification-loss weight (λ_{ce})	0.1
	RGD/SAA trade-off weight (λ_r)	0.5
	Covariance alignment weight (ρ)	0.3

E MORE RESULTS

E.1 VISUALIZATION OF MAIN-STUDY DYNAMICS

To complement the quantitative results in the main study, Fig. 4 visualizes the performance evolution of EvoBrain across three task orders and four EEG foundation models. The task-wise trajectories show how previously learned tasks evolve as new tasks are sequentially introduced, while the aggregated results summarize the overall change from immediate post-adaptation performance to final retained performance. Across different backbones and task orders, EvoBrain generally maintains stable task performance with limited forgetting, further demonstrating its robustness to sequential task expansion and order variations.

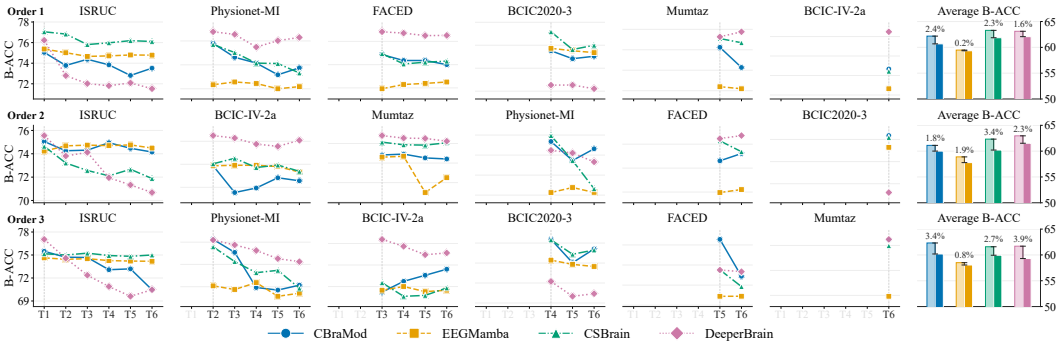

Figure 4: **Performance evolution of cross-task continual learning across three distinct task orders**. The first six columns depict the B-ACC trajectories for individual datasets as new tasks are sequentially introduced (T1–T6). The rightmost column summarizes the overall stability: the hatched (left) and solid (right) bars indicate the mean initial and final B-ACC across all tasks, respectively, with the annotated gap representing the Relative Forgetting Rate (RFR).

E.2 CONTINUAL EXPANDABILITY VERSUS STATIC FINE-TUNING

Table 5 compares single-task fine-tuning, multi-task fine-tuning, and EvoBrain using CBraMod as the shared backbone. Single-task fine-tuning achieves the highest average B-ACC by independently optimizing one backbone per task, but its training and storage costs grow linearly with the number of tasks. Multi-task fine-tuning slightly outperforms EvoBrain with one shared backbone, but requires joint access to all task data and full retraining whenever new tasks arrive. In contrast, EvoBrain adapts to tasks sequentially without full retraining, retains 91.3% of the single-task performance with a single backbone, and better supports scalable continual expansion as new EEG tasks emerge.

Table 5: Comparison of different adaptation paradigms using CBraMod as the shared EEG foundation backbone. Single-task fine-tuning trains an isolated model for each task, multi-task fine-tuning jointly optimizes all tasks, whereas EvoBrain supports sequential task expansion through continual adaptation.

Paradigm	Paradigm Property		Inference Cost		Performance						
	New-Task Update	Evolvability	Backbones	Heads	ISRUC	PhysioNet-MI	FACED	BCIC2020-3	Mumtaz	BCIC-IV-2a	Avg. B-ACC / Retained
Single-Task FT	Train new model	✗	$N\times$	$N\times$	0.787±0.010	0.642±0.009	0.550±0.009	0.537±0.011	0.956±0.006	0.514±0.007	0.664±0.009 / –
Multi-Task FT	Retrain on all tasks	✗	$1\times$	$N\times$	0.759±0.007	0.608±0.011	0.545±0.020	0.421±0.020	0.889±0.007	0.521±0.013	0.624±0.003 / 94.0%
EvoBrain	Incremental update	✓	$1\times$	$N\times$	0.729±0.004	0.546±0.007	0.475±0.004	0.482±0.007	0.876±0.010	0.527±0.008	0.606±0.002 / 91.3%

E.3 VISUALIZATION OF CROSS-TASK REPRESENTATION EVOLUTION

Fig. 5 visualizes the longitudinal evolution of target-task latent spaces and B-ACC scores during continual adaptation. We analyze PhysioNet-MI and BCIC-IV-2a with CBraMod, and FACED with DeeperBrain, covering both within-paradigm and cross-paradigm transfer. The visualization reveals two main trends. First, EvoBrain can produce positive forward transfer before direct target-task

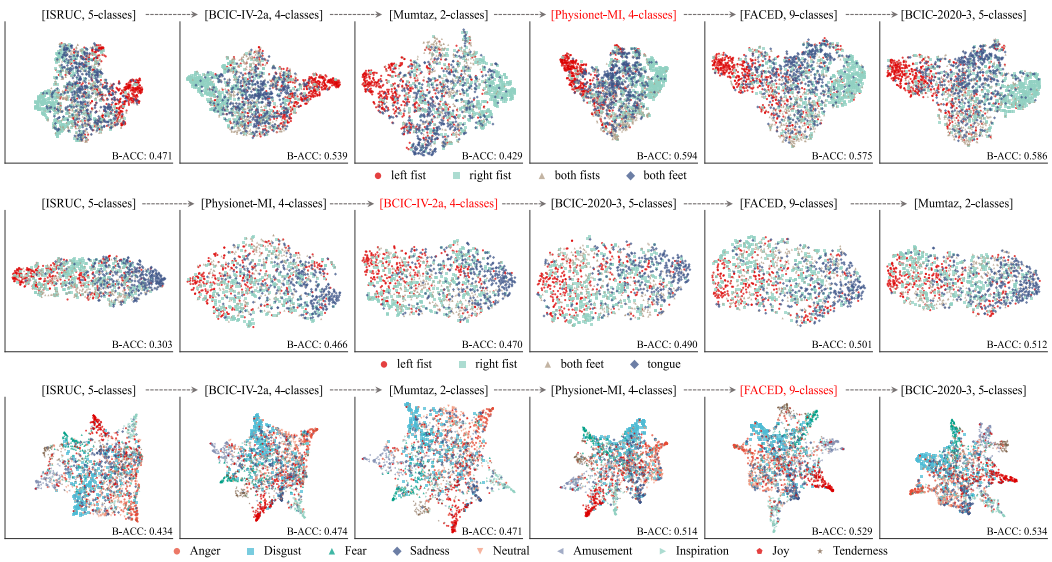

Figure 5: Longitudinal evolution of latent representations and classification performance throughout the cross-task continual learning process. For each target task, the classifier learned when that task entered the stream is held fixed and retrospectively paired with backbone checkpoints from different stages. The first and second rows visualize the PhysioNet-MI and BCIC-IV-2a datasets using CBraMod, respectively, while the third row visualizes the FACED dataset using DeeperBrain. The stages marked in red denote the onset of training for the corresponding target tasks. The reported Balanced Accuracy (B-ACC) scores show that related tasks can induce positive transfer before explicit target-task adaptation, as reflected by the early improvement between the two motor imagery datasets (PhysioNet-MI: $0.471 \rightarrow 0.539$ after BCIC-IV-2a adaptation; BCIC-IV-2a: $0.303 \rightarrow 0.466$ after PhysioNet-MI adaptation). After target-task adaptation, the final B-ACC remains high for PhysioNet-MI, BCIC-IV-2a, and FACED (0.586, 0.512, and 0.534).

training: PhysioNet-MI improves from 0.471 to 0.539 after BCIC-IV-2a, BCIC-IV-2a improves from 0.303 to 0.466 after PhysioNet-MI, and FACED improves from 0.434 to 0.514 after four non-emotion tasks. Second, target-task representations remain recoverable and can continue to improve after their own adaptation stage, with final B-ACC reaching 0.586 for PhysioNet-MI, 0.512 for BCIC-IV-2a, and 0.534 for FACED. These results suggest that EvoBrain progressively refines task-discriminative latent spaces rather than overwriting previous representations.

E.4 Visualization of Spectral-Modulation Patterns

To quantify the frequency-wise effect induced by the residual modulation, we further define a relative linear power difference:

$$\Delta_{\mathrm{rel}}(f) = \frac{P_{\mathrm{NSN}}(f) - P_{\mathrm{PTN}}(f)}{\frac{1}{|\Omega|}\sum_{f' \in \Omega} P_{\mathrm{PTN}}(f')}, \quad \Omega = [0.5, 45]\mathrm{Hz}, \tag{23}$$

where $P_{\mathrm{PTN}}(f) = |\,\mathrm{rFFT}(\mathbf{X}_{\mathrm{tn}})(f)|^2$ and $P_{\mathrm{NSN}}(f) = |\,\mathrm{rFFT}(\mathbf{X}_{\mathrm{nsn}})(f)|^2$. This quantity measures the normalized power change from the temporally normalized representation to the NSN-modulated representation. A positive value of $\Delta_{\mathrm{rel}}(f)$ indicates that SRM amplifies the response around frequency f, whereas a negative value indicates suppression. Therefore, $\Delta_{\mathrm{rel}}(f)$ provides an interpretable frequency-wise view of how the learned task-conditioned gates reshape spectral responses.

We visualize this frequency-wise effect of Neuro-Spectral Task Normalization (NSN) by examining how its Spectral Residual Modulation (SRM) component reshapes task-specific EEG representations. The visualization is conducted on CBraMod under the Order 1 continual learning stream. For each task, we use the final student model and compute the relative linear spectral modulation $\Delta_{\mathrm{rel}}(f)$ defined in Eq. 23. We report the mean curve over five repeated CBraMod runs, with the shaded region indicating the 95% confidence interval. The results suggest that SRM learns task-conditioned

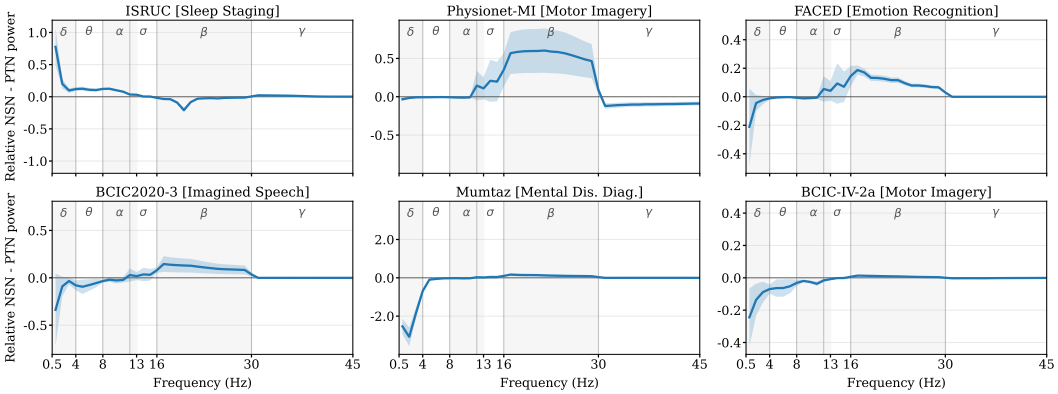

Figure 6: Visualization of NSN-induced relative linear spectral modulation over five repeated CBraMod runs under Order 1. For each task, we use the final student model after continual learning and compute the frequency-wise power change induced by the Spectral Residual Modulation (SRM) component of NSN. The y-axis reports $\Delta_{\mathrm{rel}}(f)$, as defined in Eq. 23, which measures the normalized frequency-wise power change induced by NSN relative to the temporally normalized representation. The solid line denotes the mean over five repeated CBraMod runs, and the shaded region denotes the 95% confidence interval. Vertical bands indicate canonical EEG frequency ranges: δ, θ, α, σ, β, and γ.

spectral recalibrations that are qualitatively consistent with canonical EEG rhythm priors. For sleep staging, SRM emphasizes the low-frequency δ range, aligning with the relevance of slow-wave activity in sleep analysis. For motor imagery, PhysioNet-MI shows a clear β-band enhancement, whereas BCIC-IV-2a mainly suppresses low-frequency components with only mild β-band amplification. For emotion recognition, SRM also increases the β range, suggesting that it tends to emphasize potentially discriminative spectral components for affective EEG representations. Overall, these patterns indicate that NSN performs task-conditioned residual spectral adjustment rather than uniform normalization. This analysis provides a model-level interpretation of the learned SRM behavior: the observed correspondence with EEG rhythm priors offers qualitative support for the interpretability of NSN, rather than direct neurophysiological evidence.

E.5 Visualization of Relation-Structure Preservation.

We further visualize whether RAD preserves the internal response geometry of previously learned tasks by comparing the teacher–student relation matrices on replayed old-task samples. Specifically, for the first three tasks in the Order-1 stream, i.e., ISRUC, PhysioNet-MI, and FACED, we use the corresponding teacher checkpoints $\mathcal{M}_i$ and the final student model $\mathcal{M}_6$. Specifically, we extract the intermediate responses from the teacher and student and compute the sample-wise cosine relational matrices as defined in Eq. 15. We then visualize the absolute difference $|R_{\mathrm{student}} - R_{\mathrm{teacher}}|$, where lower error indicates that the final student better preserves the old-task response geometry encoded by the frozen teacher. The reported `offdiag_mse` is computed only on the off-diagonal entries, exactly following the RGD objective in Eq. 16. Figure 7 visualizes old-task relation preservation under RAD. Full EvoBrain consistently keeps relation errors low across the first three old tasks, while the variant that removes SAA remains close to the full model, suggesting that removing SAA alone does not substantially disrupt old-task relational geometry. In contrast, removing RGD produces stronger error bands and larger off-diagonal MSE across all three datasets, with the most evident degradation on PhysioNet-MI, where the error increases from 0.0020 to 0.0051. The consistent trend observed on ISRUC and FACED further identifies RGD as the key factor for preserving old-task relational structures. This result shows that merely replaying historical samples is insufficient, since replay provides access to old-task data but does not explicitly constrain the geometry of their responses. RGD addresses this limitation by aligning the student's replay-sample relation matrix with that of the corresponding teacher, thereby suppressing inter-sample relational drift during continual adaptation.

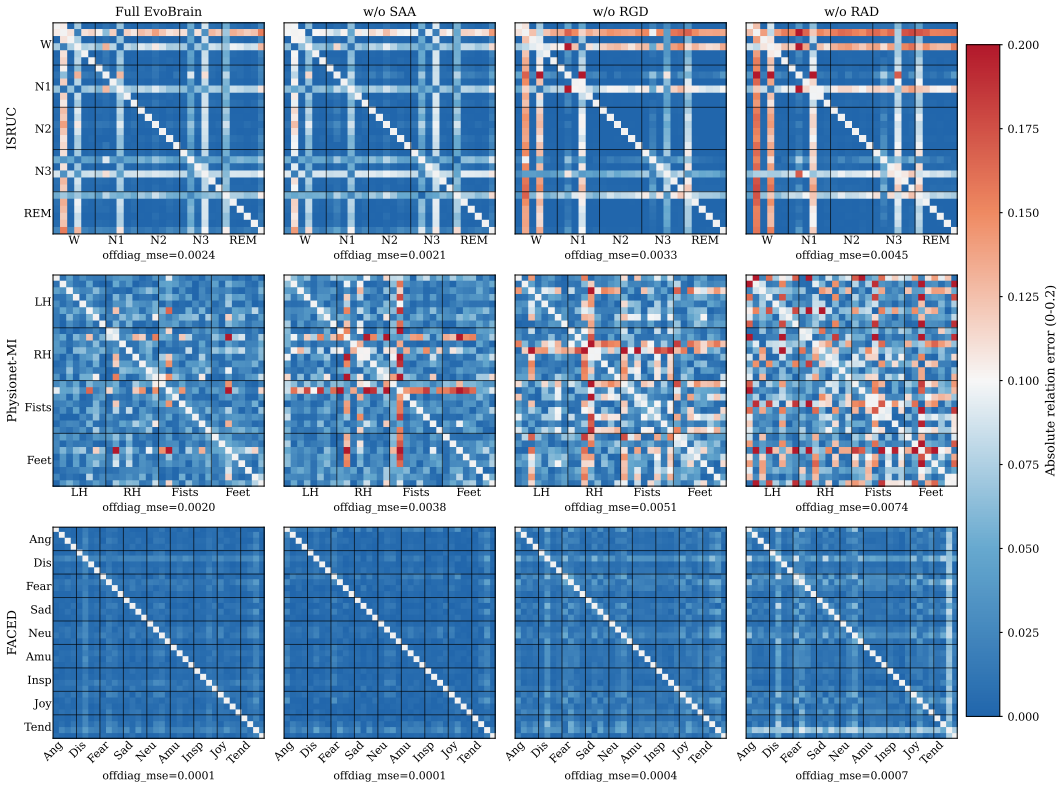

Figure 7: Visualization of old-task relation preservation under RAD. We use CBraMod under the Order-1 task stream and select the first three old tasks, ISRUC, PhysioNet-MI, and FACED. For each task T_i ($i \in \{1, 2, 3\}$), $\mathcal{M}_i$ denotes the frozen teacher after adapting to T_i, and $\mathcal{M}_{\mathcal{N}}$ denotes the final student after the full six-task sequence. Following Eqs. 15–16, we compute sample-wise cosine relation matrices from pooled intermediate responses of the teacher and student on the same replay samples, and visualize their absolute difference $|R_{\text{student}} - R_{\text{teacher}}|$. The value below each subplot reports the off-diagonal mean squared error (`offdiag_mse`), where smaller values and colder colors indicate better preservation of old-task response geometry. Full EvoBrain and w/o SAA consistently show lower relation errors than w/o RGD and w/o RAD, especially on ISRUC and PhysioNet-MI, demonstrating that RGD is the key component for stabilizing old-task relational structures, while SAA provides complementary response-statistics alignment.

F Discussion

F.1 Implications

The present results suggest that cross-task continual adaptation is a promising post-training paradigm for EEG foundation models. Unlike task-isolated fine-tuning, which repeatedly optimizes and stores separate backbones for different downstream tasks, EvoBrain enables a single backbone to evolve across sequentially arriving heterogeneous BCI tasks. Its ability to retain 91.3% of the single-task fine-tuning performance indicates that scalable one-for-all EEG decoding requires not only larger pretraining corpora, but also post-training mechanisms that support continual task expansion without restarting adaptation from the pretrained model. The effectiveness of NSN further shows that continual EEG adaptation should be treated as both a distributional and neuro-spectral alignment problem. By combining prototype-guided temporal normalization with task-conditioned spectral residual modulation, NSN stabilizes cross-task feature statistics while preserving task-specific frequency responses, moving beyond generic feature normalization toward EEG-specific spectral recalibration. Meanwhile, RAD demonstrates that stability in heterogeneous task streams depends on preserving relational response structure and enabling selective transfer, rather than merely constraining individual predictions or parameters. Together, these findings position Evo-

Brain as a neuro-spectrally grounded continual adaptation framework, where plasticity is achieved through task-adaptive spectral recalibration and stability is maintained through affinity-aware response preservation.

F.2 LIMITATIONS

Several directions could further extend EvoBrain toward more open BCI applications. The current study adopts a controlled task-incremental protocol with known task boundaries and task-specific heads, which enables systematic evaluation of cross-task continual adaptation. Future work may consider more flexible online settings with ambiguous task identity, delayed labels, dynamic head selection, or unified output spaces. In addition, although replay memory effectively improves stability, privacy-aware alternatives such as federated continual learning, prototype replay, and generative replay could further facilitate deployment in clinical or multi-institutional environments. Finally, the spectral modulation analysis provides model-level interpretability, and future studies with controlled paradigms and subject-level validation may further strengthen its neurophysiological grounding.