TaKF$^+$: A VERSATILE AND PARAMETER-EFFICIENT TUNING FOR EEG FOUNDATION MODEL

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalogram (EEG) data, widely used in brain-computer interfaces (BCIs), pose challenges for reusing deep learning models trained on specific datasets due to variations in recording configurations and domain gaps. While foundation models pre-trained on large-scale EEG datasets have emerged as a promising solution, the challenge of effectively adapting them to downstream tasks has yet to be fully explored. To address this, we propose a novel tuning method, TaKF$^+$, which consists of the Task-Adaptive Key-Feature Extractor (TaKF) and adapter modules. TaKF$^+$ is designed to efficiently extract task-relevant features from EEG foundation models for downstream tasks while preserving the model's parameters and significantly reducing computational overhead. We evaluate TaKF$^+$ across a diverse range of tasks, including motor imagery, emotion recognition, and seizure detection, and demonstrate its superior performance and adaptability compared to existing methods over publicly available datasets. Our research paves the way for more efficient and versatile applications of EEG foundation models across various domains.

1 INTRODUCTION

An electroencephalogram (EEG) is a record of changes in the membrane potential, which is obtained from electrodes placed on the scalp surface. EEG is widely used in brain-computer interfaces (BCIs) due to its non-invasive nature and ability to capture meaningful information related to neural activity, making it a valuable tool for interpreting brain states (Phyo et al., 2022; Kim et al., 2024; Panda et al., 2010). Recently, deep learning has become a standard for BCI-based EEG analysis with remarkable performance. Specifically, utilizing neural networks with supervised learning can significantly enhance the performances of diverse EEG tasks by capturing the inherent patterns of EEG (Lawhern et al., 2018; Song et al., 2022). However, despite the existence of the standard international 10-20 system for EEG data recording, heterogeneity in configurations—such as the number of channels, electrode placements, and sampling rates—across different measurement institutions creates a domain gap (Dornhege et al., 2004; Brunner et al., 2008). Additionally, there is also a gap caused by the high variability between subjects and datasets (Ko et al., 2022). These gaps ultimately limit the generalizability of dataset-specific trained models. In practice, these challenges result in limitations, such as the need for large amounts of data and increased computational costs, as neural networks often need to be retrained for new tasks or different patients to maintain performance.

Transfer learning-based deep learning models for EEG have been introduced to mitigate these challenges (Zhang et al., 2017). To enhance the generalization of models across different datasets, transfer learning utilizes prior knowledge of models, resulting in reduced dependency on large amounts of task-specific data and lower computational cost associated with retraining neural networks for new tasks or subjects (Cai et al., 2023). Generally, through the pre-training process, transfer learning acquires general representations that can be applied across subjects (Ko et al., 2024). In particular, the foundation model, a pretrained neural network that has learned general features from large datasets for various downstream tasks, can be a powerful breakthrough (Zhou et al., 2023). With this in mind, attempts have been made to develop transformer-based EEG foundation models using masked prediction as a pretext task (Yang et al., 2024; Zhang et al., 2024; Jiang et al., 2024). These efforts overcome the heterogeneities in configurations and dissolve inter-dataset variability, pointing towards developing EEG foundation models that can be easily used for new tasks.

However, while research on developing effective pretext tasks for EEG foundation model pre-training is steadily increasing, insufficient consideration has been given to how these EEG foundation models can be applied to downstream tasks. Existing EEG foundation models perform downstream tasks by fully fine-tuning all their parameters, which inevitably entails high computational costs and can lead to a degradation of generalization ability (Kumar et al., 2022). Parameter-efficient fine-tuning (PEFT) methods, which adapt the foundation model to new tasks by modifying only a tiny portion of parameters, offer a potential solution (Zaken et al., 2021). Especially in clinical applications, EEG foundation models must be adapted to specific patients due to the crucial importance of stability and accuracy. In these situations, additive fine-tuning, which updates only additional modules without altering the parameters of the foundation model (Han et al., 2024), is well-suited in terms of storage and financial costs. However, our analysis shows two limitations when applying additive fine-tuning methods to EEG foundation models. One limitation is, as shown in Figure 1, the inconsistency in performance that occurs depending on the EEG foundation model or the downstream task. In particular, these methods show a large dependency on which datasets are used for EEG foundation model pre-training. Another limitation is poor few-shot performance, a critical issue given the difficulty obtaining sufficient EEG data.

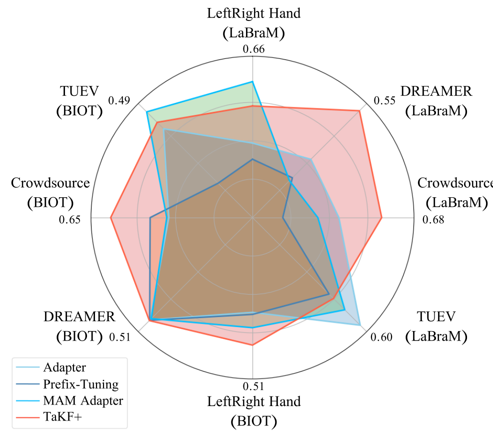

Figure 1: The overall performance comparison of TaKF$^+$ and other additive tuning methods across diverse EEG downstream tasks.

Therefore, in this paper, we propose a novel additive fine-tuning method, TaKF$^+$, which can be applied task-agnosticically and is robust in few-shot scenarios. To be a versatile tool for diverse downstream tasks, TaKF$^+$ is divided into two parts: the Task-Adaptive Key-Feature Extractor (TaKF), which enhances the neural network's expressiveness, and the adapter modules, which modify the prior knowledge of the EEG foundation model in a task-specific manner. Notably, TaKF has the advantage of data efficiency by extracting task-relevant features in a lower-dimensional space than the original dimensions of the EEG foundation model. We evaluate the effectiveness of our proposed method on a wide range of downstream tasks, such as seizure detection, emotion recognition, and motor imagery. Consequently, we demonstrate that our proposed method can serve as a versatile and task-agnostic tool and excels in few-shot scenarios. The contributions of this work are summarized as follows:

- We propose a novel additive fine-tuning method, called TaKF$^+$, specifically designed to obtain task-relevant key features from two complementary views through the TaKF and adapter modules, which can provide a versatile solution for a broad spectrum of downstream tasks with EEG.
- We achieve significant improvements in data-scarce environments and mitigate dependency on labeled data by proposing the TaKF module, which operates in a low-dimensional space using learnable low-dimensional queries.
- We validate the effectiveness and efficiency of the proposed method through extensive experimentation, demonstrating its superior performance across various applications, including seizure detection, emotion recognition, and motor imagery over four publicly available datasets.

2 Related works

2.1 Foundation models for EEG signals

Foundation models are pre-trained neural networks with large datasets to learn general representation, so that they can work well in a wide range of downstream tasks (Zhou et al., 2023). The remarkable efficacy of foundation models in enhancing EEG-based BCI systems has especially garnered notable attention. Kostas et al. (2021) developed a universal pre-trained model, BENDR, trained using contrastive self-supervised learning, for wider EEG-based analysis. Cai et al. (2023) suggested a self-supervised learning framework that can learn implicit spatial and temporal correlations through pretext tasks reflecting the characteristics of brain signals. Kostas et al. (2021) and Cai et al. (2023) attempted to advance towards large models by training on massive-sized source datasets, but they did not address diverse configurations of EEG signals. Wang et al. (2023) proposed a reusable model named BrainBERT, pre-trained using a masking strategy, to provide embeddings for intracranial recordings. Brant, devised by Zhang et al. (2024), is a general and large-scale model that learns the long-term temporal dependency and spatial correlation of intracranial neural signals for adaption across a wide range of tasks. Brant and BrainBERT are designed to handle signals with different numbers of channels using a single model, but they were pre-trained on only a single dataset. Yang et al. (2024) demonstrated that the BIOT model, trained on multiple datasets with different formats, performs well across a wide range of downstream tasks. The foundation model proposed by Jiang et al. (2024), LaBram, trained via self-supervised learning using a tokenizer on tremendous EEG data, exhibited excellent performance in out-of-source datasets. On the other hand, by pointing out the weaknesses of the masked prediction pretext task, Foumani et al. (2024) introduces a self-prediction approach, EEG2Rep, which enables the generation of rich semantic representations. Despite these advancements, there have been few in-depth discussions on how to effectively apply these models to diverse downstream tasks.

2.2 Parameter-efficient fine-tuning

Fine-tuning pre-trained models for specific tasks is a common and effective way to enhance learning by leveraging knowledge from related domains (Yosinski et al., 2014; Ko et al., 2024). However, as the parameter scale of pre-trained models continues to grow (Kaplan et al., 2020), the inefficiencies and costs associated with fine-tuning have been pointed out as critical issues (He et al., 2021). Recently, PEFT, which updates only a tiny proportion of parameters while freezing the rest of the pre-trained model, has emerged as a solution to this issue (Houlsby et al., 2019; Zaken et al., 2021). Among the proposed PEFT methods, additive fine-tuning algorithms adopt additional trainable modules and only tune these added modules while keeping the pre-trained model unchanged (Han et al., 2024). The adapter-form approach, initially introduced by the Adapter (Houlsby et al., 2019), is the most well-known concept among additive fine-tuning algorithms. An adapter module, used in the adapter-form approach, plays a key role in encoding the representations in the intermediate layers into a task-specific form. Alternatively, the soft prompt approach improves performance through the continuous embedding space (Petrov et al., 2023). Li & Liang (2021) introduces learnable vectors that are prepended to the keys and values in the multi-head attention mechanism of transformer blocks, making models more expressive. Jia et al. (2022) demonstrates that the prefix-tuning method is adaptable to transformer-based vision models. Apart from these, He et al. (2021) explores the connections between additive fine-tuning approaches and proposes a hybrid form known as the MAM Adapter, which integrates Adapter and Prefix-Tuning. However, despite many advancements, these approaches have not yet been validated for EEG-related pre-trained models or their suitability for BCI tasks.

3 Preliminary

3.1 Adapter-form approach

The adapter-form approach (Houlsby et al., 2019) employs a trainable module (adapter) which is inserted in transformer layers. The adapter generally consists of a down-projection matrix $\boldsymbol{W}_{\text{down}} \in \mathbb{R}^{d\times r}$, followed by a non-linear activation function $\sigma(\cdot)$, and an up-projection matrix $\boldsymbol{W}_{\text{up}} \in \mathbb{R}^{r\times d}$, which first projects the input vector $\boldsymbol{h}$ to a lower-dimensional space of size r, then projects it back to the original dimension d. The adapter process operates as:

$$\boldsymbol{h} \leftarrow \boldsymbol{h} + \sigma(\boldsymbol{h}\boldsymbol{W}_{\text{down}})\boldsymbol{W}_{\text{up}}. \tag{1}$$

These methods utilizing adapter modules have been shown in the natural language processing (NLP) field to achieve comparable performance to full fine-tuning with only a few tunable parameters. Depending on the purpose, the architecture and usage of adapters can be modified, as seen in Parallel Adapter (He et al., 2021), AdaMix (Wang et al., 2022), and AdapterSoup (Chronopoulou et al., 2023).

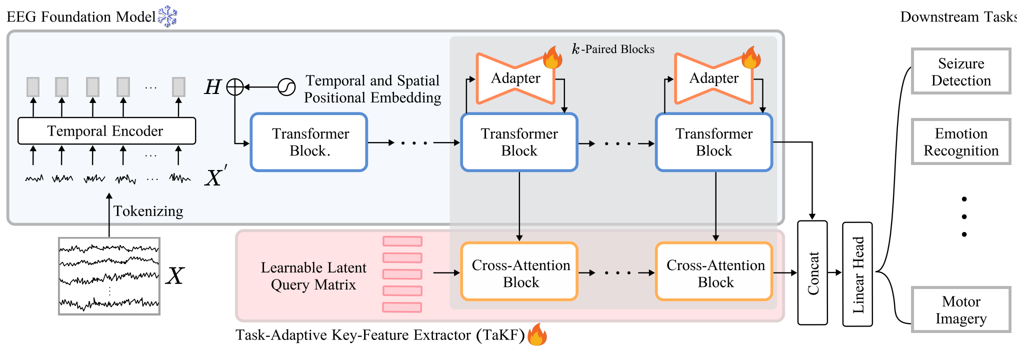

Figure 2: Overview of the proposed framework, TaKF^{+}. TaKF^{+} includes the EEG foundation model, the Task-Adaptive Key-Feature Extractor (TaKF), and adapter modules. Only the parameters of the TaKF and adapter modules are updated during downstream tasks, while the EEG foundation model remains frozen during fine-tuning

3.2 Prefix-Tuning

Motivated by the progress of prompt-based learning methods in the NLP field, prefix-tuning (Li & Liang, 2021) introduces n prefix vectors, which have a dimension of d, and concatenates (Concat) them with keys and values in multi-head attention (MHA) layer. Specifically, the n prefix vectors for the key, $\boldsymbol{P}_k \in \mathbb{R}^{n \times d}$, and for the value, $\boldsymbol{P}_v \in \mathbb{R}^{n \times d}$, are appended to the original key $\boldsymbol{K}$ and value $\boldsymbol{V}$, either at the first or at every attention (Attn) layer (Jia et al., 2022). Then, multi-head attention is applied to the concatenated keys and values, with the computation for each head being as follows:

$$\mathrm{MHA}(\boldsymbol{C}, \boldsymbol{x}, \boldsymbol{P}_k, \boldsymbol{P}_v) = \mathrm{Concat}(\mathrm{head}_1, \dots, \mathrm{head}_h)\boldsymbol{W}_o, \tag{2}$$

$$\mathrm{head}_i = \mathrm{Attn}(\boldsymbol{x}\boldsymbol{W}_q^{(i)}, \mathrm{Concat}(\boldsymbol{P}_k^{(i)}, \boldsymbol{C}\boldsymbol{W}_k^{(i)}), \mathrm{Concat}(\boldsymbol{P}_v^{(i)}, \boldsymbol{C}\boldsymbol{W}_v^{(i)})), \tag{3}$$

where $\boldsymbol{W}_o \in \mathbb{R}^{d \times d}$, $\boldsymbol{W}_q^{(i)}, \boldsymbol{W}_k^{(i)}, \boldsymbol{W}_v^{(i)} \in \mathbb{R}^{d \times d/N_d}$ denotes projection matrix of query, key, and value, $\boldsymbol{C} \in \mathbb{R}^{m \times d}$ represents a given sequence of m vectors, $\boldsymbol{x} \in \mathbb{R}^d$ indicates a query vector, and head_i refers to the i-th head's vectors, and $\boldsymbol{P}_k^{(i)}$ and $\boldsymbol{P}_v^{(i)} \in \mathbb{R}^{n \times d/N_d}$ correspond to the respective parts of $\boldsymbol{P}_k$ and $\boldsymbol{P}_v$. Notably, its superiority is particularly evident in low-data settings with a limited parameter budget.

4 Method

4.1 Overall Framework

Given diverse downstream tasks $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_S\}$, which can correspond to specific datasets or subjects, we aim to address $\mathcal{T}$ by leveraging the EEG foundation model effectively. For a specific task $\mathcal{T}_s$, our framework is divided into a frozen EEG foundation model with pre-trained weights θ and additive tunable modules with learnable parameter ϕ_s, where ϕ_s has a significantly smaller parameter size compared to θ. We only update the learnable parameters ϕ_s. Since the approach to adapting the EEG foundation model may vary depending on $\mathcal{T}_s$, we divise the additive tunable modules of two complementary components to reduce this variability: a TaKF, which makes the model more expressive for capturing new task-relevant patterns, and adapter modules, which transform the prior knowledge of the EEG foundation model into a task-specific form.

As shown in Figure 2, an input EEG signal $\boldsymbol{X} \in \mathcal{T}_s$ is passed through the EEG foundation model for decoding. The input EEG signal is embedded through two pathways: (1) transformer blocks combined with adapter modules and (2) the TaKF. The two features, embedded through separate pathways, are concatenated and used for prediction. This process updates the additive tunable parameters ϕ_s. As a result, we can handle $\mathcal{T}$ by using the pre-trained weights θ and $\phi_1, \phi_2, \dots, \phi_S$, which are obtained from each downstream process.

4.2 EEG FOUNDATION MODEL

The EEG foundation model f_θ, consisting of a patch embedding encoder and L transformer encoder blocks, is a neural network that has learned EEG-related representations through pretext tasks, such as masked patch prediction (Zhou et al., 2023). The input EEG signal $\boldsymbol{X} \in \mathbb{R}^{c\times t}$, where c represents the number of electrodes (channels) and t represents the number of time steps in the signal, is tokenized by segmenting it into patches of size $1 \times w$, resulting in a reshaped input signal $\boldsymbol{X}' \in \mathbb{R}^{(c\times\lfloor t/w \rfloor)\times w}$. The segmented input is then encoded into d-dimensional tokens $\boldsymbol{H} \in \mathbb{R}^{(c\times\lfloor t/w \rfloor)\times d}$ using a temporal encoder. Temporal and spatial positional embedding vectors, as positional encodings (Vaswani et al., 2017), are added to these tokens. The tokens are then passed through L transformer encoder blocks and concatenated with the output of the TaKF.

4.3 TASK-ADAPTIVE KEY-FEATURE EXTRACTOR

The TaKF aims to extract task-relevant information from the EEG foundation model and compress it into low-dimensional features. To perform this role, the TaKF utilizes cross-attention blocks to extract important features from the representation features and project them into the learnable latent query matrix. The TaKF has J cross-attention blocks and a learnable latent query matrix $\boldsymbol{Q}_0 \in \mathbb{R}^{N\times r}$, which consists of N learnable latent query vectors $\boldsymbol{q}_0 \in \mathbb{R}^{1\times r}$. Each learnable latent query vector $\boldsymbol{q}_0$ is a randomly initialized r-dimensional learnable parameter, where r is less than the dimension d of the representation features of the EEG foundation model. The cross-attention blocks are paired with the last J transformer blocks of the EEG foundation model, as high-level features are closely related to the targets, while low-level features capture common characteristics (Ren et al., 2023). The query matrix that passes through the J cross-attention-based blocks is used as the input to the linear head.

Cross-Attention Block. The ability of the cross-attention to extract and compress large input information into a low-dimensional array has been validated in numerous works, including Perceiver (Jaegle et al., 2021) and Q-former (Li et al., 2023). Furthermore, cross-attention offers parameter-efficient advantages in mapping high-dimensional vectors to low-dimensional vectors. Therefore, we adopt the cross-attention for extracting task-relevant features from the large and high-dimensional representation features of the EEG foundation model.

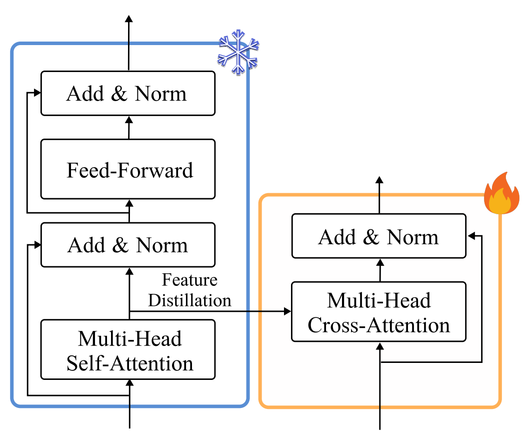

Figure 3: The illustration of a paired block in TaKF$^+$. The cross-attention block uses the output of self-attention within the paired transformer block as the key and value for cross-attention. In the figure, the adapter module attached to the transformer block was omitted.

As shown in Figure 3, the cross-attention block consists of a cross-attention and residual connection. The first cross-attention block uses the initial latent query matrix $\boldsymbol{Q}_0$, which is learnable and stored within the neural network, as queries for cross-attention. In the j-th cross-attention block, for $j > 0$, the latent query matrix $\boldsymbol{Q}_{(j-1)}$, which have passed through $(j-1)$ preceding cross-attention blocks, act as queries for cross-attention. For the keys and values in the cross-attention, the representation feature $\boldsymbol{H}_l \in \mathbb{R}^{(c\times\lfloor t/w \rfloor)\times d}$, obtained from the l-th transformer block of the EEG foundation model, is used. In this case, the l-th transformer block is paired with the j-th cross-attention block. Before using the latent query matrix and the representation feature as queries, keys, and values, Layer Normalization (Ba, 2016) is applied to each to ensure training stability (Dehghani et al., 2023). After the cross-attention process, a residual connection (He et al., 2016) is applied to the output. The process of j-th cross-attention (Cross-Attn) block can be described as follows:

$$\text{Cross-Attn}(\boldsymbol{Q}_{(j-1)}, \boldsymbol{H}_l) = \text{Softmax}\left(\frac{\text{LN}(\boldsymbol{Q}_{(j-1)})\boldsymbol{W}_q \cdot (\text{LN}(\boldsymbol{H}_l)\boldsymbol{W}_k)^T}{\sqrt{r}}\right)\text{LN}(\boldsymbol{H}_l)\boldsymbol{W}_v, \quad (4)$$

$$\boldsymbol{Q}_j = \boldsymbol{Q}_{(j-1)} + \text{Cross-Attn}(\boldsymbol{Q}_{(j-1)}, \boldsymbol{H}_l), \quad (5)$$

where $\text{LN}(\cdot)$ denotes Layer Normalization, and $\boldsymbol{W}_q \in \mathbb{R}^{r\times r}$, $\boldsymbol{W}_k \in \mathbb{R}^{d\times r}$, and $\boldsymbol{W}_v \in \mathbb{R}^{d\times r}$ are the query, key, and value projection matrices, respectively.

Feature Distillation from Before the Feed-forward Layer. When passing information from the EEG foundation model to the TaKF, we used the representation features taken from before the feed-forward layer of the transformer block in the EEG foundation model. We aimed to find a way for the EEG foundation model to pass information to the TaKF module to effectively capture task-relevant features for tasks that are not closely related to the prior knowledge of the EEG foundation model. Our analysis confirmed that the performance of TaKF varies significantly depending on whether the representation features are taken before or after the feed-forward layer, as detailed in Appendix C. Therefore, to maximize the intended functionality of TaKF, we set the feature distillation position.

4.4 Adapter Module

For tasks closely related to the pre-training dataset, the EEG foundation model already contains sufficient relevant information, while capturing task-relevant features through the TaKF module can be highly effective for tasks not directly associated with the pre-training dataset. In such cases, the adapter, which transforms the representations of intermediate layers in transformer blocks into task-specific features (Houlsby et al., 2019), is a powerful tool. Therefore, we integrated adapter modules into our framework. Inspired by Mix-adapter (He et al., 2021), we use the adapter module exclusively in parallel with the feed-forward layer of the transformer blocks.

5 Experiment

5.1 Datasets

We employ four EEG datasets to validate our proposal. 1) **Left/Right Hand Motor Imagery (Left-Right Hand)** is a binary dataset with two labels: left-hand and right-hand (Zakrzewski et al., 2022). 2) **DREAMER** (Katsigiannis & Ramzan, 2017) contains signals related to a subject's affective state, including values for valence, arousal, and dominance. 3) **Crowdsourced** is a binary dataset (eyes open vs. eyes closed) from the EMOTIV platform (Williams et al., 2023). 4) **TUEV** Obeid & Picone (2016) contains signals classified into six abnormal-related events. Details of the datasets is described in Appendix A.1. The code can be accessed on Github[1].

5.2 Preprocessing and Evaluation Metrics

We follow essential preprocessing steps from LaBraM (Jiang et al., 2024). Further details are provided in the Appendix A.2. For the data division, we followed the approach described below: 1) **DREAMER**, **LeftRight Hand**, and **Crowdsourced**: we evaluate them using 5-fold cross-validation based on subjects. 2) **TUEV**: The division for training and test sets is provided by the dataset. For the validation group, we split the training set into an 80%:20% ratio based on subjects. We use balanced accuracy (BACC) and the area under the ROC curve (AUROC) as evaluation metrics for binary classification datasets. For multi-class tasks, we adapt balanced accuracy (BACC) and Cohen's kappa (Cohen's κ).

5.3 Baseline

We evaluate our method using three categories: 1) supervised modeling methods (SMM), 2) self-supervised modeling methods (Self-SMM), and 3) additive fine-tuning methods. The supervised modeling methods, which are simply fine-tuned on each downstream dataset, include SPaRCNet

[1]https://anonymous.4open.science/r/TaKFplus

Table 1: Performance comparison to the competitors on LeftRight Hand, DREAMER, Crowdsourced, and TUEV datasets, using LaBraM as the base EEG foundation model. FT and LP denote fine-tuning and linear probing, respectively, while PT represents Prefix-Tuning. Bold values represent the best results, while underlined values indicate the second-best.

	LeftRight Hand		DREAMER	
	BACC	AUROC	BACC	AUROC
SMM SOTA	69.06 ± 5.16	77.75 ± 6.02	56.25 ± 4.10	59.04 ± 7.17
Self-SMM SOTA	60.26 ± 5.79	66.34 ± 8.18	55.63 ± 2.85	58.02 ± 2.74
LaBraM-FT (Jiang et al., 2024)	70.01 ± 4.36	77.71 ± 5.73	55.67 ± 3.64	59.60 ± 4.79
LaBraM-LP	52.06 ± 2.10	53.83 ± 3.07	50.51 ± 1.13	53.20 ± 3.23
LaBraM-Adapter (Houlsby et al., 2019)	62.91 ± 4.83	70.72 ± 7.58	$\underline{53.61 \pm 3.48}$	$\underline{57.47 \pm 6.68}$
LaBraM-PT (Li & Liang, 2021)	62.28 ± 3.11	68.88 ± 4.75	53.10 ± 3.01	57.10 ± 3.90
LaBraM-MAM Adapter (He et al., 2021)	$\mathbf{65.31 \pm 4.74}$	$\mathbf{73.02 \pm 6.99}$	53.02 ± 3.32	54.85 ± 3.94
(Ours) LaBraM-TaKF$^+$	$\underline{64.36 \pm 5.66}$	$\underline{71.66 \pm 7.72}$	$\mathbf{54.95 \pm 3.84}$	$\mathbf{59.31 \pm 4.86}$
	Crowdsourced		**TUEV**	
	BACC	AUROC	BACC	Cohen's κ
SMM SOTA	60.71 ± 11.99	74.99 ± 9.32	43.84 ± 3.49	39.12 ± 2.37
Self-SMM SOTA	65.77 ± 12.84	69.27 ± 3.08	53.37 ± 1.10	52.61 ± 2.44
LaBraM-FT	62.04 ± 10.51	65.30 ± 11.15	64.09 ± 0.65	66.37 ± 0.93
LaBraM-LP	54.95 ± 3.23	68.11 ± 7.67	34.61 ± 2.25	39.68 ± 3.29
LaBraM-Adapter	$\underline{65.37 \pm 11.17}$	$\underline{74.75 \pm 5.89}$	$\mathbf{59.86 \pm 0.98}$	$\mathbf{56.88 \pm 1.52}$
LaBraM-Prefix	63.18 ± 10.45	72.06 ± 14.11	55.56 ± 1.94	52.44 ± 3.48
LaBraM-MAM Adapter	64.55 ± 9.87	71.93 ± 14.79	$\underline{57.73 \pm 0.59}$	51.72 ± 2.05
(Ours) LaBraM-TaKF$^+$	$\mathbf{67.04 \pm 14.20}$	$\mathbf{75.46 \pm 12.74}$	56.17 ± 1.45	$\underline{54.27 \pm 1.14}$

(Jing et al., 2023), ContraWR (Yang et al., 2023), CNN-Transformer (Peh et al., 2022), FFCL (Li et al., 2022), and ST-Transformer (Song et al., 2021). Self-supervised modeling methods, which first undergo pre-training to learn semantic representations from EEG data followed by fine-tuning, include BIOT (Yang et al., 2024), EEG2Rep (Foumani et al., 2024), and LaBraM (Jiang et al., 2024). Although EEG2Rep is capable of cross-domain transfer learning, we evaluate them solely on in-domain tasks to ensure consistency in comparison and due to limitations in channel configurations. BIOT and LaBraM are evaluated using the released pre-trained weights. The final category, additive fine-tuning methods, includes Adapter (Houlsby et al., 2019), Prefix-Tuning (Li & Liang, 2021), and MAM Adapter (He et al., 2021). We evaluate these methods using an equal tunable parameter ratio of 3% and the same EEG foundation model with the publicly available parameters. To demonstrate the effectiveness of our method across EEG foundation models, we adopt two EEG foundation models, LaBraM and BIOT. Details about baseline models are described in Appendix A.4.

6 Experimental Results

6.1 Main Results

We present the summary results in Tables 1 and 2. In the analysis, while additive fine-tuning methods entirely depend on the potential of the EEG foundation model, direct comparisons between the proposed and baseline methods, such as supervised and self-supervised modeling methods, are not necessarily fair. Instead, it is important to note that both the proposed and existing additive fine-tuning methods guarantee performance comparable to the EEG foundation model or maintain the EEG foundation model's outperformance compared to other baseline methods. To this end, we primarily compare our proposal with the additive fine-tuning methods. We briefly present the state-of-the-art (SOTA) for both SMMs and Self-SMMs separately in the tables, with detailed results in Appendix B. Bold values represent the best results, while underlined values indicate the second-best, both shown only within the context of additive fine-tuning results without fine-tuning.

Table 2: Performance comparison to the competitors on LeftRight Hand, DREAMER, Crowdsourced, and TUEV datasets, using BIOT as the base EEG foundation model. FT and LP denote fine-tuning and linear probing, respectively, while PT represents Prefix-Tuning. Bold values represent the best results, while underlined values indicate the second-best.

	LeftRight Hand		**DREAMER**	
	BACC	AUROC	BACC	AUROC
SMM SOTA	69.06 ± 5.16	77.75 ± 6.02	56.25 ± 4.10	59.04 ± 7.17
Self-SMM SOTA	60.74 ± 3.51	41.44 ± 5.32	55.67 ± 3.64	59.60 ± 4.79
BIOT-FT (Yang et al., 2024)	49.32 ± 0.70	49.55 ± 0.91	49.04 ± 1.94	48.88 ± 3.13
BIOT-LP	50.11 ± 0.61	50.53 ± 1.09	49.78 ± 0.95	49.78 ± 2.66
BIOT-Adapter (Houlsby et al., 2019)	49.85 ± 0.90	49.81 ± 1.46	50.35 ± 1.73	49.32 ± 2.35
BIOT-PT (Li & Liang, 2021)	49.89 ± 0.43	$\underline{51.53 \pm 0.69}$	50.40 ± 1.31	49.57 ± 1.69
BIOT-MAM Adapter (He et al., 2021)	$\underline{50.15 \pm 0.42}$	$\mathbf{51.81 \pm 1.51}$	50.23 ± 1.38	$\mathbf{51.49 \pm 3.51}$
(Ours) BIOT-TaKF$^+$	$\mathbf{50.49 \pm 0.75}$	50.61 ± 0.87	$\mathbf{50.43 \pm 1.55}$	$\underline{49.92 \pm 1.20}$

	Crowdsourced		**TUEV**	
	BACC	AUROC	BACC	Cohen's κ
SMM SOTA	60.71 ± 11.99	74.99 ± 9.32	43.84 ± 3.49	39.12 ± 2.37
Self-SMM SOTA	62.04 ± 10.51	65.30 ± 11.15	64.09 ± 0.65	66.37 ± 0.93
BIOT-FT	57.71 ± 8.11	69.42 ± 9.39	52.81 ± 2.25	52.73 ± 2.49
BIOT-LP	$\underline{61.87 \pm 8.16}$	68.62 ± 9.17	37.47 ± 1.25	46.66 ± 2.48
BIOT-Adapter	58.16 ± 8.24	65.44 ± 5.70	45.54 ± 2.77	$\underline{51.12 \pm 4.33}$
BIOT-Prefix	59.98 ± 7.56	$\mathbf{72.55 \pm 8.30}$	36.01 ± 1.45	35.09 ± 2.67
BIOT-MAM Adapter	58.36 ± 7.76	70.08 ± 6.90	$\mathbf{48.49 \pm 1.83}$	47.58 ± 3.46
(Ours) BIOT-TaKF$^+$	$\mathbf{63.83 \pm 7.06}$	$\underline{70.44 \pm 4.70}$	$\underline{46.66 \pm 1.22}$	$\mathbf{51.56 \pm 2.03}$

LaBraM Case. Table 1 presents the results of using LaBraM as the EEG foundation model for applying additive fine-tuning methods. The results show that TaKF$^+$ outperformed or delivered performance comparable to other methods across most tasks. Especially in the Crowdsourced dataset, we observed that TaKF$^+$ achieved better performance than fine-tuning. It is notable that while other additive fine-tuning methods exhibit high variability depending on the dataset, our method maintained low variability across task types. Although the Adapter performed more stably than other baselines, it did not achieve the versatility of TaKF$^+$. Consequently, TaKF$^+$ is a versatile and effective tuning method when applied to LaBraM.

BIOT Case. Although BIOT is pre-trained on multiple datasets with a pretext task, it is noteworthy that BIOT is smaller, with only 3.3 million parameters, and is pre-trained on fewer datasets, specifically related to sleep and seizure, compared to LaBraM. The results of applying additive fine-tuning methods to BIOT are reported in Table 2. The results show that TaKF$^+$ achieves improved performance than other baselines. While Prefix-Tuning and MAM Adapter show better performance than Adapter due to their ability to enhance the expressiveness of the neural network, TaKF$^+$ surpasses them with a single setting across most datasets. Moreover, while fine-tuning underperforms except on non-seizure datasets due to the restricted prior knowledge of BIOT, TaKF$^+$ improves performance beyond fine-tuning. In particular, BIOT-TaKF$^+$ surpasses the Self-SMM SOTA, which is the case for LaBraM-FT, in the Crowdsourced dataset. As a result, This result demonstrates that TaKF$^+$ also functions as an effective additive fine-tuning method on BIOT.

Compare Two Cases In both cases, TaKF$^+$ performs well in terms of performance and variability compared to other additive fine-tuning methods. Specifically, in the case of LaBraM, the Adapter shows better stability than other baselines, likely due to LaBraM's potential, as it contains abundant information from diverse datasets. In contrast, the Adapter's ability is completely underutilized for BIOT, which has a narrower prior knowledge base. On the other hand, TaKF$^+$ consistently shows stable performance regardless of the EEG foundation model. Furthermore, the architecture of TaKF$^+$ is more suitable for EEG foundation models than the MAM Adapter. This difference likely arises from using TaKF, which operates in a low-dimensional space, in contrast to the MAM Adapter, which uses Prefix-Tuning to make the neural network more expressive. In conclusion, given the absence of an additive tuning method that can be broadly applied to a wide range of downstream tasks in the BCI field, we believe that TaKF$^+$ could be a versatile and effective solution.

6.2 Ablation on Few-Shot Learning.

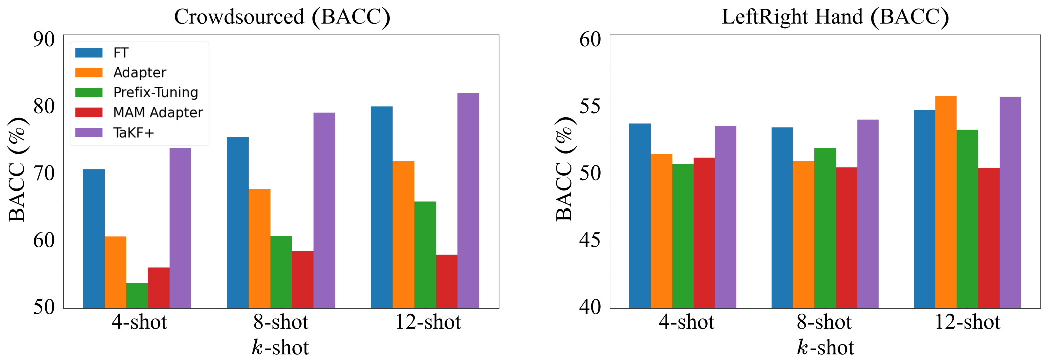

Figure 4: A comparison of the performance between our proposed method and existing additive fine-tuning methods under different few-shot settings (4-shot, 8-shot, and 12-shot). FT in the figure denotes fine-tuning.

While collecting a large amount of labeled EEG data can be expensive and time-consuming, data efficiency is a crucial issue (Ahuja & Sethia, 2024). To verify the data efficiency of each tuning method, we evaluate the performance of our proposed method and existing additive fine-tuning methods under an extremely limited training budget. We compare the methods across three few-shot settings (4-shot, 8-shot, and 12-shot) in a subject-dependent scenario using the Crowdsourced and LeftRight Hand datasets. The detailed numerical results can be found in the Appendix D. The results, illustrated in Figure 4, show that TaKF$^+$ is comparable to or better than other methods, including fine-tuning, in few-shot scenarios. Considering the comparison with the MAM Adapter, extracting task-relevant features in a low-dimensional space through introducing the TaKF module likely plays a key role. As a result, this verifies that TaKF$^+$ is strong in low-data regimes and has high data efficiency.

6.3 Ablation on Additive Tunable Modules.

Table 3: Ablation study of additive tunable modules on LaBraM. Bold values represent the best results, while underlined values indicate the second-best.

	TUEV		**DREAMER**		**Crowdsourced**	
	BACC	Cohen's κ	BACC	AUROC	BACC	AUROC
TaKF$^+$	56.17 ± 1.45	54.27 ± 1.14	54.95 ± 3.84	59.31 ± 4.86	$\mathbf{67.04 \pm 14.20}$	$\mathbf{75.46 \pm 12.74}$
PA(FF)	41.13 ± 1.65	46.00 ± 1.85	52.21 ± 3.44	55.98 ± 4.55	60.51 ± 10.96	71.82 ± 14.53
PA	$\mathbf{59.40 \pm 2.23}$	$\mathbf{56.35 \pm 1.71}$	52.21 ± 3.44	55.98 ± 4.55	58.85 ± 7.95	69.10 ± 8.78
TaKF	52.98 ± 2.03	49.81 ± 0.97	53.38 ± 2.53	59.36 ± 2.49	65.97 ± 10.05	74.96 ± 12.75
TaKF(+FF)	53.21 ± 2.44	51.90 ± 1.72	$\mathbf{56.85 \pm 4.29}$	$\mathbf{60.01 \pm 5.52}$	62.30 ± 8.22	68.76 ± 6.87

We conduct an ablation study to verify the contribution of each additive tunable module. We use an equal tunable parameter ratio of 3%. To evaluate the contribution of the TaKF module, we introduce a new architecture, TaKF(+FF), which adds a feed-forward layer to the cross-attention block to preserve the low-dimensional nature of TaKF with a fixed tunable parameter ratio. PA(FF) is equivalent to the tuning method that excludes a TaKF module from TaKF$^+$. Since having adapter modules attached only to the feed-forward network may negatively impact performance evaluation for comparison, Parallel Adapter (PA) (He et al., 2021), which adds additional adapter modules to the attention layer within a limited parameter budget, is included as a comparison target. Table 3 presents the comparison results on the TUEV, DREAMER, and Crowdsourced datasets. The results demonstrate that each additive tunable module performs well as intended. Specifically, while the PA outperforms in TUEV, the TaKF(+FF) demonstrates outstanding ability in DREAMER. As a result, TaKF$^+$ successfully fuses the advantages of both the adapter-form approach and the TaKF, achieving synergy.

7 Discussion

Limitations. First of all, although we verify the superiority of TaKF$^+$ on LaBraM and BIOT, there remains a need to confirm its adaptability to other EEG foundation models. Secondly, we have provided an explanation and experimental data for the scenarios where the additive fine-tuning methods each perform well, but there is a lack of theoretical analysis to support this explanation. Lastly, there is room for improving synergy considering the Table 3.

Future works. Considering the above limitations, we pave the way for future research directions. 1) Investigate enhancing the synergy or developing architectures tailored to the characteristics of EEG. 2) Develop the pre-training process by adapting additive fine-tuning methods to EEG foundation models that learn from different pretext tasks. Analyzing how prior knowledge is formed based on the pretext task through additive fine-tuning methods may provide key insights into the weaknesses of previous pre-training processes. 3) Discover the anatomical and physiological mechanisms of the brain by utilizing dataset- or subject-relevant key features extracted from TaKF$^+$.

8 Conclusion

We highlight the necessity of exploring methods to apply EEG foundation models to downstream tasks and identify areas for improvement by adapting existing methods from other domains. Based on this analysis, we propose a versatile tuning method, TaKF$^+$, which adapts the EEG foundation model to downstream tasks without altering the pre-trained weights. TaKF$^+$ is a tuning method that can be broadly applied across a wide range of downstream tasks, with a design that includes the TaKF module, which focuses on adaptive feature extraction for tasks, and adapter modules, which primarily leverage the prior knowledge of the EEG foundation model. Through experiments, we demonstrate that TaKF$^+$ achieves performance comparable to state-of-the-art methods for each dataset using a single architecture across most datasets. The proposed TaKF$^+$ presents a breakthrough in efficiently adapting EEG foundation models in a task-agnostic manner. Additionally, the ablation study verifies that TaKF$^+$ performs well in low-data scenarios and has the potential to significantly alleviate the challenges of sample and label efficiency in real-world medical applications, where available data is often limited.

Acknowledgments

References

Chirag Ahuja and Divyashikha Sethia. Harnessing few-shot learning for EEG signal classification: A survey of state-of-the-art techniques and future directions. *Frontiers in Human Neuroscience*, 18:1421922, 2024.

Jimmy Lei Ba. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

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, 2008.

Donghong Cai, Junru Chen, Yang Yang, Teng Liu, and Yafeng Li. Mbrain: A multi-channel self-supervised learning framework for brain signals. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 130–141, 2023.

Alexandra Chronopoulou, Matthew E Peters, Alexander Fraser, and Jesse Dodge. Adaptersoup: Weight averaging to improve generalization of pretrained language models. *arXiv preprint arXiv:2302.07027*, 2023.

Heng Cui, Aiping Liu, Xu Zhang, Xiang Chen, Kongqiao Wang, and Xun Chen. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, 205:106243, 2020.

Mostafa Dehghani, Josip Djolonga, Basil Mustafa, Piotr Padlewski, Jonathan Heek, Justin Gilmer, Andreas Peter Steiner, Mathilde Caron, Robert Geirhos, Ibrahim Alabdulmohsin, et al. Scaling vision transformers to 22 billion parameters. In *International Conference on Machine Learning*, pp. 7480–7512. PMLR, 2023.

Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Muller. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6):993–1002, 2004.

Navid Mohammadi Foumani, Geoffrey Mackellar, Soheila Ghane, Saad Irtza, Nam Nguyen, and Mahsa Salehi. EEG2Rep: Enhancing self-supervised EEG representation through informative masked inputs. *arXiv preprint arXiv:2402.17772*, 2024.

Zeyu Han, Chao Gao, Jinyang Liu, Sai Qian Zhang, et al. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024.

Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. Towards a unified view of parameter-efficient transfer learning. *arXiv preprint arXiv:2110.04366*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *International Conference on Machine Learning*, pp. 2790–2799. PMLR, 2019.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *International Conference on Machine Learning*, pp. 4651–4664. PMLR, 2021.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pp. 709–727. Springer, 2022.

Weibang Jiang, Liming Zhao, and Bao liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024.

Jin Jing, Wendong Ge, Shenda Hong, Marta Bento Fernandes, Zhen Lin, Chaoqi Yang, Sungtae An, Aaron F Struck, Aline Herlopian, Ioannis Karakis, et al. Development of expert-level classification of seizures and rhythmic and periodic patterns during EEG interpretation. *Neurology*, 100 (17):e1750–e1762, 2023.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107, 2017.

Jungye Kim, Seungwoo Jeong, Jaehyun Jeon, and Heung-Il Suk. Unveiling diagnostic potential: EEG microstate representation model for alzheimer's disease and frontotemporal dementia. In *2024 12th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4. IEEE, 2024.

Wonjun Ko, Eunjin Jeon, Jee Seok Yoon, and Heung-Il Suk. Semi-supervised generative and discriminative adversarial learning for motor imagery-based brain-computer interface. *Scientific Reports*, 12(1):4587, 2022.

Wonjun Ko, Seungwoo Jeong, Sa-Kwang Song, and Heung-Il Suk. EEG-oriented self-supervised learning with triple information pathways network. *IEEE Transactions on Cybernetics*, 2024.

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.

Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. Fine-tuning can distort pretrained features and underperform out-of-distribution. *arXiv preprint arXiv:2202.10054*, 2022.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

Hongli Li, Man Ding, Ronghua Zhang, and Chunbo Xiu. Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network. *Biomedical Signal Processing and Control*, 72:103342, 2022.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language–image pre-training with frozen image encoders and large language models. In *International Conference on Machine Learning*, pp. 19730–19742. PMLR, 2023.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*, 2021.

Iyad Obeid and Joseph Picone. The temple university hospital EEG data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

Rajanikant Panda, PS Khobragade, PD Jambhule, SN Jengthe, PR Pal, and TK Gandhi. Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction. In *2010 International Conference on Systems in Medicine and Biology*, pp. 405–408. IEEE, 2010.

Wei Yan Peh, Yuanyuan Yao, and Justin Dauwels. Transformer convolutional neural networks for automated artifact detection in scalp EEG. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 3599–3602. IEEE, 2022.

Aleksandar Petrov, Philip HS Torr, and Adel Bibi. When do prompting and prefix-tuning work? a theory of capabilities and limitations. *arXiv preprint arXiv:2310.19698*, 2023.

Jaeun Phyo, Wonjun Ko, Eunjin Jeon, and Heung-II Suk. Enhancing contextual encoding with stage-confusion and stage-transition estimation for EEG-based sleep staging. In *2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1301–1305. IEEE, 2022.

Jian-Xin Ren, Yu-Jie Xiong, Xi-Jiong Xie, and Yu-Fan Dai. Learning transferable feature representation with swin transformer for object recognition. *Neural Processing Letters*, 55(3):2211–2223, 2023.

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

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. BrainBERT: Self-supervised representation learning for intracranial recordings. *arXiv preprint arXiv:2302.14367*, 2023.

Yaqing Wang, Subhabrata Mukherjee, Xiaodong Liu, Jing Gao, Ahmed Hassan Awadallah, and Jianfeng Gao. Adamix: Mixture-of-adapter for parameter-efficient tuning of large language models. *arXiv preprint arXiv:2205.12410*, 1(2):4, 2022.

Nikolas S Williams, William King, Geoffrey Mackellar, Roshini Randeniya, Alicia McCormick, and Nicholas A Badcock. Crowdsourced EEG experiments: A proof of concept for remote EEG acquisition using emotivpro builder and emotivlabs. *Heliyon*, 9(8), 2023.

Chaoqi Yang, Cao Xiao, M Brandon Westover, Jimeng Sun, et al. Self-supervised electroencephalogram representation learning for automatic sleep staging: Model development and evaluation study. *JMIR AI*, 2(1):e46769, 2023.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36, 2024.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27, 2014.

Elad Ben Zaken, Shauli Ravfogel, and Yoav Goldberg. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language models. *arXiv preprint arXiv:2106.10199*, 2021.

Stanislaw Zakrzewski, Bartlomiej Stasiak, and Adam Wojciechowski. EEG-based left-hand/right-hand/rest motor imagery task classification. In *2022 IEEE 16th International Scientific Conference on Informatics*, pp. 368–372. IEEE, 2022.

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, 2024.

Xing-Zan Zhang, Wei-Long Zheng, and Bao-Liang Lu. EEG-based sleep quality evaluation with deep transfer learning. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part IV 24*, pp. 543–552. Springer, 2017.

Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. *arXiv preprint arXiv:2302.09419*, 2023.