

# DLGE: Dual Local-Global Encoding for Generalizable Cross-BCI-Paradigm

Jingyuan Wang, *Graduate Student Member, IEEE*, Junhua Li, *Senior Member, IEEE*

**Abstract**—Deep learning models have been frequently used to decode a single brain-computer interface (BCI) paradigm based on electroencephalography (EEG). It is challenging to decode multiple BCI paradigms using one model due to diverse barriers, such as different channel configurations and disparate task-related representations. In this study, we propose Dual Local-Global Encoder (DLGE), enabling the classification across different BCI paradigms. To address the heterogeneity in EEG channel configurations across paradigms, we employ an anatomically inspired brain-region partitioning and padding strategy to standardize EEG channel configuration. In the proposed model, the local encoder is designed to learn shared features across BCI paradigms within each brain region based on time-frequency information, which integrates temporal attention on individual channels with spatial attention among channels for each brain region. These shared features are subsequently aggregated in the global encoder to form respective paradigm-specific feature representations. Three BCI paradigms (motor imagery, resting state, and driving fatigue) were used to evaluate the proposed model. The results demonstrate that our model is capable of processing diverse BCI paradigms without retraining and retuning, achieving average macro precision, recall, and F1-score of 60.16%, 59.88%, and 59.56%, respectively. We made an initial attempt to develop a general model for cross-BCI-paradigm classification, avoiding retraining or redevelopment for each paradigm. This study paves the way for the development of an effective but simple model for cross-BCI-paradigm decoding, which might benefit the design of portable devices for universal BCI decoding.

**Index Terms**—Brain–computer interface (BCI), cross-BCI-paradigm, deep learning, electroencephalography (EEG).

## I. INTRODUCTION

RAIN-COMPUTER interface (BCI) has emerged in recent decades as a promising communication system, which enables a direct pathway between the human brain and external devices or environments [11]. Among various neurophysiological signals adopted in BCI, electroencephalography (EEG) is popular with its excellent temporal resolution, low cost, ease of use, and non-invasive advantage [8], [9], [30], [34]. By the use of EEG, different BCI paradigms have been developed for diverse applications, such as motor imagery decoding [1], [14], [17], [19], [40], resting-state monitoring [33], and fatigue detection [12], [13], [21], [26], [35]. Currently, decoding models are usually designed based on a particular BCI paradigm. Even if the same model can be used, it must be retrained when applied to a new BCI paradigm. This is mainly due to two crucial challenges that hinder cross-BCI-paradigm classification. These two challenges are the lack of consistent configurations of EEG channels across BCI paradigms and diverse cognitive or behavioural tasks that result in varying brain activation patterns and EEG characteristics. Both challenges collectively make it difficult to design a structured and generalizable model capable of effective cross-BCI-paradigm classification.

A large body of EEG decoding models relies on convolutional neural networks (CNNs), which are well suited for capturing local

Jingyuan Wang and Junhua Li are with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

Corresponding author: Junhua Li (e-mail: junhua.li@essex.ac.uk).

spatial and temporal features. Pioneering works such as ConvNet [24], EEGNet [15], and TSception [6] demonstrate high performance on a single paradigm. These models operate under the assumption of a fixed number of EEG channels arranged in a predefined order and cannot be directly used for the case of cross-BCI-paradigm, in which channel configurations differ from each other. Extending these models by incorporating attention mechanisms for global modeling, Transformer-based models, including Conformer [29] and Deformer [5] were proposed. While these models achieve excellent performance on a single paradigm, they still employ CNN-based encoders or positional embeddings tied to specific channel configurations, implying that these models require a prior knowledge of channel configurations, making them unable to generalize to new paradigms without retraining. Moreover, to address these limitations in a more flexible manner, graph-based models such as EmT [7] and HNDGST [3] leverage graph convolutional networks (GCNs) to represent inter-channel relationships. However, these models require a fixed adjacency matrix based on the original channel configuration, which again assumes consistent channel configuration across paradigms. Consequently, these models exhibit limited adaptability to the heterogeneous channel configuration that characterize cross-BCI-paradigm.

Besides the challenge of channel configuration heterogeneity, another major limitation of current EEG models is their inability to generalize across BCI paradigms involving different cognitive or behavioural tasks. Although this limitation is acknowledged as a significant challenge, most studies focus on the relatively easier tasks of improving generalization across subjects or sessions within the same BCI paradigm, where the task types and channel configurations remain consistent. For example, several studies aimed to enhance cross-subject generalization in a single paradigm [30]. In an earlier work, Li et al. explored a wide range of linear and nonlinear EEG features to improve cross-subject decoding [18]. Shen et al. proposed a contrastive learning method CLISA for subject-invariant representation learning using convolutional alignment [28]. Liu et al. employed a domain-adaptive multi-branch capsule network DA-CapsNet to improve subject transferability [20]. Similarly, an ensemble of neural networks optimized by evolutionary programming (EPNNE) was introduced to boost cross-subject classification performance [42]. Other studies have tackled cross-session variability, which is caused by intra-subject changes over time due to factors such as tiredness, electrode shift, or ambient change. For instance, the JCSFE [22] utilized a semi-supervised framework to jointly learn session-invariant and session-specific features. To jointly handle cross-subject and cross-session variability, a multi-source domain adaptation network was proposed, which aligns both shared and specific representations across domains [27]. More recently, FMLAN [41] introduced a fine-grained mutual learning adaptive network that simultaneously enhances emotion-discriminative feature learning and domain alignment across both subjects and sessions. While these methods have advanced subject/session-level generalization, they are limited by their assumption of a consistent task type. In practical applications, BCI paradigms differ widely in diverse cognitive or behavioral tasks, such as motor imagery, resting state, or fatigue state, each engaging distinct neural activity [16], [38], [43]. For example, motor imagery paradigm typically involves sensorimotor cortex and

strengthens neural activities in mu and beta frequency bands [39]. In contrast, the driving fatigue paradigm is associated with frontal pole and frontal regions in the lower alpha band (8–10 Hz) [31], [36]. Resting state paradigm exhibits a distinct right hemisphere topography in the theta band [2].

The aforementioned challenges are barriers to the development of cross-BCI-paradigm classification. The existing models are not able to achieve reasonable performance unless they are retrained or redeveloped to adapt to a new paradigm. In order to develop a generalizable classification model fitting different models, the model must be compatible with diverse channel configurations and different tasks across BCI paradigms. To address this issue, it is crucial to understand the major challenges in cross-BCI-paradigm classification, which can be summarized as follows:

- Lack of effective solutions for handling channel heterogeneity cross-BCI-paradigm. Previously proposed models are inherently restricted by architectural constraints, preventing them from processing EEG signals with varying configuration of channels. Even for models that are structurally capable of accepting heterogeneous inputs, channel inconsistency often introduces noise and impairs the model's ability to learn reliable and discriminative representations. Therefore, it is crucial to develop a module that can address inconsistent EEG channel configurations across different paradigms while reducing noise induced by channel heterogeneity.
- Lack of appropriate strategies to address interference between various types of tasks. In scenarios involving multiple types of tasks, the neural activity associated with different task types often occurs in distinct frequency bands. This often results in a model trained to perform well on one task type may fail to generalize effectively to others. Therefore, it is essential to develop mechanisms that alleviate this interference, enabling the model to learn more stable and task-related representations across different types of EEG tasks.

To tackle the challenge posed by inconsistent channel configurations across paradigms, we employ an anatomically inspired brain-region partitioning and padding strategy. By standardizing EEG channels into a predefined anatomical brain-region configuration and padding missing channels accordingly, we transformed different channel configurations of each paradigm into a unified form. This step enables the model to process heterogeneous channel configurations from different paradigms in a unified form. In addition, motivated by the success of spatio-spectral modeling across brain regions in EEG-based classification [10], [37], our proposed model incorporates a dual-level encoder framework to learn more stable and task-related representations. The Local Brain Encoder (LBE) extracts core regional features that are consistent across paradigms. Subsequently, the Global Brain Encoder (GBE) aggregates these shared fundamental regional features into high-level representations that are specific to each task. This dual decoding framework allows the model to learn both shared and task-dependent characteristics across diverse paradigms. The main contributions of this work are summarized as follows:

- We propose an anatomical inspired brain-region partitioning and padding strategy to address inconsistencies in channel configurations across paradigms. Our approach allows the model to handle varied channel configurations across paradigms in a consistent manner, without any dependence on paradigm-specific adjustments or retraining.
- We design a dual-level encoder to mitigate task-specific interference and capture both shared and task-specific patterns. The LBE learns fundamental features within each brain region

using spatial and temporal attention, while the GBE aggregates these features into high-level task-specific representations across paradigms.

- We demonstrate the effectiveness of the proposed model DLGE is effective on the classification across BCI paradigms according to the evaluation results based on three different BCI paradigms.

The remainder of this paper is organized as follows. Section II introduces the BCI paradigms used in this study and presents a comprehensive description of the proposed DLGE model. Section III reports the experimental results. Finally, Section IV concludes the paper and outlines directions for future research.

## II. METHODS

### A. Datasets

To evaluate the feasibility of our model in cross-BCI-paradigm classification, three paradigms are included, which are the motor imagery paradigm [32], the resting state paradigm [33], and the driving fatigue paradigm [21]. A setting comparison of the three BCI paradigms is summarized in Table I.

**1) Motor Imagery Paradigm:** The data of the motor imagery paradigm are from BCI Competition IV-2a dataset [32]. There are four tasks of motor imagery (i.e., left hand, right hand, both feet, and tongue). Nine subjects participated in the experiment, each of whom performed two sessions on separate days. Each participant completed two sessions. Each session consisted of six runs, with short breaks between them. Each run included 48 trials (4 tasks  $\times$  12 trials), resulting in 288 trials per session. EEG signals were recorded using 22 active Ag/AgCl electrodes positioned according to the international 10–20 system, with a sampling rate of 250 Hz.

**2) Resting State Paradigm:** The resting-state paradigm utilizes the Resting-State Dataset [33], which includes under eyes open and eyes closed conditions. EEG signals were recorded using a 64-channel Biosemi ActiveTwo system, with electrodes placed according to the international 10–10 system and a sampling rate of 2048 Hz. Each subject completed two sessions corresponding to the open and closed eyes conditions. Each session lasted 2.5 minutes. Ten healthy individuals participated in the paradigm, including six females and four males, with a mean age of  $30.25 \pm 6.95$  years and an age range from 22 to 45.5 years. In addition, all subjects were right-handed, had normal or corrected normal vision, and had not taken any hypnotic medication during the experiment.

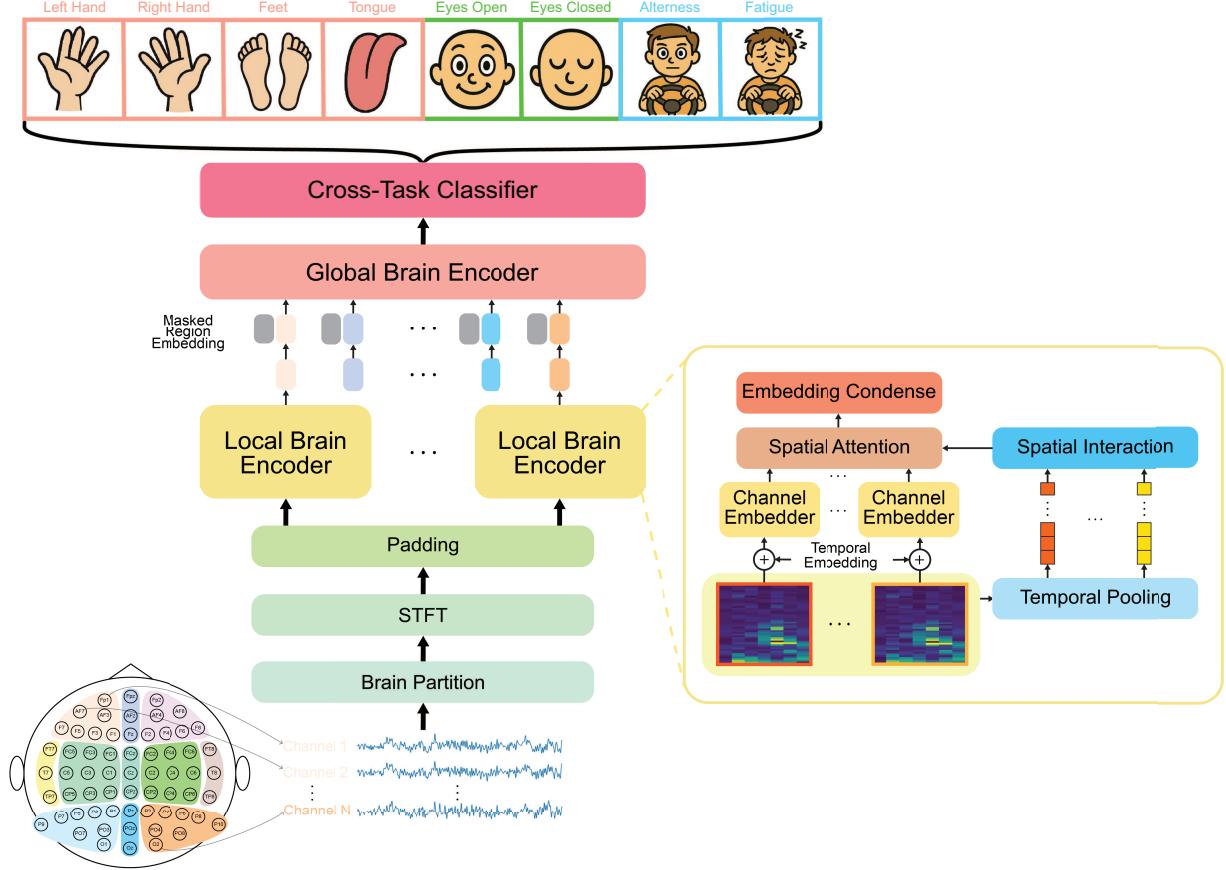
**3) Driving Fatigue Paradigm:** For the driving fatigue paradigm, we employ the Driver Fatigue Dataset [21], which contains EEG signals collected under two mental states: alertness and fatigue. To ensure data quality, all 12 healthy male participants (aged 12 to 24 years) were instructed to maintain adequate sleep and refrain from taking stimulants or medications before the experiment. EEG signals were recorded using a 40-channel Neuroscan amplifier, with 32 active Ag/AgCl electrodes placed according to the international 10–20 system. For each subject, 5-minute EEG segments were collected at a sampling rate of 1000 Hz under both alertness and fatigue conditions.

### B. Data Standardization

All EEG data from the three paradigms are resampled to 250 Hz and band-pass filtered between 0.5–45 Hz. Each trial is segmented into non-overlapping 3-second segments, which serve as individual samples in this study. In the motor imagery paradigm, each subject complete 576 trials (144 per class), yielding a total of 5,184 samples (1296 per class) from 9 subjects across four classes. The resting-state paradigm involve 5 minutes per subject. After segmentation, around 1000 samples (500 per class) are obtained in total. For the driving

TABLE I  
SETTING COMPARISON OF THREE BCI PARADIGMS

Paradigm	#Subjects	#Channels	Sampling Rate (Hz)	#Classes	Class Labels
Motor Imagery Paradigm	9	22	250	4	Left Hand, Right Hand, Feet, Tongue
Resting State Paradigm	10	64	2048	2	Eyes Open, Eyes Closed
Driving Fatigue Paradigm	12	32	1000	2	Alertness, Fatigue



**Fig. 1.** The schematic of the proposed Dual Local-Global Encoder (DLGE) model for cross-BCI-paradigm classification. Each padded brain region is used as the input to the local brain encoder (LBE) that will extract fundamental shared feature within each brain by channel embedder and spatial attention. The input of channel embedder is the time–frequency features of each channel. As for the spatial attention, the input is spatial-frequency features of each brain region. Then the global brain encoder (GBE) is applied to integrate fundamental regional features into task-specific high level representations. After that, the cross-task classifier is employed to obtain classification results.

fatigue paradigm, segmentation of 5-minute sessions under alertness and fatigue states resulted in 1,200 samples per state, totaling 2,400 samples across subjects. Finally, we shuffle all the samples.

### C. Dual Local-Global Encoder

As shown in Fig. 1, our proposed model consists of four main components. The first component addresses the heterogeneity of channel configuration, caused by different settings across BCI paradigms, by introducing an anatomically inspired brain region partitioning and padding strategy. The second component is the LBE, which is designed to capture fundamental regional features shared across paradigms. Both local spatial attention and temporal attention are embedded into the LBE. The third component, named GBE, is designed to integrate fundamental regional features into high-level feature

representations associated with each task. The final component is the cross-task classifier, which performs the classification based on the inputted high-level feature representations.

Let  $\mathbf{X} \in \mathbb{R}^{C \times T}$  denotes a sample with  $C$  channels and  $T$  time points. EEG channels are grouped into 11 sets (corresponding to 11 brain regions,  $R = 11$ ) according to their anatomically spatial locations. These regions are the left frontal region, middle frontal region, right frontal region, left temporal region, left central region, middle motor cortex region, right motor cortex region, right temporal region, left parietal-occipital region, middle parietal-occipital region, and right parietal-occipital region. The regional grouping can be expressed as follows

$$\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(R)}\}, \quad \mathbf{X}^{(r)} \in \mathbb{R}^{C_r \times T}. \quad (1)$$

We subsequently extract time-frequency features from each EEG channel using Short-Time Fourier Transform (STFT) with the sliding time window of 250 data points and an overlap rate of 50%. After STFT, we obtain a feature tensor  $\tilde{\mathbf{X}}^{(r)} \in \mathbb{R}^{C_r \times F \times T'}$ , where  $F$  denotes the number of frequencies and  $T'$  is the number of time points.

In order to unify the different numbers of channels across paradigms, we align the channels within a region  $r$  to a standardized configuration of  $\tilde{C} = 9$  channels. To align with the standardized channel configuration, any missing channels in a paradigm are padded with zeros.

$$\forall r \in \{1, \dots, R\}, \quad \tilde{\mathbf{X}}^{(r)} \in \mathbb{R}^{\tilde{C} \times F \times T'}. \quad (2)$$

Therefore, each  $\tilde{\mathbf{X}}^{(r)}$  contains the time-frequency representations of all valid channels within a specific brain region, with zero padding applied to align with the standardized channel configuration. The paradigms channel configurations, brain region partitioning strategy and the corresponding channel padding scheme are illustrated in Fig. 2, which demonstrates the number of electrodes assigned to each region and the zero-padding strategy.

**1) Local Brain Encoder:** The Local Brain Encoder is designed to extract essential fundamental regional features that were shared across different paradigms within each brain region. It combines local spatial attention with temporal attention mechanisms to enhance its ability to capture shared fundamental local regional representations.

Initially, each channel is fed into the Channel Embedder, which is utilized to extract the representations of all valid (non-padded) EEG channels within each brain region. The Channel Embedder follows a conventional Transformer architecture, but employs single-head self-attention instead of multi-head attention in our implementation. This design preserves the natural frequency structure of EEG signals, as multi-head attention would split the frequency dimension across different heads, potentially breaking the meaningful relationships between frequency bands. Each channel within a region is treated as a sequence of tokens along the temporal axis. Specifically, the time-frequency matrix  $\mathbf{X}_i^{(r)} \in \mathbb{R}^{F \times T'}$  of the  $i$ -th channel in  $r$ -th brain region is transposed and then incorporate temporal positional embeddings to encode temporal information as:

$$\mathbf{Z}_i^{(r)} = \text{PE} + (\tilde{\mathbf{X}}_i^{(r)})^\top \in \mathbb{R}^{T' \times F}, \quad (3)$$

where  $\text{PE} \in \mathbb{R}^{T' \times F}$  is a learnable temporal positional embedding. Channel Embedder processes each  $\mathbf{Z}_i^{(r)}$  to yield the channel-level feature as

$$\mathbf{h}_i^{(r)} = \text{Transformer}(\mathbf{Z}_i^{(r)}), \quad \bar{\mathbf{h}}_i^{(r)} = \frac{1}{T'} \sum_{t=1}^{T'} \mathbf{h}_i^{(r)}[t]. \quad (4)$$

Next, to capture inter-channel relationships within a brain region, we compute pairwise channel correlations. First, we apply temporal pooling on each region as

$$\mathbf{v}^{(r)} = \text{MeanPool}_t(\tilde{\mathbf{X}}^{(r)}) \in \mathbb{R}^{\tilde{C} \times F}. \quad (5)$$

Then, following the attention mechanism, we project the pooled vectors into queries and keys as

$$\mathbf{q}^{(r)} = \mathbf{W}^q \mathbf{v}^{(r)}, \quad \mathbf{k}^{(r)} = \mathbf{W}^k \mathbf{v}^{(r)}, \quad (6)$$

where  $\mathbf{W}^q, \mathbf{W}^k \in \mathbb{R}^{F \times d_{\text{local}}}$ . The attention scores across all channels in the region are computed as

$$\mathbf{a}^{(r)} = \text{softmax} \left( \frac{\mathbf{q}^{(r)} \mathbf{k}^{(r)\top}}{\sqrt{d_{\text{local}}}} \right) \in \mathbb{R}^{\tilde{C} \times \tilde{C}}. \quad (7)$$

This procedure is referred to as Spatial Interaction, which quantifies the channel-wise relevance within a region. Then, the attention scores are used to reweight the channel embeddings as

$$\tilde{\mathbf{h}}^{(r)} = \mathbf{a}^{(r)} \bar{\mathbf{h}}^{(r)} \in \mathbb{R}^{\tilde{C} \times d_{\text{local}}}. \quad (8)$$

This step is referred to as spatial attention. The final representation for the region  $r$  is obtained by applying a feed-forward network (FFN) followed by a channel-wise pooling operation to condense and aggregate the representations of all channels within the region as

$$\mathbf{z}^{(r)} = \frac{1}{C'} \sum_{i=1}^{C'} \text{FFN}(\tilde{\mathbf{h}}_i^{(r)}) \in \mathbb{R}^{d_{\text{local}}}, \quad (9)$$

where  $C'$  is the number of valid channels. This step is referred to as Embedding Condensation.

**2) Global Brain Encoder:** The Global Brain Encoder is designed to integrate fundamental regional features into comprehensive high-level representations tailored for each specific task. For each brain region, we firstly add a learnable masked region embedding to its representation, which encodes the positional information of the region. The resulting region-wise representations are then passed into the Global Brain Encoder, which is implemented as a standard multi-head attention Transformer. The Transformer follows the general formulation as

$$\mathbf{H} = \text{Transformer}(\mathbf{z}) \in \mathbb{R}^{R \times d_{\text{global}}}. \quad (10)$$

To prevent noise introduced by padded zeros in brain regions lacking valid channels during the self-attention computation, the attention scores corresponding to these padded positions are assigned a value of negative infinity before the softmax operation. This ensures that after applying softmax, the attention weights for the padded channels are effectively zeroed out, thereby eliminating their influence on the attention mechanism. The output is aggregated via mean pooling over the token (region) dimension to obtain a high-level representation  $\bar{\mathbf{H}}$  as

$$\bar{\mathbf{H}} = \frac{1}{R'} \sum_{r=1}^{R'} \mathbf{H}[r], \quad (11)$$

where  $R'$  is the number of valid regions.

**3) Cross-Task Classifier:** In the final step, the cross-task classifier consists of two fully-connected layers with ReLU activation in between. It takes the high-level representation as input and outputs the final prediction after *Softmax* function as

$$\hat{\mathbf{y}} = \text{Softmax}(\text{FC}_2(\text{ReLU}(\text{FC}_1(\bar{\mathbf{H}}))). \quad (12)$$

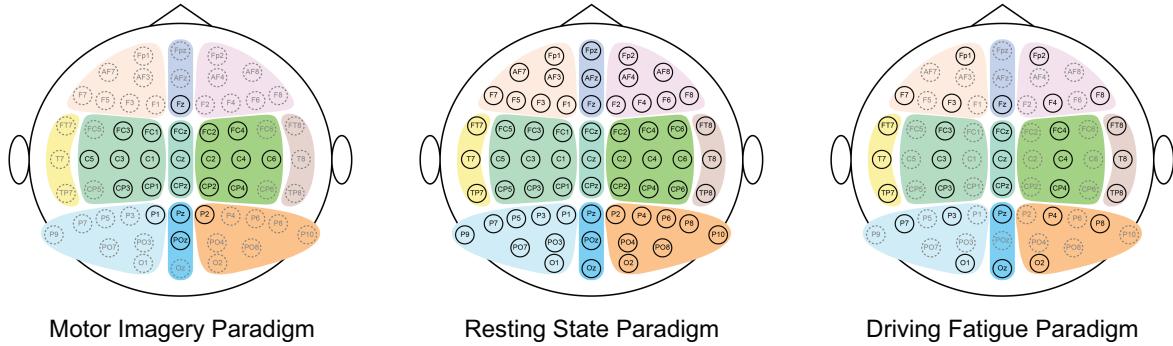
#### D. Evaluation

For reliable performance evaluation, we addressed class imbalance by implementing a class-weighted cross-entropy loss function during training. This approach prevents the model from developing bias toward majority classes. Our experimental design utilized five-fold cross-validation, with the dataset partitioned into five folds maintaining approximately equal class proportion.

To provide a thorough evaluation of the model's performance, we adopted macro precision, macro recall, and macro F1-score obtained via five-fold cross-validation as evaluation metrics. The formulas are given as

$$\text{Macro Precision} = \frac{1}{C_n} \sum_{c=1}^{C_n} \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}, \quad (13)$$

$$\text{Macro Recall} = \frac{1}{C_n} \sum_{c=1}^{C_n} \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \quad (14)$$



**Fig. 2.** Illustration of the recorded electrode layouts and the padded electrodes for each paradigm, as well as the brain region partition in our study. Solid circles represent electrodes actually used for data recording, while dashed circles represent the missing electrodes that are padded in our study. Background colors indicate the partitioned brain regions.

**TABLE II**  
THE RESULTS OF DLGE ON CROSS-BCI-PARADIGM CLASSIFICATION

Fold	Mac. Prec.	Mac. Rec.	Mac. F1
1	58.58%	58.58%	58.39%
2	58.72%	58.89%	58.37%
3	60.12%	59.43%	59.39%
4	61.34%	61.00%	60.68%
5	62.06%	61.49%	60.97%
<b>Mean</b>	<b>60.16%</b>	<b>59.88%</b>	<b>59.56%</b>

Mac. Prec. = Macro Precision; Mac. Rec. = Macro Recall; Mac. F1 = Macro F1-Score.

$$\text{Macro F1-Score} = \frac{1}{C_n} \sum_{c=1}^{C_n} \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \quad (15)$$

where  $C_n$  is the number of classes,  $\text{TP}_c, \text{FP}_c, \text{FN}_c$  denote the number of true positives, false positives, and false negatives for class  $c$ , respectively. This macro averaging scheme ensures that the evaluation reflects the model's ability to perform well across all classes, including those with fewer samples.

### III. RESULTS

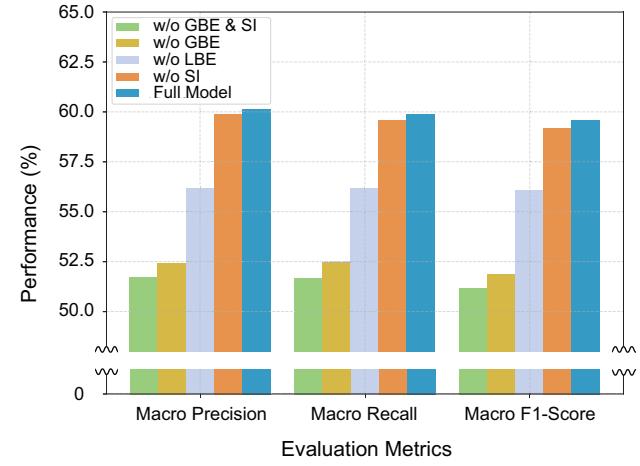
#### A. Classification Performance

To demonstrate the effectiveness of the DLGE, we evaluate our model on a new dataset that we constructed by shuffling all samples from three different paradigms. The macro precision, macro recall, and macro F1-score achieved in each fold are presented in Table II.

The evaluation results demonstrate the feasibility and effectiveness of the proposed DLGE model for cross-BCI-paradigm classification. It achieved an average macro precision of 60.16%, macro recall of 59.88%, and macro F1-score of 59.56% according to the five-fold cross-validation on the data of the three different BCI paradigms. In particular, the relatively small standard deviation between folds (approximately 1.2%) reflects consistent performance across data splits, demonstrating the robustness of the proposed approach. These consistent results across folds suggest that DLGE is feasible to various paradigms, benefiting from its local-global high level task-specific modeling and brain-region-based EEG standardization.

#### B. Ablation Study

To better understand the contribution of each module in DLGE to the overall performance, we conducted an ablation study. Specifically, we conducted ablation studies by comparing four model configurations including DLGE without the GBE, DLGE without both the LBE



**Fig. 3.** Ablation study results of the proposed Dual Local-Global Encoder (DLGE) model across three BCI paradigms. We evaluate the impact of removing key architectural components, which are global brain encoder (GBE), local brain encoder (LBE) and its internal spatial interaction (SI) module, SI module, and both GBE and SI module. The removal of GBE leads to the most significant performance drop across all evaluation metrics, underscoring its critical role in modeling global context. Excluding LBE+SI results in moderate degradation, highlighting the importance of localized feature extraction. The absence of only SI causes marginal impact, suggesting its auxiliary role in enhancing regional interactions.

and its internal Spatial Interaction (SI) module, DLGE without the SI module, and DLGE without both GBE and SI module, as illustrated in Fig. 3.

The Fig. 3 shows that the removal of the GBE has the most significant negative impact on our model performance. Specifically, the macro precision dropped by 7.73%, macro recall by 7.4%, and macro F1-Score by 7.67%, highlighting the critical role of global contextual modeling in integrating task-relevant representations. However, removing LBE (along with SI) leads to a less severe decrease than that observed with the removal of GBE, but it is still notable that macro precision decreased by 3.99%, macro recall by 3.71%, and macro F1-score by 3.48%. This indicates that fundamental regional features shared across paradigms within each brain region also play a crucial role in enhancing the our model's representation capability. Furthermore, removing only the SI module resulted in marginal performance reductions, less than 1% across all three metrics, suggesting that although SI contributes to our model's

generalisation ability, its impact is relatively limited compared to LBE and GBE.

These results demonstrate that the GBE and LBE modules effectively complement each other in capturing hierarchical representations at different levels. The LBE captures fundamental regional features shared across paradigms within each brain region, while the GBE focuses on integrating fundamental regional features into high-level feature representations associated with each task. Together, they significantly enhance our model's generalization ability and overall performance.

### C. Parameter Sensitivity

We explored the impact of important architectural parameters of the proposed DLGE model, which are the number of self-attention layers in both the LBE and GBE and the number of attention heads in GBE's multi-head self-attention. In Transformer-based architectures, the depth of attention layers and the number of heads play a critical role in balancing representational capacity and generalization. To investigate the impact of architectural depth, we varied the number of layers in both LBE and GBE components across the range 2, 4, 6, 8. Additionally, we examined the sensitivity of the multi-head attention mechanism by increasing the number of heads in GBE from 1 to 16, with results presented in Fig. 4.

As illustrated in Fig. 4(a), the increase in the number of attention layers leads to a decline in the classification performance. The best performance is achieved when the number of attention layers are set to two, which outperformed the worst case with 8 layers by approximately 2.17%, 2.18%, and 2.26% in terms of macro precision, macro recall, and macro F1-score, respectively. This suggests that a shallow architecture is sufficient for capturing relevant dependencies in our study, while deeper stacks may lead to overfitting and reduced generalization.

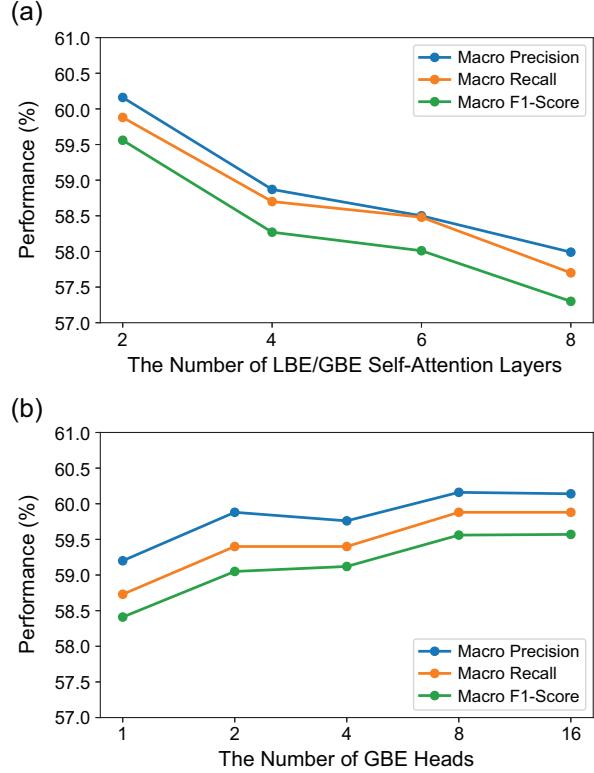
In contrast, the performance change is relatively small when the number of attention heads varied (see Fig. 4(b)). The best choice is eight heads, which surpasses the worst case of 1 head by approximately 0.96%, 1.15%, and 1.15% in terms of macro precision, macro recall, and macro F1-score, respectively. This indicates that while multi-head attention contributes to marginal gain, the use of more heads beyond a fixed number offers limited additional benefit in this context.

### D. Visualization

To investigate how different brain regions contribute to classification performance across paradigms and classes, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) [25] for visualization. This method computes the gradients of the target class score with respect to the input embeddings or intermediate representations, generating class-discriminative localization maps that highlight the most influential brain regions. The resulting visualizations, shown in Fig. 5, illustrate how the model attends to distinct brain regions for different paradigms and classes.

The motor imagery paradigm results in Fig. 5(a) show that the most significant contributions are concentrated in the motor cortex region, particularly in the right motor cortex region, which consistently shows higher importance than the left. The parietal-occipital region contributes significantly less, and all four classes in this paradigm exhibit a highly consistent importance distribution centered on the motor cortex region, in agreement with the shared policy-related areas across tasks.

For the resting state paradigm shown in Fig. 5(b), activity predominantly localizes to the right hemisphere, especially the right motor cortex. The left parietal region shows the lowest contribution across



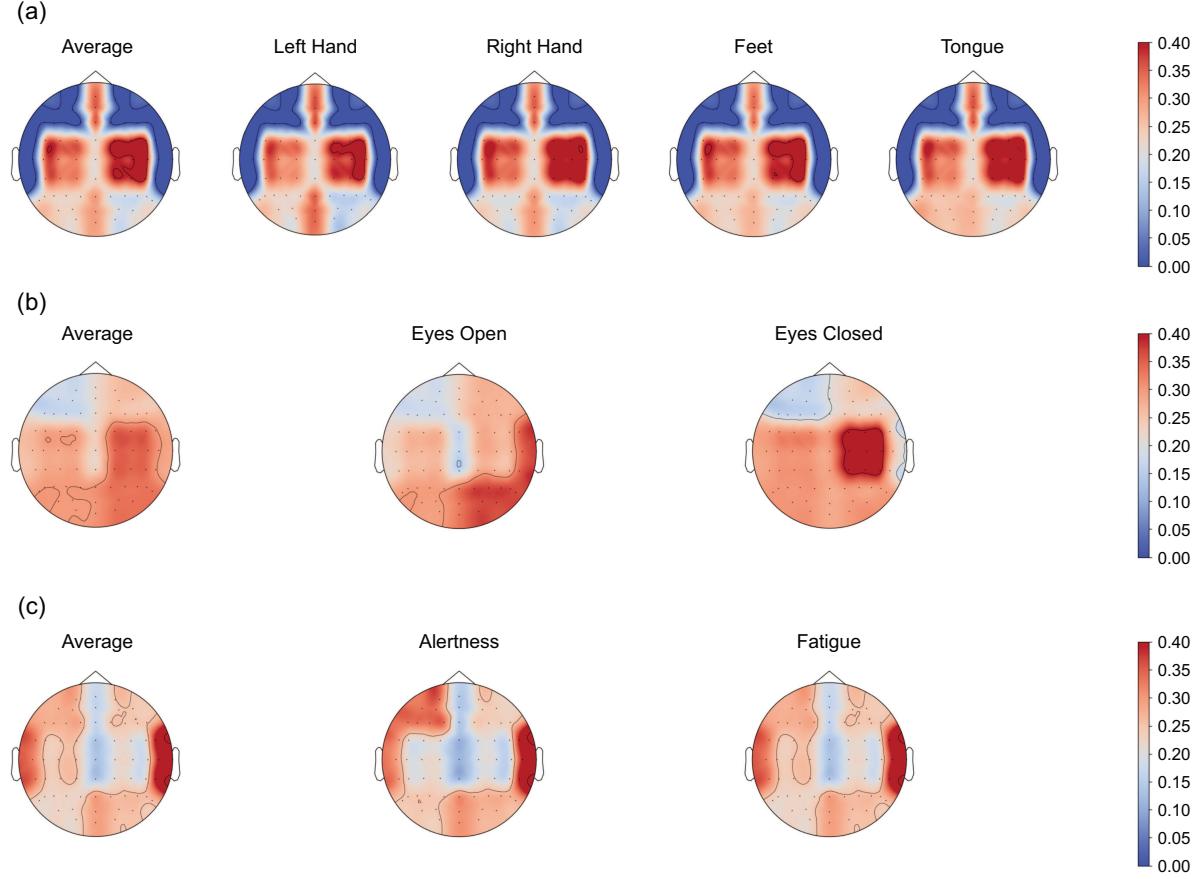
**Fig. 4.** Parameter sensitivity analysis of the proposed Dual Local-Global Encoder (DLGE) model on classification performance. (a) Effect of varying the number of self-attention layers in both the local brain encoder (LBE) and global brain encoder (GBE). The model performs best with two layers, with performance gradually degrading as the number increases, likely due to overfitting. (b) Effect of changing the number of attention heads in GBE's multi-head self-attention. Performance varies marginally, with the best results observed at eight heads, indicating that excessive heads offer limited benefit in this context.

both classes. Specifically, in the “eyes open” class, the right temporal and right parietal-occipital regions show the highest contribution, whereas in the “eyes closed” class, the right motor cortex becomes the dominant region. This pattern may reflect a shift in internal cognitive processing which is with eyes closed, the lack of visual input may increase reliance on motor-related imagination, enhancing motor cortex activity and in contrast, with eyes open, external visual information remains available, reducing motor-related activation and increasing contributions from temporal and parietal-occipital regions associated with relaxation.

Fig. 5(c) presents the performance on the driving fatigue paradigm that the temporal regions, especially the right temporal region, show the strongest contributions. This is followed by the left frontal region, which shows higher importance in the alertness state. In the fatigue state, the left frontal contribution decreases notably.

Overall, the Grad-CAM visualization confirms that our model is able to focus on physiologically relevant brain regions under different tasks. For instance, motor imagery tasks yield high contributions from the motor cortex, while fatigue detection relies more heavily on parietal and occipital regions, aligning with established neuroscience findings [4], [23]. Resting-state conditions show a distinct shift in dominant regions depending on whether the eyes are open or closed.

Furthermore, the integration of dual-level encoder framework to learn more stable and task-related representations. The Grad-CAM visualization confirms that our model is able to focus on physiologi-



**Fig. 5.** Topographical distribution of task-specific brain region contributions based on Gradient-weighted Class Activation Mapping (Grad-CAM). The leftmost column shows the average contribution map across all classes, while the remaining columns are class-specific contribution maps. (a) Motor Imagery Paradigm. Strong activation in the motor cortex across all classes. (b) Resting State Paradigm. “Eyes open” emphasizes right temporal and parietal-occipital regions, “eyes closed” emphasizes right motor cortex. (c) Driving Fatigue Paradigm. Right temporal region dominates and left frontal region is more active in the alertness state.

cally relevant brain regions under different tasks. For instance, motor imagery tasks yield high contributions from the motor cortex, while fatigue detection relies more heavily on parietal and occipital regions, aligning with established neuroscientific findings [4], [23]. Resting-state conditions show a distinct shift in dominant regions depending on whether the eyes are open or closed.

#### IV. CONCLUSIONS

We conducted an initial study to investigate the feasibility of the cross-BCI-paradigm classification. A novel model DLGE is proposed to classify different paradigms without retraining. In this study, we also provide a solution to address two significant challenges (i.e., channel heterogeneity and interference between tasks) to enable cross-BCI-paradigm classification. The proposed anatomical inspired brain-region partitioning and padding strategy significantly enhances model compatibility across paradigms with varying channel configurations. The dual-level encoder architecture enables DLGE to learn stable high-level task-related representations. Through evaluation across three different paradigms, the results demonstrate that DLGE is feasible and effective for cross-BCI-paradigm classification. Ablation and sensitivity analyses confirm the contributions and robustness of each component, while Grad-CAM visualizations reveal task-relevant neural activation patterns. While effective, DLGE faces challenges in maintaining computational efficiency, suggesting future directions for lightweight and adaptive implementations.

#### REFERENCES

- [1] Phairot Autthasan, Rattanaphon Chaisaen, Thapanun Sudhawiyangkul, Phurin Rangpong, Suktipol Kiatthaveepong, Nat Dilokthanakul, Gun Bhakdisongkhram, Huy Phan, Cuntai Guan, and Theerawit Wilairaprasitorn. MIN2Net: End-to-End Multi-Task Learning for Subject-Independent Motor Imagery EEG Classification. *IEEE Transactions on Biomedical Engineering*, 69(6):2105–2118, June 2022.
- [2] Robert J. Barry, Adam R. Clarke, Stuart J. Johnstone, Christopher A. Magee, and Jacqueline A. Rushby. EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765–2773, December 2007. Publisher: Elsevier BV.
- [3] Cheng Cheng, Zikang Yu, Yong Zhang, and Lin Feng. Hybrid Network Using Dynamic Graph Convolution and Temporal Self-Attention for EEG-Based Emotion Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 35(12):18565–18575, December 2024.
- [4] Chin-Teng Lin, Ruei-Cheng Wu, Sheng-Fu Liang, Wen-Hung Chao, Yu-Jie Chen, and Tzzy-Ping Jung. EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(12):2726–2738, December 2005.
- [5] Yi Ding, Yong Li, Hao Sun, Rui Liu, Chengxuan Tong, Chenyu Liu, Xinliang Zhou, and Cuntai Guan. EEG-Deformer: A Dense Convolutional Transformer for Brain-Computer Interfaces. *IEEE Journal of Biomedical and Health Informatics*, 29(3):1909–1918, March 2025.
- [6] Yi Ding, Neethu Robinson, Su Zhang, Qiuahao 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, July 2023.
- [7] Yi Ding, Chengxuan Tong, Shuailei Zhang, Muyun Jiang, Yong Li,

- Kevin Lim Jun Liang, and Cuntai Guan. EmT: A Novel Transformer for Generalized Cross-subject EEG Emotion Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2025. arXiv:2406.18345 [cs].
- [8] Stefan K. Ehrlich, Kat R. Agres, Cuntai Guan, and Gordon Cheng. A closed-loop, music-based brain-computer interface for emotion mediation. *PLOS ONE*, 14(3):e0213516, March 2019.
- [9] Sharlene N. Flesher, John E. Downey, Jeffrey M. Weiss, Christopher L. Hughes, Angelica J. Herrera, Elizabeth C. Tyler-Kabara, Michael L. Boninger, Jennifer L. Collinger, and Robert A. Gaunt. A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science*, 372(6544):831–836, May 2021.
- [10] Sim Kuan Goh, Hussein A. Abbass, Kay Chen Tan, Abdullah Al-Mamun, Nitish Thakor, Anastasios Bezerianos, and Junhua Li. Spatio-Spectral Representation Learning for Electroencephalographic Gait-Pattern Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(9):1858–1867, September 2018. Publisher: Institute of Electrical and Electronics Engineers (IEEE).
- [11] Shu Gong, Kaibo Xing, Andrzej Cichocki, and Junhua Li. Deep Learning in EEG: Advance of the Last Ten-Year Critical Period. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2):348–365, June 2022. Publisher: Institute of Electrical and Electronics Engineers (IEEE).
- [12] Zizheng Guo, Ming Zhou, and Guofa Li. Driver Vigilance State Estimation Based on Multisource Data. *IEEE Sensors Journal*, 23(3):2592–2604, February 2023.
- [13] Jonathan Harvy, Anastasios Bezerianos, and Junhua Li. Reliability of EEG Measures in Driving Fatigue. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2743–2753, 2022. Publisher: Institute of Electrical and Electronics Engineers (IEEE).
- [14] Wei-Yen Hsu and Ya-Wen Cheng. EEG-Channel-Temporal-Spectral-Attention Correlation for Motor Imagery EEG Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1659–1669, 2023.
- [15] 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, October 2018.
- [16] Gang Li, Jie Wang, Wanxiu Xu, Kuijun Wu, Yufeng Liu, Anastasios Bezerianos, and Yu Sun. Self-Regulation Phenomenon Emerged During Prolonged Fatigue Driving: An EEG Connectivity Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:4895–4906, 2023.
- [17] Junhua Li, Jianyi Liang, Qibin Zhao, Jie Li, Kan Hong, and Liqing Zhang. Design of Assistive Wheelchair System Directly Steered by Human Thoughts. *International Journal of Neural Systems*, 23(03):1350013, June 2013. Publisher: World Scientific Pub Co Pte Ltd.
- [18] Xiang Li, Dawei Song, Peng Zhang, Yazhou Zhang, Yuexian Hou, and Bin Hu. Exploring EEG Features in Cross-Subject Emotion Recognition. *Frontiers in Neuroscience*, 12:162, March 2018.
- [19] Guoyang Liu, Rui Zhang, Lan Tian, and Weidong Zhou. Fine-Grained Spatial-Frequency-Time Framework for Motor Imagery Brain-Computer Interface. *IEEE Journal of Biomedical and Health Informatics*, pages 1–13, 2025.
- [20] Shuaiqi Liu, Zeyao Wang, Yanling An, Bing Li, Xinrui Wang, and Yudong Zhang. DA-CapsNet: A multi-branch capsule network based on adversarial domain adaption for cross-subject EEG emotion recognition. *Knowledge-Based Systems*, 283:111137, January 2024.
- [21] Jianliang Min, Ping Wang, and Jianfeng Hu. Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLOS ONE*, 12(12):e0188756, December 2017.
- [22] Yong Peng, Honggang Liu, Junhua Li, Jun Huang, Bao-Liang Lu, and Wanzenge Kong. Cross-Session Emotion Recognition by Joint Label-Common and Label-Specific EEG Features Exploration. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:759–768, 2023.
- [23] Gert Pfurtscheller and Christa Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters*, 239(2-3):65–68, December 1997.
- [24] 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, November 2017.
- [25] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization.
- [26] Mohammad Shahbakhti, Matin Beiramvand, Izabela Rejer, Piotr Augustyniak, Anna Broniec-Wojcik, Michal Wierzchon, and Vaidotas Marozas. Simultaneous Eye Blink Characterization and Elimination From Low-Channel Prefrontal EEG Signals Enhances Driver Drowsiness Detection. *IEEE Journal of Biomedical and Health Informatics*, 26(3):1001–1012, March 2022.
- [27] Qingshan She, Chenqi Zhang, Feng Fang, Yuliang Ma, and Yingchun Zhang. Multisource Associate Domain Adaptation for Cross-Subject and Cross-Session EEG Emotion Recognition. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023.
- [28] Xinhe Shen, Xianggen Liu, Xin Hu, Dan Zhang, and Sen Song. Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition. *IEEE Transactions on Affective Computing*, 14(3):2496–2511, July 2023.
- [29] 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, 2023.
- [30] Wu Sun and Junhua Li. AdaptEEG: A Deep Subdomain Adaptation Network With Class Confusion Loss for Cross-Subject Mental Workload Classification. *IEEE Journal of Biomedical and Health Informatics*, 29(3):1940–1949, March 2025.
- [31] Yu Sun, Julian Lim, Kenneth Kwok, and Anastasios Bezerianos. Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks. *Brain and Cognition*, 85:220–230, March 2014. Publisher: Elsevier BV.
- [32] Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot R. Müller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI Competition IV. *Frontiers in Neuroscience*, 6, 2012.
- [33] Mastaneh Torkamani-Azar, Sumeyra Demir Kanik, Serap Aydin, and Mujdat Cetin. Prediction of Reaction Time and Vigilance Variability From Spatio-Spectral Features of Resting-State EEG in a Long Sustained Attention Task. *IEEE Journal of Biomedical and Health Informatics*, 24(9):2550–2558, September 2020.
- [34] Kaido Värbu, Naveed Muhammad, and Yar Muhammad. Past, Present, and Future of EEG-Based BCI Applications. *Sensors*, 22(9):3331, April 2022.
- [35] Hongtao Wang, Andrei Dragomir, Nida Itrat Abbasi, Junhua Li, Nitish V. Thakor, and Anastasios Bezerianos. A novel real-time driving fatigue detection system based on wireless dry EEG. *Cognitive Neurodynamics*, 12(4):365–376, August 2018.
- [36] Hongtao Wang, Xucheng Liu, Junhua Li, Tao Xu, Anastasios Bezerianos, Yu Sun, and Feng Wan. Driving Fatigue Recognition With Functional Connectivity Based on Phase Synchronization. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3):668–678, September 2021. Publisher: Institute of Electrical and Electronics Engineers (IEEE).
- [37] Tian Wang, Anastasios Bezerianos, Andrzej Cichocki, and Junhua Li. Multikernel Capsule Network for Schizophrenia Identification. *IEEE Transactions on Cybernetics*, 52(6):4741–4750, June 2022.
- [38] Yijun Wang, Bo Hong, Xiaorong Gao, and Shangkai Gao. Phase Synchrony Measurement in Motor Cortex for Classifying Single-trial EEG during Motor Imagery. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 75–78, New York, NY, August 2006. IEEE.
- [39] Hongli Yu, Sidi Ba, Yuxue Guo, Lei Guo, and Guizhi Xu. Effects of Motor Imagery Tasks on Brain Functional Networks Based on EEG Mu/Beta Rhythm. *Brain Sciences*, 12(2):194, January 2022.
- [40] Hyunsoo Yu, Suwhan Baek, Jiwoon Lee, Illsoo Sohn, Bosun Hwang, and Cheolsoo Park. Deep Neural Network-Based Empirical Mode Decomposition for Motor Imagery EEG Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:3647–3656, 2024.
- [41] Peng Yu, Xiaopeng He, Haoyu Li, Haowen Dou, Yeyu Tan, Hao Wu, and Badong Chen. FMLAN: A novel framework for cross-subject and cross-session EEG emotion recognition. *Biomedical Signal Processing and Control*, 100:106912, February 2025.
- [42] Hanzhong Zhang, Tiennu Zuo, Zhiyang Chen, Xin Wang, and Poly Z.H. Sun. Evolutionary Ensemble Learning for EEG-Based Cross-Subject Emotion Recognition. *IEEE Journal of Biomedical and Health Informatics*, 28(7):3872–3881, July 2024.
- [43] Catharina Zich, Stefan Debener, Cornelia Kranczioch, Martin G. Bleichner, Ingmar Gutberlet, and Maarten De Vos. Real-time EEG feedback

during simultaneous EEG–fMRI identifies the cortical signature of motor imagery. *NeuroImage*, 114:438–447, July 2015.