This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/JBHI.2026.3689121, IEEE Journal of Biomedical and Health Informatics

Subject-Independent Deep Learning Framework for Motor Imagery Electroencephalogram Decoding in Neurorehabilitation

Xicheng Lou, Xinwei Li, Hongying Meng, *Senior Member, IEEE*, and Zhangyong Li

***Abstract*— Motor imagery (MI) has emerged as a pivotal paradigm in non-invasive brain-computer interfaces (BCIs) for neurorehabilitation, enabling motor function restoration through mental rehearsal of movements. However, traditional MI electroencephalogram (EEG) classification models face significant challenges due to high inter-subject variability and the expensive requirement of annotated EEG data for each new subject. To tackle these limitations, we introduce a deep learning framework, the Dual-branch Subject-aligned Generalization Network (DSGNet). DSGNet simultaneously extracts temporal and spectral EEG features through dual complementary convolutional branches and incorporates a novel class alignment loss to enforce domain-invariant representation across subjects, enabling generalization to unseen individuals without requiring subject-specific labeled data. We evaluate DSGNet on four public MI-EEG datasets—OpenBMI, BCI Competition IV 2a, SHU Version 5, and BCI Competition IV 2b—under a rigorous leave-one-subject-out cross-validation protocol. Experimental results show that DSGNet achieves the highest accuracy on the three-class and four-class datasets, with improvements of 0.22% and 2.15% over the strongest baselines, respectively, while maintaining comparable performance on the binary-class dataset. These findings highlight the effectiveness of class-structure alignment in developing reliable subject-independent BCI systems for neurorehabilitation. The source code is available at: `https://github.com/xicheng105/DSGNet`.**

***Index Terms*— motor imagery (MI), electroencephalogram (EEG), brain–computer interfaces (BCIs), domain generalization**

Manuscript received XXXX XX, XXXX; revised XXXX XX, XXXX. This work was supported by the National Natural Science Foundation of China (grant number 62171073, 62311530103, and 62576066); Natural Science Foundation of Chongqing, China (grant number CSTB2023NSCQ-LZX0064); Key Project of Science and Technology Research Program of Chongqing Municipal Education Commission (grant number KJZD-K202400602); Chongqing Chuying Project (grant number CY240610); Chongqing Scientific Research Innovation Project for Postgraduate Students (grant number CYB23240); and the Doctoral Training Program of Chongqing University of Posts and Telecommunications (grant number BYJS202317). *(Corresponding author: Zhangyong Li)*

Xicheng Lou is with the School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: louxicheng@ctbu.edu.cn)

Xinwei Li and Zhangyong Li are with the Research Center of Biomedical Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: lixinwei@cqupt.edu.cn; lizy@cqupt.edu.cn)

Hongying Meng is with the Department of Electrical, Brunel University London, UB8 3PH London, U.K. (e-mail: hongying.meng@brunel.ac.uk)

I. Introduction

In recent years, brain–computer interfaces (BCIs) based on electroencephalography (EEG) have shown promising applications in neurorehabilitation by enabling motor-impaired individuals to interact with external devices through mental activities [1]. Among various paradigms, motor imagery (MI) has attracted considerable interest due to its noninvasive nature and its capacity to elicit consistent neural activation in the sensorimotor cortex [2], [3]. Despite these advantages, building subject-independent MI-BCI systems that generalize well across users remains an open challenge [4].

A key difficulty lies in the considerable variability in EEG signals across subjects, arising from anatomical, physiological, and cognitive differences [5]. These variations impair the consistency of neural representations and pose significant challenges to feature extraction models. As a result, deep learning models trained on pooled data often suffer severe performance degradation when applied to unseen subjects [6].

Although recent advances in domain adaptation and domain generalization have aimed to alleviate such cross-subject variability, many of these methods exhibit limitations that hinder their effectiveness in MI-EEG decoding. Specifically, domain adaptation approaches often rely on adversarial training or pseudo-labeling strategies, which can be unstable or error-prone in the presence of noisy or limited target data [7]. Domain generalization methods, while not dependent on target domain access, typically focus on global feature alignment and neglect the modeling of class-level semantics—a crucial factor for maintaining discriminability across domains [8]. Consequently, features from different classes may become entangled, resulting in suboptimal classification performance.

To address these limitations, we propose a deep learning framework named Dual-branch Subject-aligned Generalization Network (DSGNet) for subject-independent MI-EEG classification. DSGNet consists of two complementary convolutional branches that separately capture temporal and spectral characteristics of EEG signals: the temporal branch models sequential dynamics over time, while the spectral branch focuses on simulated frequency-specific patterns through learned convolutional filters. This dual-branch design enables the model to extract richer and more complementary representations for motor imagery decoding.

More importantly, to improve cross-subject generalization,

we introduce a novel class alignment loss that explicitly preserves semantic class structure across subjects. By simultaneously promoting intra-class compactness, inter-class separability, and cross-subject class-center consistency, this loss encourages the learning of feature representations that are both discriminative and domain-invariant. As a result, DSGNet can better mitigate subject-level variability and maintain robust performance on unseen individuals.

Our main contributions are threefold:

1) We introduce DSGNet, a subject-independent MI classification model that unifies temporal–spectral feature extraction with structure-aware alignment objectives.
2) We propose a class alignment loss that preserves semantic structure across subjects, enabling the model to learn more consistent and generalizable representations.
3) We validate our method on four benchmark MI-EEG datasets—BCI Competition IV 2a [9], BCI Competition IV 2b [9], SHU Version 5 [10], and OpenBMI [11]—under a rigorous leave-one-subject-out protocol, achieving superior performance over competitive baselines.

The remainder of this paper is structured as follows: Section II reviews related work on MI-BCI and subject-independent learning. Section III details the proposed DSGNet framework. Section IV presents experimental design and results. Section V discusses the insights and implications of our findings. Section VI concludes the paper.

II. Related Work

A. Motor Imagery EEG Classification

Motor imagery (MI) EEG classification is a fundamental task in noninvasive brain–computer interface (BCI) systems. Traditional machine learning approaches typically involve two stages: feature extraction and classification. Widely used methods include common spatial pattern (CSP) and its extension, filter bank common spatial pattern (FBCSP), which decompose EEG signals to extract discriminative spatial features across frequency bands [12], [13]. These handcrafted features are often combined with shallow classifiers such as linear discriminant analysis (LDA) or support vector machines (SVM) to perform MI decoding [14], [15]. Although these pipelines perform well in subject-dependent scenarios, their generalization to unseen subjects is limited due to inter-subject variability in brain dynamics. Moreover, their performance often heavily relies on handcrafted features and expert-designed preprocessing pipelines, which may not be optimal across different datasets or users [16], [17].

To address these limitations, deep learning models have been introduced to learn discriminative representations directly from raw or minimally processed EEG data. Architectures such as EEGNet [18], ConvNet [19], and EEGNeX [20] extract hierarchical temporal–spatial patterns and have been widely applied to MI classification tasks. More recent models integrate attention mechanisms [21]–[23], residual connections [24], [25], or multi-branch structures [26], [27] to enhance representational power. However, deep models trained on pooled data may still overfit to specific subject distributions and suffer performance drops when applied to unseen individuals [28].

B. Domain Adaptation Methods

Domain adaptation (DA) has become a prominent strategy for improving subject-independent EEG classification by leveraging labeled data from source domains to enhance performance on unlabeled or sparsely labeled target subjects. A typical approach is to align the marginal feature distributions across domains using adversarial learning frameworks. For instance, Zhao *et al*. [29] introduced a deep representation-based domain adaptation (DRDA) model that employs a feature extractor, domain discriminator, and center loss to reduce both domain discrepancy and intraclass variation. Similar adversarial strategies are adopted in MI-CAT [30], which integrates a Transformer-based temporal encoder with a domain classifier to achieve domain-invariant sequence modeling.

Beyond single-source settings, recent efforts have extended to multi-source domain adaptation (MSDA). Methods like DAMSDAF [31] and MSDDAEF [32] aggregate information from multiple source domains to mitigate the negative effects of source heterogeneity. These frameworks often rely on domain-specific classifiers or attention mechanisms to weigh the relevance of each source with respect to the target. For instance, MI-DAGSC [33] incorporates both marginal and conditional distribution alignment, together with semantic consistency constraints, to improve MI decoding across datasets.

Despite their success, most existing DA methods focus primarily on aligning global distributions, without explicitly modeling the semantic class structures. As a result, features from different classes may become indistinguishable after adaptation, especially when class boundaries differ across domains. Moreover, the reliance on adversarial training or target pseudolabeling may lead to instability or error propagation in the presence of noisy target data.

C. Domain Generalization Methods

Unlike DA, which assumes access to unlabeled or partially labeled target data during training, domain generalization (DG) aims to build models that generalize to unseen domains without any exposure to target domain samples. This makes DG especially attractive for real-world MI-BCI applications where collecting new subject data is costly or impractical.

Recent DG methods explore architectural and objective-level strategies to mitigate inter-subject variability. Transformer-based models, such as ST-DG [34], utilize spatial and temporal attention blocks to capture discriminative patterns while incorporating domain generalization via a gradient reversal layer. Other works leverage ensemble-style designs; for example, Song *et al*. [35] proposed a multi-branch classification network with a score and gate mechanism to adaptively integrate features from latent distributions. Some approaches incorporate feature-level regularization: FDCL [36] enforces category-oriented feature decorrelation and consistency across augmented views to promote subject-invariant representations. Contrastive learning is adopted in in SCLDGN [37] to pull together samples of the same class from different domains, aided by domain-specific batch normalization. EEG-DG [38] enhances cross-subject generalization by combining shared and domain-specific

branches with distribution-level constraints that encourage intra-domain compactness and inter-domain separation.

Compared with DA, DG avoids reliance on potentially unstable pseudo-labeling or target adaptation, and instead emphasizes intrinsic robustness through domain-invariant learning objectives and cross-domain consistency constraints.

Despite the progress achieved by existing domain generalization methods, several challenges remain unresolved. Many approaches rely on complex architectures or auxiliary objectives that may increase computational burden or complicate optimization. Others neglect the explicit modeling of class-level semantic structures across domains, which is crucial for maintaining discriminability in cross-subject MI-EEG decoding. These limitations highlight the need for a unified framework that can extract complementary features while aligning class semantics across subjects in a stable and interpretable manner.

III. Method

We propose a novel deep learning framework, DSGNet, to address the challenge of subject-independent motor imagery EEG classification. As illustrated in Fig. 1, the model adopts a dual-branch architecture that captures temporal and spectral representations in parallel, enabling the extraction of rich and complementary features. To enhance generalization across subjects, a class alignment loss is further introduced to encourage the learning of discriminative and domain-invariant representations.

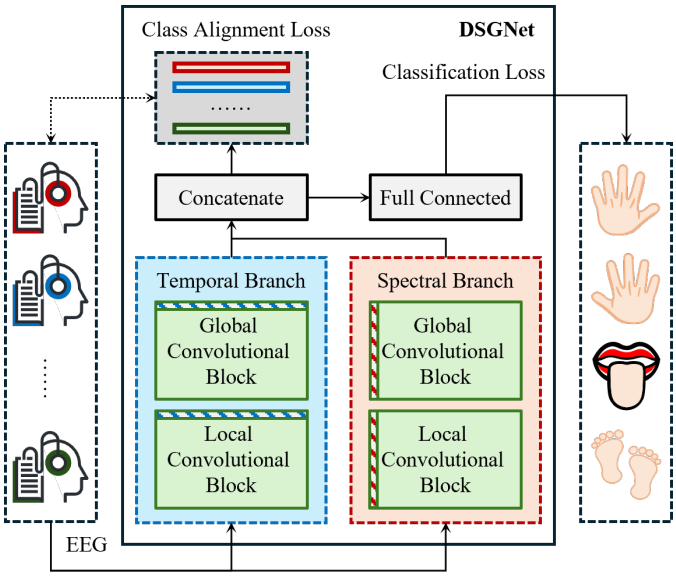


Fig. 1: Overview of the proposed MI-EEG classification framework.

DSGNet is trained on a set of labeled EEG segments $\{\mathbf{X}_i, y_i\}_{i=1}^{m}$, where each input $\mathbf{X}_i \in \mathbb{R}^{1\times E\times T}$ represents a single EEG segment with E electrodes and T timepoints, and y_i denotes the corresponding class label. After training, the model is used to perform classification on unseen test segments.

A. Dataset and Preprocessing

We evaluate our model on four publicly available MI-EEG datasets: the BCI Competition IV 2a (BCI42a) [9], BCI Competition IV 2b (BCI42b) [9], SHU Version 5 (SHUv5) [10], and OpenBMI [11].

The BCI42a dataset consists of four-class MI-EEG recordings (left hand, right hand, feet, and tongue) from 9 subjects, using 22 EEG channels at a sampling rate of 250 Hz. The BCI42b dataset includes two-class MI recordings (left and right hand) from 9 subjects, recorded with 3 bipolar EEG channels (C3, Cz, C4), also sampled at 250 Hz. Following the official protocol, files ending with "T" are used for training and those ending with "E" for validation. Under the leave-one-subject-out (LOSO) evaluation, all T and E files of the held-out subject are merged as the test set.

The SHUv5 dataset was collected from 62 healthy participants performing upper-limb or upper-and-lower-limb MI tasks. In this study, we use only the 11 subjects who participated in the three-class MI paradigm. Each subject contains three sessions; sessions 1–2 are used for training, and session 3 is used for validation, while all three sessions of the held-out subject are used as the test set. SHUv5 was originally recorded at 1000 Hz; we downsample the signals to 250 Hz and apply a 0.5–40 Hz band-pass filter before trial segmentation.

The OpenBMI dataset provides large-scale EEG recordings for ERP, MI, and SSVEP paradigms. We only use the MI portion, which includes 54 subjects, each with two MI sessions recorded at 1000 Hz using 20 EEG channels. Each session contains multiple trials of left- and right-hand motor imagery. Following the LOSO protocol, data from session 1 and session 2 of all subjects except the held-out subject are used for training and validation, and all MI trials from both sessions of the test subject are used as the test set. All OpenBMI EEG signals are downsampled to 250 Hz and filtered using a 0.5–40 Hz band-pass filter.

For all datasets, we extract 4-second EEG segments aligned to each motor imagery trial after preprocessing. A summary of dataset characteristics and preprocessing procedures is provided in Table I.

TABLE I: Summary of the Four MI-EEG Datasets and Preprocessing Protocols

Dataset	Subjects	S.R. (Hz)	Channels	Classes	S.P.
OpenBMI	54	250♮	20	2	LOSO♯
BCI42b	9	250	3	2	LOSO†
SHUv5	11	250♮	58	3	LOSO‡
BCI42a	9	250	22	4	LOSO†

S.R.: Sampling Rate; S.P.: Split Protocol.
♮ Original sampling rate is 1000 Hz and resampled to 250 Hz.
♯ Session 1 and session 2 of all non-held-out subjects are used for training and validation; all MI trials from both sessions of the held-out subject are used as the test set.
† Files ending with "T" are used for training and those ending with "E" for validation; all recordings of the held-out subject are combined as the test set.
‡ Sessions 1–2 for training, session 3 for validation; all three sessions of the held-out subject are used as the test set.

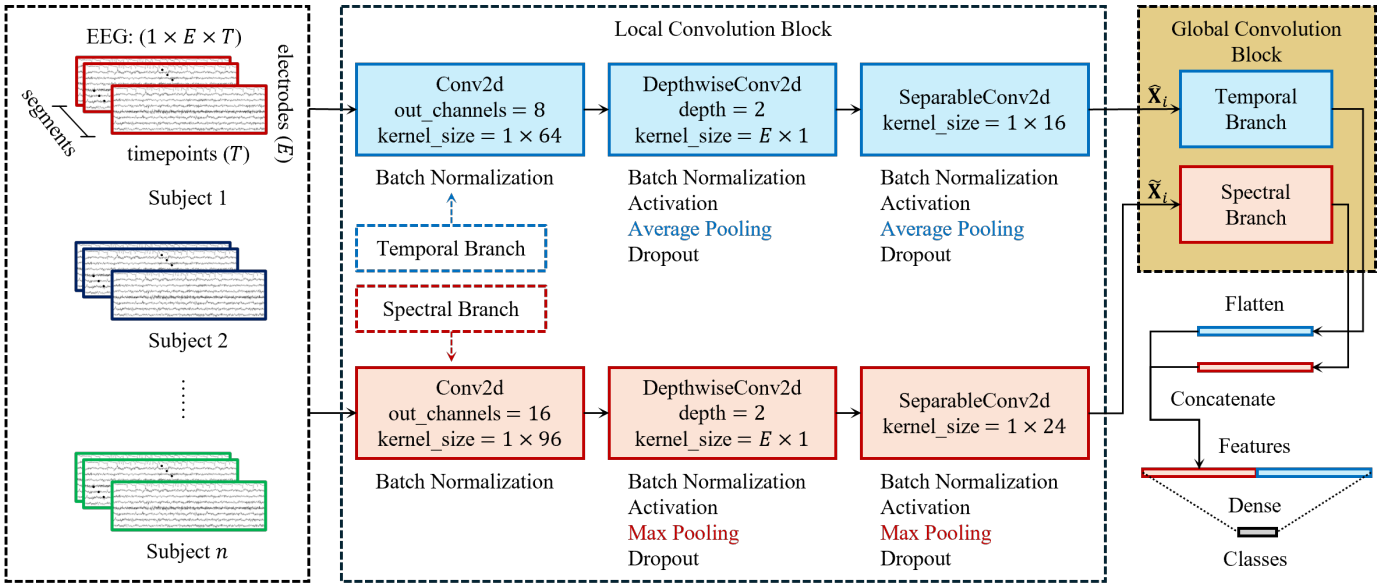


Fig. 2: The architecture of the Local Convolution Block in DSGNet.

B. Architecture of the DSGNet Model

The proposed DSGNet is designed to extract discriminative and generalizable features from motor imagery EEG signals in a subject-independent setting. DSGNet adopts a dual-branch architecture, consisting of a temporal branch and a spectral branch, which respectively focus on capturing temporal dynamics and convolutionally simulated frequency-sensitive patterns. Each branch is composed of two stages: a local convolution block for extracting low-level spatial-temporal or spatial-spectral features, and a global convolution block for modeling higher-order dependencies via attention and causal convolutions. By separating the learning of temporal and spectral representations and enhancing them with global modeling mechanisms, DSGNet enables more expressive and complementary feature extraction, which is crucial for handling the substantial inter-subject variability inherent in EEG data. The outputs of the two branches are concatenated and projected into the class space through a fully connected layer to perform final classification.

1) Local Convolution Block: The local convolution block is designed to extract low-level modality-specific features from EEG signals. In DSGNet, two separate local convolution blocks are employed for the temporal and spectral branches, each tailored to its respective modality. Both branches share a similar structure that consists of three types of convolutional layers applied in sequence: a standard convolution layer to model local receptive fields, a depthwise convolution [39] layer to capture spatial relationships across electrodes, and a separable convolution [39] layer to increase nonlinear modeling capacity with fewer parameters.

The overall structure of the local convolution block is illustrated in Fig. 2, where the temporal and spectral branches are independently constructed but structurally analogous. Specifically, the temporal branch employs an initial convolution with 8 filters and a kernel size of 1×64, while the spectral branch uses 16 filters and a kernel size of 1×96, allowing each branch to focus on the dominant patterns in its corresponding domain. Following the depthwise convolution with a kernel size of $E \times 1$ and a depth of 2, a separable convolution is applied with kernel sizes of 1×16 (temporal) and 1×24 (spectral). Each convolutional layer is followed by batch normalization [40], ELU activation [41], dropout [42], and a pooling layer—average pooling in the temporal branch and max pooling in the spectral branch. This asymmetric pooling strategy reflects the differing statistical characteristics of temporal and spectral EEG features and has been found beneficial for enhancing modality-specific discriminability. We will further analyze the impact of different pooling types on model performance in Section V-B.

The output of each branch preserves essential temporal or spectral resolution, and serves as the input to the corresponding global convolution block for high-level feature modeling. Specifically, the output of the temporal branch is denoted as $\hat{\mathbf{X}}_i \in \mathbb{R}^{\hat{C} \times \hat{T}}$, and that of the spectral branch as $\tilde{\mathbf{X}}_i \in \mathbb{R}^{\tilde{C} \times \tilde{T}}$, where $\hat{C} = 16$ and $\tilde{C} = 32$ are the number of feature channels, and $\hat{T}$ and $\tilde{T}$ represent the sequence lengths of the temporal and spectral features, respectively.

2) Global Convolution Block: The global convolution block is designed to extract high-level features from the modality-specific representations generated by the local convolution block. Given the localized nature of EEG dynamics, directly applying global convolutions may dilute discriminative temporal or spectral patterns. To mitigate this, we adopt a sliding window strategy with stride 1 that segments the feature map into N overlapping subsequences, enabling finer-grained modeling. Taking the temporal branch as an example, the n-th subsequence can be denoted as $\hat{\mathbf{X}}_{n,i} \in \mathbb{R}^{\hat{C} \times \left(\hat{T}-N+1\right)}$, where $n \in \{1, 2, \cdots, N\}$ and $\hat{C}$ is the number of feature channels. These overlapping slices provide diverse and complementary views of the signal, encouraging the network to learn position-aware representations. An overview of the global convolution block structure is illustrated in Fig. 3(a).

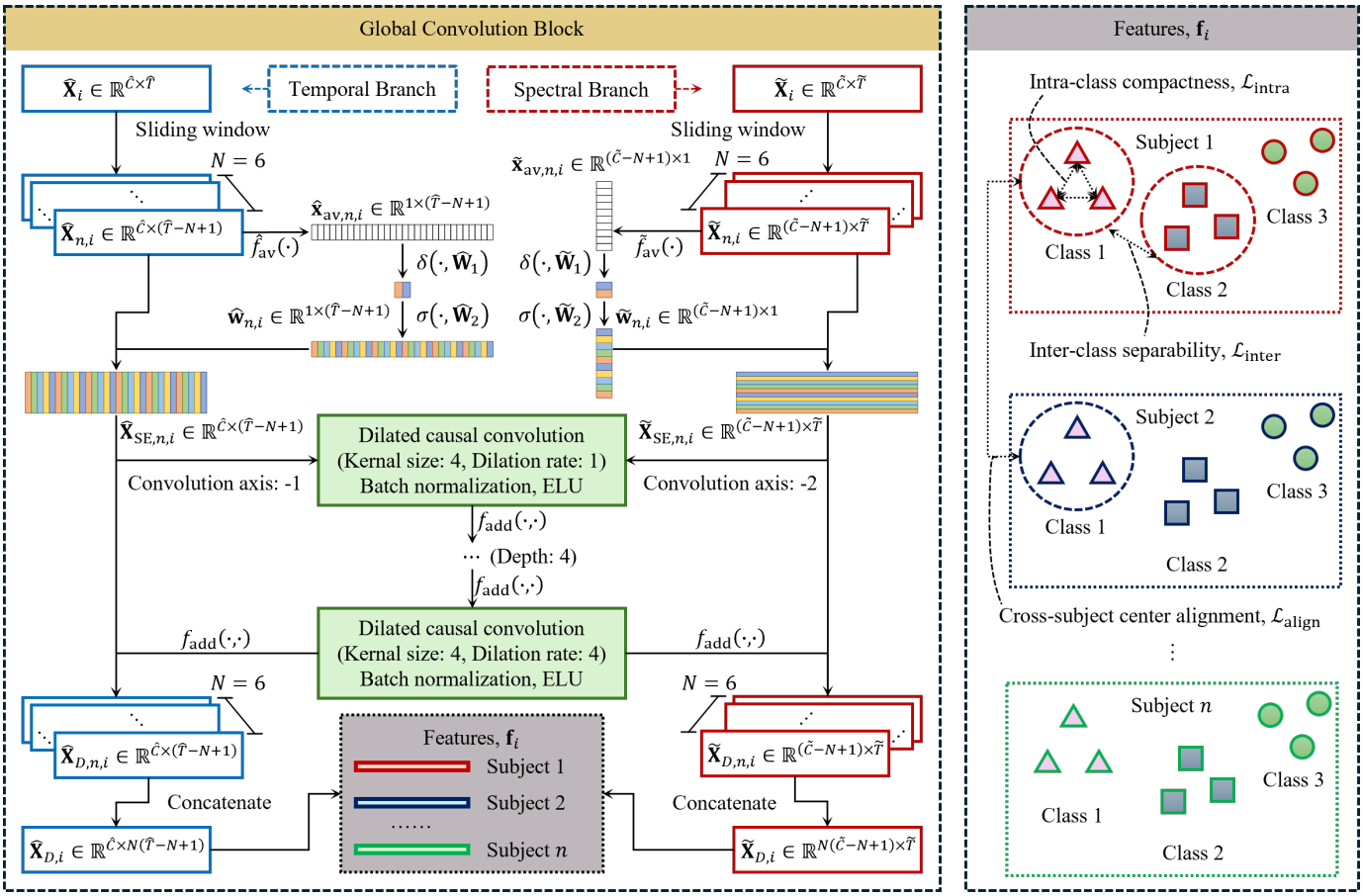


Fig. 3: The architecture of the Global Convolution Block and the associated class alignment loss in DSGNet. (a) Structural overview of the Global Convolution Block in both temporal and spectral branches. (b) Illustration of the proposed class alignment constraints, including intra-class compactness, inter-class separability, and cross-subject class-center consistency.

To process each subsequence individually and emphasize discriminative content, we introduce a dimension-reduced squeeze-and-excitation (SE) [43] module for adaptive feature recalibration. Since not all temporal (or spectral) regions contribute equally to classification, it is necessary to highlight informative patterns while suppressing noise and redundancy. The SE module achieves this by generating a data-driven attention mask that selectively amplifies salient activations within each subsequence.

Concretely, for a given subsequence $\hat{\mathbf{X}}_{n,i}$, we first compute the temporal average across channels:

$$\begin{aligned}\hat{\mathbf{x}}_{\text{av},n,i} &= \hat{f}_{\text{av}}\left(\hat{\mathbf{X}}_{n,i}\right) \\ &= \frac{1}{\hat{C}}\sum_{j=1}^{\hat{C}}\hat{\mathbf{X}}_{n,i}\left(j\right),\end{aligned} \tag{1}$$

producing a descriptor $\hat{\mathbf{x}}_{\text{av},n,i} \in \mathbb{R}^{1\times\left(\hat{T}-N+1\right)}$. This vector is passed through two fully connected layers with ReLU (δ) [44] and Sigmoid (σ) [45] serving as the activation functions:

$$\hat{\mathbf{w}}_{n,i} = \sigma\left(\delta\left(\hat{\mathbf{x}}_{\text{av},n,i}\cdot\hat{\mathbf{W}}_1\right)\cdot\hat{\mathbf{W}}_2\right). \tag{2}$$

This produces a temporal weighting vector $\hat{\mathbf{w}}_{n,i} \in \mathbb{R}^{1\times\left(\hat{T}-N+1\right)}$, with $\hat{\mathbf{W}}_1$ and $\hat{\mathbf{W}}_2$ denoting the weight matrices of the two fully connected layers. Finally, the recalibrated subsequence is obtained via element-wise multiplication:

$$\hat{\mathbf{X}}_{\text{SE},n,i} = \hat{\mathbf{X}}_{n,i}\odot\hat{\mathbf{w}}_{n,i}, \tag{3}$$

where $\odot$ denotes broadcasting along the channel dimension.

This recalibration enables the model to focus on discriminative fragments within each window, which is especially valuable in EEG signals where salient patterns are transient and localized. The SE-enhanced outputs thus provide a more informative input to the subsequent dilated convolution layers for global context modeling.

The recalibrated subsequences are then passed through a stack of d dilated causal convolutional (DCC) layers with increasing dilation rates. Each layer is followed by batch normalization, ELU (ζ) activation, and residual connections. The cumulative receptive field after applying d DCC layers with kernel size k is computed as:

$$R_d = R_{d-1} + \left[\left(d-1\right)\left(k-1\right)+k\right]-1, \tag{4}$$

with $R_0 = 1$, and must satisfy $R_d \geq \hat{T}-N+1$ to cover the entire subsequence.

To mitigate gradient vanishing and facilitate deep feature learning, residual connections are incorporated across the DCC layers. Specifically, the output of each DCC layer is added to the original SE-enhanced subsequence via residual connections before being passed to the next layer:

$$\begin{aligned}\hat{\mathbf{X}}_{\mathrm{DCC},d} &= f_{\mathrm{add}}\left(\hat{\mathbf{X}}_{\mathrm{SE},n,i}, \hat{\mathbf{X}}_{\mathrm{DCC},d-1}\right) \\ &= \zeta\left(\hat{\mathbf{X}}_{\mathrm{SE},N,i} + \hat{\mathbf{X}}_{\mathrm{DCC},d-1}\right),\end{aligned} \tag{5}$$

where $\hat{\mathbf{X}}_{\mathrm{DCC},d}$ is the output of the last DCC layer, After passing through all d DCC layers, the final output of the residual stack is again combined with the original input to form the refined representation for the n-th subsequence:

$$\begin{aligned}\hat{\mathbf{X}}_{D,n,i} &= f_{\mathrm{add}}\left(\hat{\mathbf{X}}_{\mathrm{SE},n,i}, \hat{\mathbf{X}}_{\mathrm{DCC},d}\right) \\ &= \zeta\left(\hat{\mathbf{X}}_{\mathrm{SE},n,i} + \hat{\mathbf{X}}_{\mathrm{DCC},d}\right),\end{aligned} \tag{6}$$

where $\hat{\mathbf{X}}_{D,n,i} \in \mathbb{R}^{\hat{C}\times\left(\hat{T}-N+1\right)}$ denotes the final output of the n-th subsequence after residual refinement. After DCC processing, n output subsequences are concatenated to yield the global representation $\hat{\mathbf{X}}_{D,i} \in \mathbb{R}^{\hat{C}\times N\left(\hat{T}-N+1\right)}$.

For the spectral branch, the input $\tilde{\mathbf{X}}_i \in \mathbb{R}^{\tilde{C}\times\tilde{T}}$ is processed in an analogous manner, with a key difference in the modeling direction. Since this branch focuses on spectral-domain dynamics, we segment the input along the spectral axis using a sliding window. Each n-th subsequence is denoted as $\tilde{\mathbf{X}}_{n,i} \in \mathbb{R}^{\left(\tilde{C}-N+1\right)\times\tilde{T}}$. The subsequent steps, including SE and DCC, are applied to these spectral-domain sequences accordingly.

After global feature extraction, the outputs from the temporal and spectral branches are vectorized into one-dimensional feature vectors: $\hat{\mathbf{x}}_i \in \mathbb{R}^{\hat{C}\cdot N\left(\hat{T}-N+1\right)}$ and $\tilde{\mathbf{x}}_i \in \mathbb{R}^{N\left(\tilde{C}-N+1\right)\cdot\tilde{T}}$. These vectors are then concatenated to form a unified representation for the i-th segment, which is passed through a fully connected layer to produce the class logits:

$$\mathbf{z}_i = \mathbf{W}\left(\hat{\mathbf{x}}_i \frown \tilde{\mathbf{x}}_i\right) + \mathbf{b}, \tag{7}$$

where $\frown$ denotes vector concatenation, and $\mathbf{W}$ and $\mathbf{b}$ are the weight and bias parameters of the fully connected layer. The output $\mathbf{z}_i \in \mathbb{R}^C$ represents the unnormalized logits over C classes for the i-th input segment, which are used for cross-entropy-based training. The predicted class label can be obtained as:

$$\check{y}_i = \arg\max_c z_{i,c}. \tag{8}$$

C. Class Alignment Loss

Cross-subject classification of MI-EEG remains challenging due to inherent variability in neural representations among individuals. These variations, caused by anatomical and cognitive differences, often lead to feature distributions that differ significantly between subjects, thereby hindering the generalization of learned models. To address this issue, it is essential to learn discriminative features that can separate different motor imagery classes effectively, while also ensuring that these features are domain-invariant, i.e., robust to subject-specific differences.

To this end, we propose a class alignment loss that jointly optimizes three objectives: encouraging features within the same class to be compact regardless of subject, i.e., intra-class compactness; pushing features from different classes apart, i.e., inter-class separability; and aligning class centers across subjects to reduce inter-subject variability, i.e., cross-subject class-center consistency. This composite loss guides the model to learn representations that are both semantically meaningful and generalizable, facilitating reliable decoding on unseen subjects, as shown in Fig. 3(b).

We define the final feature vector for the i-th training sample as the concatenation of vectorized outputs from the temporal and spectral branches:

$$\mathbf{f}_i = \hat{\mathbf{x}}_i \frown \tilde{\mathbf{x}}_i. \tag{9}$$

Each sample is associated with a class label $y_i \in \{1, 2, \cdots, C\}$ and a subject identity $s \in \{1, 2, \cdots, S\}$. For each class c and subject s, we define the subject-specific class center as:

$$\mu_c^{(s)} = \frac{1}{|\mathcal{D}_c^{(s)}|}\sum_{i\in\mathcal{D}_c^{(s)}} \mathbf{f}_i, \tag{10}$$

where $\mathcal{D}_c^{(s)}$ is the set of indices of samples belonging to class c from subject s.

We further define the global class center $\bar{\mu}_c$ as the average of subject-specific centers:

$$\bar{\mu}_c = \frac{1}{S_c}\sum_{s=1}^{S_c} \mu_c^{(s)}, \tag{11}$$

where S_c denotes the number of subjects with class c samples.

1) Intra-class Compactness Loss: Intra-class compactness aims to minimize the dispersion of feature representations within each class. Since EEG signals from the same motor imagery class may still vary due to subject-specific factors, we compute a subject-specific class center $\mu_c^{(s)}$ and encourage each sample $\mathbf{f}_i$ from subject s and class c to be close to $\mu_c^{(s)}$. This helps the model reduce within-class variance locally (per subject), laying a foundation for global alignment across subjects.

Formally, the intra-class compactness loss is defined as:

$$\mathcal{L}_{\mathrm{intra}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{f}_i - \mu_{y_i}^{(s_i)}\right\|_2^2, \tag{12}$$

where N is the total number of training samples, y_i and s_i denote the class and subject identity of sample i, respectively.

2) Inter-class Separability Loss: While intra-class compactness promotes the clustering of same-class features, it is equally important to ensure sufficient separability between different classes. To this end, we define an inter-class separability loss that measures the average distance between class centers within each subject. This encourages features from different classes to remain distinguishable at the subject level, improving within-domain discriminability.

The inter-class separability loss is defined as:

$$\mathcal{L}_{\mathrm{inter}} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{C_s\left(C_s-1\right)}\sum_{c_1\neq c_2}\left\|\mu_{c_1}^{(s)} - \mu_{c_2}^{(s)}\right\|_2, \tag{13}$$

where S is the total number of subjects, C_s is the number of classes present in subject s, and $\mu_c^{(s)}$ is the class center of class c, computed from samples of subject s.

3) Cross-subject Center Alignment Loss: Although intra-class compactness is enforced within each subject, the same class may still be represented by different feature distributions across subjects due to individual-specific neural patterns. To address this issue and promote subject-invariant learning, we introduce a cross-subject center alignment loss. This loss encourages the subject-specific class centers to align with a global center, thereby guiding the model to extract common, subject-independent representations for each class. Formally, the cross-subject alignment loss is defined as:

$$\mathcal{L}_{\text{align}} = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{S_c}\sum_{s=1}^{S_c}\left\|\mu_c^{(s)} - \bar{\mu}_c\right\|_2^2. \tag{14}$$

4) Total Class Alignment Loss: The three loss components described above work in a complementary fashion to promote discriminative and subject-invariant feature learning. Specifically, intra-class compactness encourages samples within the same class and subject to cluster tightly, inter-class separability pushes different classes apart within each subject, and cross-subject center alignment minimizes the distributional shift of class features across subjects. By jointly optimizing these objectives, the model is guided to extract features that are both class-discriminative and robust to subject variability.

To balance the opposing behaviors of intra-class compactness and inter-class separability—whose values tend to decrease simultaneously during training—we introduce a weighting factor $\lambda_1 > 0$ to amplify the separability term. The final class alignment loss is formulated as:

$$\mathcal{L}_{\text{CA}} = \mathcal{L}_{\text{intra}} - \lambda_1\mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{align}}, \tag{15}$$

where λ_1 controls the relative emphasis placed on maximizing class separability versus minimizing intra-class variance.

After obtaining the class logits $\mathbf{z}_i$ as (7), we apply the standard cross-entropy loss to supervise the classification task:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\exp\left(z_{i,y_i}\right)}{\sum_{c=1}^{C}\exp\left(z_{i,c}\right)}\right), \tag{16}$$

where z_{i,y_i} is the logit corresponding to the ground-truth class of the i-th sample. To jointly optimize classification and class alignment, we define the total training loss as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda_2\mathcal{L}_{\text{CA}}. \tag{17}$$

This encourages the model not only to correctly classify input segments, but also to learn a feature space that is compact within classes, well-separated between classes, and aligned across subjects. To balance the contributions of the two objectives while maintaining the dominance of the classification signal, we empirically set $\lambda_2 = 0.01$, taking into account both scale compatibility and the importance of preserving classification performance.

IV. Experiments

This section presents a comprehensive evaluation of the proposed DSGNet on four public MI-EEG datasets: OpenBMI, BCI42b, SHUv5, and BCI42a. We aim to assess both the classification performance and the cross-subject generalization capability of our model under a rigorous LOSO validation protocol. To ensure fair comparisons, we benchmark DSGNet against several state-of-the-art baselines and report multiple performance metrics, including classification accuracy (Acc), macro-F1 (F1) and Cohen's Kappa (Kappa).

A. Experimental Configuration

All experiments are implemented using PyTorch [46]. Training and inference are conducted on a workstation equipped with Intel(R) Xeon(R) Gold 6326 CPU (2.90 GHz) and four NVIDIA GeForce RTX 4090 GPUs (24 GB each). The proposed model is trained using the Adam optimizer [47] with a learning rate of 0.0001. For standard classification experiments, a batch size of 128 and 500 training epochs are used. For domain adaptation and generalization methods, which are evaluated in a subject-wise manner, the batch size is set to 32 per subject. The details of dataset splits and preprocessing procedures are described in Section III-A.

To evaluate the classification performance of different methods, we report three commonly used metrics: Acc, F1 and Kappa. Acc measures the proportion of correctly classified samples among all test samples:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{18}$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

The macro-F1 score is calculated by first computing the F1-score for each class individually and then taking the unweighted average over all classes, making it more suitable for evaluating multi-class MI-EEG performance under potential class imbalance. For a given class, F1 is the harmonic mean of precision and recall:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{19}$$

where

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{20}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{21}$$

Cohen's Kappa is introduced to measure the agreement between predictions and ground truth while correcting for chance, thus providing an additional reliability assessment for model predictions:

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}, \tag{22}$$

where p_o denotes the observed agreement and p_e the expected agreement by chance.

To validate the effectiveness of the proposed DSGNet, we compare it with seven representative baseline models. EEGNet [18] is a widely adopted compact convolutional neural network

TABLE II: Classification performance of different methods on four MI-EEG datasets.

	OpenBMI [11]			BCI42b [9]			SHUv5 [10]			BCI42a [9]		
	Acc	F1	Kappa	Acc	F1	Kappa	Acc	F1	Kappa	Acc	F1	Kappa
EEGNet [18]	**0.7558***	0.7493*	**0.5116***	0.6851	0.6819	0.3707	0.6492*	0.6488*	0.4737*	0.5694	0.5591	0.4261
EEGNeX [20]	0.7532*	0.7492*	0.5064*	**0.7069**	0.7037	**0.4155**	0.6488*	0.6469*	0.4731*	0.5804	0.5715	0.4408
EEGInception [27]	0.7034*	0.6978*	0.4069*	0.6702	0.6676	0.3414	0.5809*	0.5778*	0.3713*	0.4156*	0.3945*	0.2210*
ATCNet [24]	0.7553*	**0.7524***	0.5106*	0.7069	**0.7057**	0.4144	0.6834	0.6826	0.5275	0.5818	0.5770	0.4424
EEG-Deformer [23]	0.7175*	0.7101*	0.4349*	0.6485*	0.6461*	0.2973*	0.6529*	0.6503*	0.4793*	0.5233*	0.5156*	0.3643*
MDGEEG [35]	0.7413	0.7363*	0.4827	0.6849	0.6791	0.3721	0.6688	0.6634	0.5031	0.5481*	0.5417*	0.3978*
EEG-DG [38]	0.6663*	0.6610*	0.3327*	0.6826	0.6819	0.3652	0.6509*	0.6488*	0.4763*	0.4995*	0.4912*	0.3327*
DSGNet	0.7350	0.7252	0.4699	0.6878	0.6837	0.3754	**0.6856**	**0.6833**	**0.5284**	**0.6033**	**0.5985**	**0.4710**

Results are reported as the average performance across all subjects in the LOSO evaluation setting.
Bold values indicate the best performance.
* Asterisks indicate statistically significant differences compared to DSGNet ($P < 0.05$, one-tailed paired t-test), regardless of performance direction.

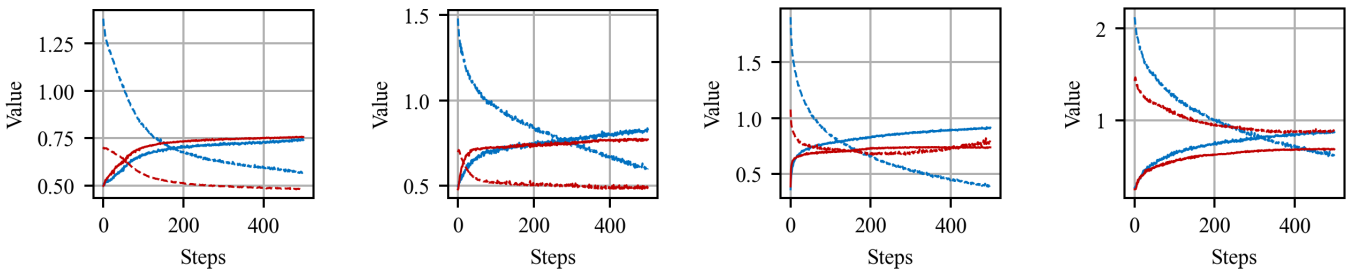


Fig. 4: Training and validation dynamics of DSGNet on four MI-EEG datasets: (a) OpenBMI, (b) BCI42b, (c) SHUv5, and (d) BCI42a. Blue curves indicate the training set, and red curves indicate the validation set. Dashed lines correspond to loss, whereas solid lines correspond to accuracy.

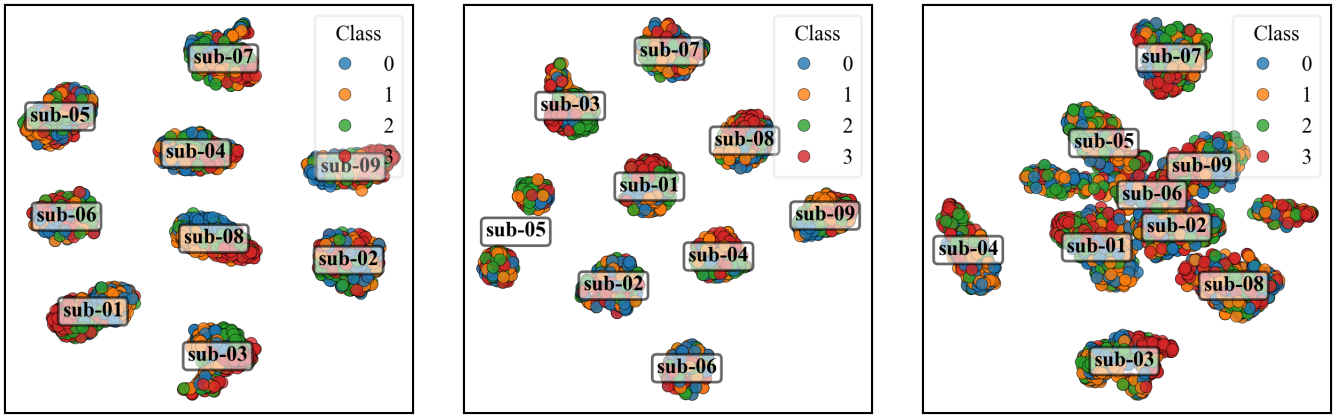
specifically designed for EEG decoding. EEGNeX [20] is an improved version of EEGNet that introduces enhanced depthwise separable convolutions and refined architectural design. EEG-Inception [27] employs a multi-scale inception architecture to capture both short- and long-term temporal dependencies in EEG signals, showing strong performance in MI classification tasks. ATCNet [24] achieves state-of-the-art classification accuracy on several MI-EEG datasets under subject-dependent settings by leveraging attention mechanisms and temporal convolutional blocks. EEG-Deformer [23] incorporates a Transformer-based deformable attention mechanism to better model the temporal dynamics of EEG signals. Additionally, we include two recent domain generalization approaches: EEGDG [38], which employs adversarial learning and class-conditional feature alignment, and MDGEEG [35], a multi-branch generalization framework that explicitly models inter-subject variability. These baselines span compact CNNs, attention-based architectures, and domain generalization paradigms, offering a comprehensive comparison for evaluating the performance and cross-subject robustness of DSGNet.

B. Performance Comparison

Table II reports the classification performance of all compared methods on the four MI-EEG datasets, while Table III summarizes their computational complexity in terms of the number of trainable parameters and FLOPs. As shown in Table II, DSGNet achieves the best performance on the SHUv5 (three-class) and BCI42a (four-class) datasets in terms of accuracy, macro-F1, and Cohen's Kappa, demonstrating its strong capability in handling more challenging multi-class classification scenarios. On the binary-class BCI42b and OpenBMI datasets, DSGNet also maintains competitive performance, suggesting that its advantage becomes more evident as the classification task becomes more complex.

Domain generalization methods such as MDGEEG and EEGDG, which are designed to enhance robustness against distribution shifts, do not consistently outperform conventional deep networks across datasets. In particular, EEGDG shows limited performance on the BCI42a dataset, possibly because its feature-alignment strategy was originally proposed for cross-dataset adaptation and may not fully address the finer-grained subject-level variability involved in our experiments.

Table III further provides a comparison of computational complexity on the BCI42b dataset. DSGNet contains 6.764×10^4 trainable parameters, which is substantially smaller than recent multi-branch or attention-enhanced architectures such as EEG-Deformer (1.681×10^6) and EEGDG (1.274×10^7). Although its FLOPs are moderately higher due to the global convolutional modules, the overall model size remains lightweight. From a practical deployment perspective, real-time feedback in neurorehabilitation systems is mainly determined by inference-time latency, and the relatively higher FLOPs of DSGNet do not constitute a bottleneck under the several-second sliding-window setting commonly used in motor imagery EEG decoding. In addition, as illustrated in



(a) (b) (c)

Fig. 5: t-SNE visualization of the learned feature distributions under different loss configurations on the BCI42a dataset. All visualizations are generated using the complete network architecture. (a) With full loss: $\mathcal{L}_{\text{intra}} - \lambda_1 \mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{align}}$. (b) Without inter-class separability loss: $\mathcal{L}_{\text{intra}} + \mathcal{L}_{\text{align}}$. (c) Without intra-class compactness loss: $-\lambda_1 \mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{align}}$. Feature embeddings are colored by class label, and subject-level clusters are annotated.

Fig. 4 the training and validation curves across all datasets exhibit smooth loss reduction and steadily increasing accuracy, further indicating stable optimization without overfitting.

TABLE III: Computational complexity comparison of different methods on the BCI42b datasets.

	Parameters †	FLOPs‡	TT (h)
EEGNet [18]	4.168×10^3	2.298×10^7	3.560
EEGNeX [20]	8.443×10^4	2.096×10^8	4.168
EEGInception [27]	2.544×10^6	2.544×10^9	3.860
ATCNet [24]	1.452×10^5	2.741×10^7	6.302
EEG-Deformer [23]	1.681×10^6	2.787×10^8	8.683
MDGEEG [35]	1.047×10^5	1.809×10^8	3.271
EEG-DG [38]	1.274×10^7	2.165×10^8	14.186
DSGNet	6.764×10^4	3.832×10^8	3.906

TT: Training time.
† Parameters are the number of trainable parameters.
‡ FLOPs are approximated using Multiply–Accumulate operations (MACs).

TABLE IV: Ablation study evaluating the effect of different loss components and architectural modules on classification performance on the BCI42a dataset.

Loss Components			Model Blocks				BCI42a		
$\mathcal{L}_{\text{intra}}$	$\mathcal{L}_{\text{inter}}$	$\mathcal{L}_{\text{align}}$	SE	GC	SB	TB	Acc	F1	Kappa
-	-	-	√	√	√	√	0.5848	0.5722	0.4463
-	-	√	√	√	√	√	0.5783	0.5686	0.4393
-	√	√	√	√	√	√	0.5671	0.5626	0.4384
√	-	√	√	√	√	√	0.5824	0.5726	0.4572
√	√	√	-	√	√	√	0.5890	0.5824	0.4597
√	√	√	-	-	√	√	0.5744	0.5683	0.4488
√	√	√	√	√	-	√	0.5543	0.5478	0.4056
√	√	√	√	√	√	-	0.5593	0.5522	0.4127
√	√	√	√	√	√	√	**0.6033**	**0.5979**	**0.4710**

Bold values indicate the best performance.
Results are reported as the average performance across all subjects in the LOSO evaluation setting.
SE: Squeeze-and-excitation block; GC: Global convolution block; SB: Spectral branch; TB: Temporal branch.

C. Ablation Study

To investigate the contribution of each loss component and architectural module to the model's performance, we conducted a systematic ablation study on the BCI42a dataset, with results summarized in Table IV.

We first examine the effect of the loss components. When all three auxiliary terms—$\mathcal{L}_{\text{intra}}$, $\mathcal{L}_{\text{inter}}$, and $\mathcal{L}_{\text{align}}$—are removed, the model is trained solely with the standard classification objective. This configuration leads to notably lower performance, indicating that cross-entropy alone is insufficient for producing discriminative and stable representations under cross-subject variability. Removing either $\mathcal{L}_{\text{intra}}$ or $\mathcal{L}_{\text{inter}}$ also causes a clear performance drop, demonstrating that both compactness and separability are essential for maintaining well-structured latent features. When only $\mathcal{L}_{\text{align}}$ is retained while the other two are disabled, the performance remains suboptimal, suggesting that global alignment without local structural constraints is inadequate for effective feature organization. These findings are further supported by the t-SNE [48] visualizations in Fig. 5, which show that stronger structural supervision makes the shape and internal class organization of each subject's cluster more consistent. When the structural constraints weaken, these clusters become increasingly irregular, distorted, or internally mixed, reflecting a loss of coherent class structure.

We then assess the contribution of key architectural elements. Removing the SE block yields a moderate decline in performance, as eliminating adaptive recalibration reduces the model's ability to emphasize informative temporal–spectral regions. Disabling the global convolution block leads to a more substantial degradation, underscoring its importance in capturing long-range dependencies beyond local receptive fields. The results of the single-branch variants further highlight the advantage of the proposed dual-branch design. Both the

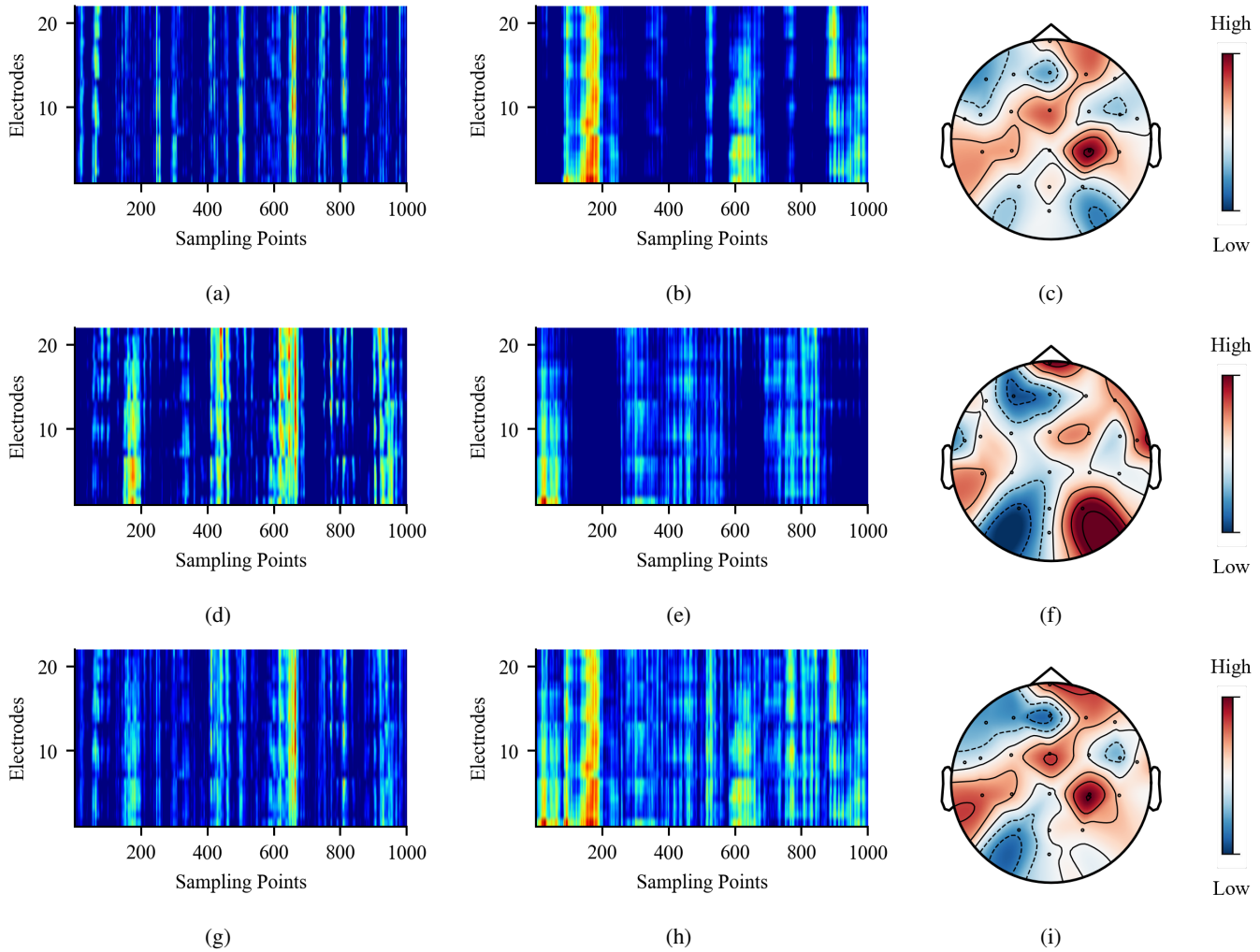


Fig. 6: Grad-CAM visualizations and corresponding topographic maps for representative samples of feet and right-hand motor imagery in the BCI42a dataset. (a) Temporal-branch CAM for feet MI; (b) Temporal-branch CAM for right-hand MI; (c) Topographic map for temporal-branch; (d) Spectral-branch CAM for feet MI; (e) Spectral-branch CAM for right-hand MI; (f) Topographic map for spectral-branch; (g) Fused CAM for feet MI; (h) Fused CAM for right-hand MI; (i) Fused topographic map.

temporal-only and spectral-only networks show considerable decreases in all three metrics compared with the full model, confirming that temporal dynamics and spectral characteristics provide complementary information. Their integration is therefore essential for achieving robust and generalizable cross-subject classification.

V. Discussion

This section presents an in-depth analysis of the proposed model's structural and functional design, with the aim of understanding the underlying mechanisms that drive its performance. We organize the discussion into two major components. First, we investigate the effectiveness of the dual-branch architecture by visualizing the spatial-temporal attention maps generated by Grad-CAM [49], revealing the complementary nature of the temporal and spectral pathways. Second, we analyze the influence of key hyperparameters on classification accuracy, including temporal segment length, pooling strategy, and feature aggregation schemes. These insights provide both theoretical and empirical justification for the architectural choices and training configurations adopted in this study.

A. Structural Insights from the Dual-Branch Design

The dual-branch architecture serves as the foundation of our model, designed to extract and integrate complementary features from the EEG signal's temporal and spectral dimensions. To gain insight into the internal behavior of each branch, we employ Grad-CAM to visualize the attention distribution over input segments, as illustrated in Fig. 6.

As shown in Fig. 6(a)–(b), the temporal branch exhibits sustained and distributed activation patterns across a wide range of sampling points and channels, indicating its capacity to model temporally evolving dynamics. In contrast, the spectral branch (Fig. 6(d)–(e)) presents sparse and sharply localized activations, suggesting a focus on transient and frequency-specific responses.

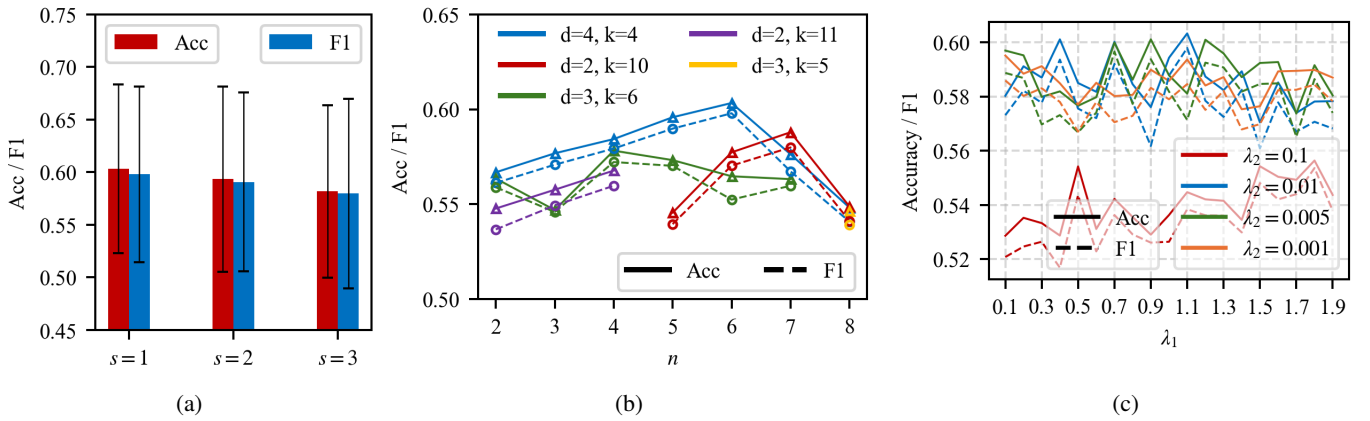


Fig. 7: Performance sensitivity analysis with respect to key hyperparameters. (a) Effects of stride s between sliding windows. (b) Impact of the number of sliding windows n, the depth d, and the kernel size k of dilated convolutional blocks. (c) Sensitivity of model performance to λ_1 and λ_2 on the BCI42a dataset.

The fused attention maps (Fig. 6(g)–(h)) integrate these distinct patterns, resulting in a more spatially balanced and robust representation. As shown in Fig. 6(g), the fused CAM for feet MI exhibits vertically aligned and spatially consistent activation around electrodes Cz, FCz, and CPz (indices 10, 4, and 16), consistent with the somatotopic representation of lower-limb movement in the midline sensorimotor cortex. For right-hand MI (Fig. 6(h)), the activation focuses on electrodes C3, C1, FC3, and CP3 (indices 8, 9, 2, and 14), corresponding to the contralateral left-hemisphere motor cortex.

In addition to the Grad-CAM analysis, the topographic maps in Fig. 6(c), (f), and (i) visualize the spatial distribution of the learned convolutional spatial filters. As shown in the fused topographic map (Fig. 6(i)), the spatial weights are predominantly concentrated over the frontal–central regions. Motor imagery is known to modulate μ/β rhythms primarily over the premotor and sensorimotor cortices, with frontal areas contributing to movement planning and central regions reflecting sensorimotor activation. The consistency between these spatial weight distributions and known cortical organization further supports the interpretability and neurophysiological plausibility of DSGNet.

B. Impact of Pooling Strategies on Performance

We explore the impact of different pooling strategies on the performance of the DSGNet model. The results, presented in Table V, show that when both branches use average pooling, the model achieves an accuracy of 0.5879 and a macro-F1 of 0.5762. When average pooling is applied to the temporal branch and max pooling to the spectral branch, the model performs best, with an accuracy of 0.6033 and a macro-F1 of 0.5979. This suggests that max pooling in the spectral branch effectively selects the most prominent features, improving the model's ability to distinguish between motor imagery classes.

When the temporal branch uses max pooling, the model's performance drops significantly, with an accuracy of 0.5542 and a macro-F1 of 0.5469. This result indicates that max pooling is less suitable for the temporal branch, as it may discard important temporal dynamics of the EEG signals, which are more subtle and require averaging for better feature preservation.

TABLE V: Performance comparison of DSGNet under different pooling strategies on the BCI42a dataset.

Temporal branch pooling type Spectral branch pooling type	Average Average	Average Max	Max Max	Max Average
Accuracy	0.5879	**0.6033**	0.5542	0.5788
macro-F1	0.5762	**0.5979**	0.5469	0.5692
Kappa	0.4476	**0.4710**	0.4171	0.4295

Bold values indicate the best performance.

These findings suggest that max pooling is more effective in the spectral branch, while average pooling is better suited for capturing temporal features.

C. Impact of Key Hyperparameters

To better understand the performance dynamics of the proposed DSGNet, we systematically investigate the impact of several key hyperparameters that govern both the structural composition and the training behavior of the model. These hyperparameters include the stride (s) between sliding windows, the number of sliding windows (n), the depth of dilated convolutional blocks (d), the kernel size of dilated convolutional blocks (k), and the weighting factor λ_1 used in the class alignment loss.

The stride s controls the overlap between adjacent temporal windows. As shown in Fig. 7(a), setting $s = 1$ yields the best performance, while larger strides lead to reduced accuracy and F1 scores. This suggests that overlapping windows help preserve temporal continuity, enabling the model to capture more stable and informative patterns. In contrast, larger strides may introduce discontinuities, weakening temporal feature extraction.

As shown in Fig. 7(b), performance improves steadily as n increases, peaking at $n = 6$. This suggests that dividing the feature map into moderately overlapping segments enhances local pattern extraction without over-fragmenting the input. Depth $d = 4$ and kernel size $k = 4$ consistently yield better

results across different window settings, indicating that they offer a good trade-off between model capacity and generalization. Larger values may lead to overfitting or unstable training, while smaller ones may under-represent key temporal or spectral patterns.

To further validate the robustness of the proposed weighting strategy, we performed a joint sensitivity analysis of λ_1 and λ_2 on the BCI42a dataset. As shown in Fig. 7(c), the value of λ_2—which controls the overall contribution of the class alignment loss relative to the classification loss—has a decisive influence on model performance. When $\lambda_2 = 0.01$, the magnitude of $\mathcal{L}_{\mathrm{CA}}$ becomes comparable to that of $\mathcal{L}_{\mathrm{cls}}$, and this setting achieves the overall best performance, with its peak accuracy and F1-score obtained at $\lambda_1 = 1.1$. A smaller value ($\lambda_2 = 0.005$) leads to moderately reduced performance. In contrast, $\lambda_2 = 0.1$ or $\lambda_2 = 0.001$ significantly degrades performance, as the imbalance in scale either overwhelms or underweights the structural alignment constraints.

Despite the promising performance demonstrated by DSGNet across multiple MI-EEG benchmarks, several limitations should be acknowledged. The proposed framework incorporates multiple architectural components and hyperparameters, which may increase implementation complexity and require careful tuning to achieve optimal performance across different datasets. While these design choices are motivated by the need to capture complementary temporal–spectral characteristics and enforce semantic alignment, they also suggest opportunities for further simplification and optimization. In future work, we plan to investigate lightweight variants of DSGNet to reduce model complexity, as well as explore self-supervised or few-shot learning strategies to further enhance cross-subject generalization with limited labeled data, which have recently shown effectiveness in other EEG analysis tasks [50]–[52]. Extending the proposed alignment mechanism to other EEG-based paradigms also represents a promising direction for future research.

VI. Conclusion

In this study, we proposed DSGNet, a subject-independent deep learning framework for motor imagery EEG classification, which integrates dual convolutional branches for temporal–spectral representation learning with a structure-aware class alignment loss. By explicitly enforcing intra-class compactness, inter-class separability, and cross-subject class-center consistency, DSGNet effectively mitigates inter-subject variability without requiring labeled data from unseen individuals. Extensive experiments conducted on four public MI-EEG datasets under a rigorous leave-one-subject-out protocol demonstrate that DSGNet achieves state-of-the-art or highly competitive performance, particularly in challenging multi-class scenarios. Ablation studies and visualization analyses further validate the complementary roles of the temporal and spectral branches, as well as the effectiveness of semantic structure alignment in enhancing generalization. These results highlight the potential of DSGNet as a reliable and scalable solution for subject-independent MI-BCI systems in neurorehabilitation applications.

References

[1] J. Wang, L. Bi, and W. Fei, "EEG-based motor BCIs for upper limb movement: current techniques and future insights," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4413–4427, 2023.

[2] D. L. Eaves, M. Riach, P. S. Holmes, and D. J. Wright, "Motor imagery during action observation: a brief review of evidence, theory and future research opportunities," *Frontiers in neuroscience*, vol. 10, p. 514, 2016.

[3] B. D. Berman, S. G. Horovitz, G. Venkataraman, and M. Hallett, "Self-modulation of primary motor cortex activity with motor and motor imagery tasks using real-time fMRI-based neurofeedback," *Neuroimage*, vol. 59, no. 2, pp. 917–925, 2012.

[4] H. Sun, Y. Ding, J. Bao, K. Qin, C. Tong, J. Jin, and C. Guan, "Leveraging temporal dependency for cross-subject-MI BCIs by contrastive learning and self-attention," *Neural Networks*, vol. 178, p. 106470, 2024.

[5] D. Hagemann, "Individual differences in anterior EEG asymmetry: methodological problems and solutions," *Biological psychology*, vol. 67, no. 1-2, pp. 157–182, 2004.

[6] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2021.

[7] L. Zhou, N. Li, M. Ye, X. Zhu, and S. Tang, "Source-free domain adaptation with class prototype discovery," *Pattern recognition*, vol. 145, p. 109974, 2024.

[8] H. Wang, T. Shen, W. Zhang, L.-Y. Duan, and T. Mei, "Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation," in *European conference on computer vision*. Springer, 2020, pp. 642–659.

[9] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz *et al.*, "Review of the BCI competition IV," *Frontiers in neuroscience*, vol. 6, p. 55, 2012.

[10] F. R. B. Yang, "WBCIC-SHU motor imagery dataset," https://doi.org/10.25452/figshare.plus.22671172.v5, 2024.

[11] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "Eeg dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.

[12] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery BCI systems," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 15–29, 2015.

[13] C. F. Blanco-Diaz, J. M. Antelis, and A. F. Ruiz-Olaya, "Comparative analysis of spectral and temporal combinations in CSP-based methods for decoding hand motor imagery tasks," *Journal of Neuroscience Methods*, vol. 371, p. 109495, 2022.

[14] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.

[15] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14 681–14 722, 2023.

[16] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.

[17] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE reviews in biomedical engineering*, vol. 14, pp. 204–218, 2020.

[18] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[19] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[20] X. Chen, X. Teng, H. Chen, Y. Pan, and P. Geyer, "Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX," *Biomedical Signal Processing and Control*, vol. 87, p. 105475, 2024.

[21] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2022.

[22] Z. Miao, M. Zhao, X. Zhang, and D. Ming, "LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability," *NeuroImage*, vol. 276, p. 120209, 2023.

[23] Y. Ding, Y. Li, H. Sun, R. Liu, C. Tong, C. Liu, X. Zhou, and C. Guan, "EEG-Deformer: A dense convolutional transformer for brain-computer interfaces," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[24] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE transactions on industrial informatics*, vol. 19, no. 2, pp. 2249–2258, 2022.

[25] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 2958–2965.

[26] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 10, pp. 2164–2177, 2019.

[27] C. Zhang, Y.-K. Kim, and A. Eskandarian, "EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046014, 2021.

[28] A. Apicella, P. Arpaia, G. D'Errico, D. Marocco, G. Mastrati, N. Moccaldi, and R. Prevete, "Toward cross-subject and cross-session generalization in EEG-based emotion recognition: Systematic review, taxonomy, and methods," *Neurocomputing*, p. 128354, 2024.

[29] H. Zhao, Q. Zheng, K. Ma, H. Li, and Y. Zheng, "Deep representation-based domain adaptation for nonstationary EEG classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 535–545, 2020.

[30] D. Zhang, H. Li, and J. Xie, "MI-CAT: A transformer-based domain adaptation network for motor imagery classification," *Neural Networks*, vol. 165, pp. 451–462, 2023.

[31] D.-q. Xu and M.-a. Li, "A dual alignment-based multi-source domain adaptation framework for motor imagery EEG classification," *Applied Intelligence*, vol. 53, no. 9, pp. 10 766–10 788, 2023.

[32] M. Miao, Z. Yang, Z. Sheng, B. Xu, W. Zhang, and X. Cheng, "Multi-source deep domain adaptation ensemble framework for cross-dataset motor imagery EEG transfer learning," *Physiological Measurement*, vol. 45, no. 5, p. 055024, 2024.

[33] D. Zhang, H. Li, J. Xie, and D. Li, "MI-DAGSC: A domain adaptation approach incorporating comprehensive information from MI-EEG signals," *Neural Networks*, vol. 167, pp. 183–198, 2023.

[34] S. Liu, L. An, C. Zhang, and Z. Jia, "A spatial-temporal transformer based on domain generalization for motor imagery classification," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2023, pp. 3789–3794.

[35] H. Song, Q. She, F. Fang, S. Liu, Y. Chen, and Y. Zhang, "Domain generalization through latent distribution exploration for motor imagery EEG classification," *Neurocomputing*, vol. 614, p. 128889, 2025.

[36] S. Liang, C. Xuan, W. Hang, B. Lei, J. Wang, J. Qin, K.-S. Choi, and Y. Zhang, "Domain-generalized EEG classification with category-oriented feature decorrelation and cross-view consistency learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3285–3296, 2023.

[37] H. Zhi, T. Yu, Z. Gu, Z. Lin, L. Che, Y. Li, and Z. Yu, "Supervised contrastive learning-based domain generalization network for cross-subject motor decoding," *IEEE Transactions on Biomedical Engineering*, 2024.

[38] X.-C. Zhong, Q. Wang, D. Liu, Z. Chen, J.-X. Liao, J. Sun, Y. Zhang, and F.-L. Fan, "EEG-DG: A multi-source domain generalization framework for motor imagery EEG classification," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[41] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[44] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[45] S. Narayan, "The generalized sigmoid activation function: Competitive supervised learning," *Information sciences*, vol. 99, no. 1-2, pp. 69–82, 1997.

[46] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.

[47] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[48] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[49] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[50] X. Chen, Y. Zhang, Q. Chen, L. Zhou, H. Chen, H. Wu, Y. Xu, K. Chen, B. Yin, W. Chen *et al.*, "ASTGSleep: Attention based spatial-temporal graph network for sleep staging," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[51] X. Chen, Y. Zhang, Z. Tang, H. Wu, C. Chen, and W. Chen, "MtRBD: Advancing iRBD analysis with multi-task learning for joint sleep staging and RSWA detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 12, pp. 8743–8750, 2025.

[52] L. Zhang, H. Shi, Z. Li, W.-L. Zheng, and B.-L. Lu, "Multi-view self-supervised domain adaptation for EEG-based emotion recognition," *IEEE Transactions on Affective Computing*, 2025.