

Breaking the Depth Barrier in Motor Imagery Classification via a Residual Depthwise-Separable Network

Ruiyu Zhao¹, Ian Daly², Xinjie He³, Ruitian Xu⁴, Chongfeng Wang, Xingyu Wang⁵,
Andrzej Cichocki⁶, *Life Fellow, IEEE*, and Jing Jin⁷, *Senior Member, IEEE*

Abstract—Lightweight networks that include depthwise-separable convolution are widely used in motor imagery (MI) electroencephalogram (EEG) decoding of brain-computer interface (BCI). Many established MI classification networks are relatively shallow, preventing them from benefiting from the hierarchical feature extraction capabilities of deeper structures. Due to suboptimal residual connection structures, the mismatched residual baseline layer design, and the poor compatibility between data preprocessing and residual modules, the deepening of networks cannot be effectively combined with residual structures. This creates a depth barrier that hinders further performance improvements. To address these challenges, we propose a novel method, residual depthwise-separable deep neural network (ResDSNet), built upon an unraveled view-path analysis of residual connection structures. The analysis reveals that the residual mechanism achieves optimal performance when the layer distribution across different paths approximates a binomial distribution. Furthermore, we design a residual depthwise-separable convolution module and a tailored data-preprocessing module that effectively integrate with the residual structure, filtering noise and retaining MI task features. We evaluate ResDSNet on three publicly available datasets, including the BCI Competition IV Dataset IIa, the BCI Competition IV Dataset IIb, and the PhysioNet dataset, which collectively

contain EEG signals recorded from 127 human subjects. ResDSNet achieves accuracies of 79.36%, 84.95%, and 64.13%, outperforming state-of-the-art methods by 3.16%, 1.59%, and 8.40% with statistical significance. Experimental results indicate that ResDSNet fully unlocks the hierarchical representation capabilities of deep networks for MI-EEG decoding, achieving robust performance and demonstrating substantial potential to overcome the inherent challenges in BCIs.

Index Terms—Brain-computer interface (BCI), convolutional neural network (CNN), motor imagery (MI), residual mechanism.

I. INTRODUCTION

BRAIN-COMPUTER interface (BCI) is a communication system that enables direct information transmission between the brain and external devices without relying on peripheral neural control [1]. BCIs are generally classified as either invasive or noninvasive. Noninvasive BCIs have attracted significant attention due to their ease of operation and user-friendliness. Common paradigms used within noninvasive BCIs include event-related potentials (ERP) [2], steady-state visual evoked potentials (SSVEPs) [3], and motor imagery (MI) [4]. Different paradigms can be combined and applied to various scenarios [5], [6], [7], [8]. These include controlling robotic systems [9], hand kinematics reconstruction [10], [11], and enabling user interaction with smart wearables or mixed-reality platforms on mobile devices [12]. The MI paradigm is based on the imagination of bodily movements without actual movement execution and is widely used within noninvasive electroencephalographic (EEG)-based BCIs [13].

Conventional approaches to MI classification often employ common spatial patterns (CSPs) and their variants and applications [14], [15], [16], [17]. MI classification networks are inspired by computer vision (CV) and natural language processing (NLP). A variety of network structures have been successfully employed for MI classification, including convolutional neural networks (CNNs) [18], [19], [20], [21], [22], [23], recurrent neural networks (RNNs) [24], [25], graph neural networks [26], and Transformers [27], [28], [29], [30]. These network structures integrate diverse machine learning mechanisms, such as reinforcement learning [31], transfer learning [1], [32], and self-supervised learning [33], to perform MI classification. However, the direct application of network structures and mechanisms from CV and NLP to MI signal recognition does not necessarily improve accuracy (Acc.) and may even decrease it [28], [34].

Received 25 May 2025; revised 27 August 2025, 9 January 2026, and 19 April 2026; accepted 2 May 2026. This work was supported by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project 2022ZD0208900, in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX; This research is also supported by key core technologies under Grant BE2022064-1, in part by the Lingang Laboratory under Grant No. LGL8998. This article was recommended by Associate Editor L. Jiao. (*Corresponding author: Jing Jin.*)

Ruiyu Zhao, Xinjie He, Ruitian Xu, Chongfeng Wang, and Xingyu Wang are with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China (e-mail: zhaoruiyu@mail.ecust.edu.cn; xinjiehe@mail.ecust.edu.cn; xrt_1016@163.com; 806237081@qq.com; xywang@ecust.edu.cn).

Ian Daly is with the Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K. (e-mail: i.daly@essex.ac.uk).

Andrzej Cichocki is with the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wakō 351-0198, Japan, also with the Systems Research Institute of Polish Academy of Science, 01-447 Warsaw, Poland, and also with the Department of Informatics, Nicolaus Copernicus University, 87-100 Toruń, Poland (e-mail: a.cichocki@riken.jp).

Jing Jin is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, Shanghai 200237, China, also with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China, and also with the School of Math, East China University of Science and Technology, Shanghai 200237, China (e-mail: jinjingat@gmail.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2026.3690707>.

Unlike the deep CNNs used in CV and NLP applications, the MI-EEG classification networks generally benefit from shallow networks [18], [19], [34], [35]. Data preprocessing in MI classification networks continues to incorporate traditional EEG signal processing methods, with network structures deliberately designed to align with these preprocessing pipelines. This codesigned structure aims to mitigate the challenge of extracting discriminative features from inherently weak MI-EEG signals [19], [20], [36], [37], [38]. This structural choice is primarily dictated by the challenges associated with MI-EEG data, including low signal-to-noise ratios and the inherent risk of overfitting deep networks on small datasets.

Compared with the abundant large-scale annotated datasets available in CV or NLP [39], [40], publicly available datasets for MI task classification remain relatively scarce [41], [42]. The scarcity of MI-EEG data is attributable to significant collection challenges, including costly and time-consuming experimental sessions, as well as substantial data variability between subjects and across sessions for the same subject, which arises from the inherent noise, nonlinearity, and non-stationarity of EEG signals [43], [44]. Consequently, most MI datasets contain a limited number of subjects. For instance, the widely used BCI Competition datasets in BCI research also encompass EEG data from a relatively small cohort of subjects [42].

As a result of these data constraints, many established MI classification models, ranging from lightweight convolutional networks to Transformer-based structures, remain relatively shallow. This structural depth barrier prevents them from benefiting from the depth-driven hierarchical representation learning that has driven significant advances in CV and NLP. While residual mechanisms are widely utilized to construct deep structures in these fields [45], directly transferring them to MI classification is problematic. For instance, because the self-attention mechanism in Transformers lacks structural bias, deep Transformer models without pretraining often severely overfit on small- or medium-sized EEG datasets [46]. Moreover, even when attempting to mitigate this through large-scale pretraining, the EEG foundation models still require fine-tuning for specific downstream tasks, and their MI decoding capability remains substantially limited, often falling short of state-of-the-art (SOTA) MI classification models [47], [48].

Due to these fundamental differences in data characteristics and model structures, the baseline layers typically used in CV and NLP residual structures are suboptimal for MI data. Simply adding residual connections to MI networks often fails to yield consistent improvements. Effectively deepening these networks instead requires the simultaneous alignment of three critical factors, including the residual connection structure, the residual baseline layer design, and the compatibility between data preprocessing and residual modules.

To address these three challenges, we develop a residual depthwise separable deep neural network (ResDSNet). This network comprises a data preprocessing module, a residual depthwise separable convolution module, and a classification module. To integrate the preprocessed data with the subsequent residual mechanism, we employ a filter bank for dual-frequency band segmentation to process the aligned data.

This multiband feature extraction enhances the capture of effective information within low signal-to-noise ratio MI-EEG signals and reduces intertrial distribution variations and input redundancy. This improves the recognizability of the model's input data [19], [49]. To improve data extraction efficiency, we utilize 1-D convolutional operations [19], [35]. In the residual depthwise separable convolution module, to match the 4–16 and 16–40 Hz nonoverlapping frequency bands, we employ a design combining depthwise separable convolution layers with a residual mechanism. Through an unraveled view comparison with various residual structures, we statistically analyze the distribution of paths across different layers and find that the residual depthwise separable mixed (RDSM) layer achieves optimal performance when the residual connections approximate a binomial distribution. We utilize larger kernels in shallower layers and smaller kernels in deeper layers, enabling a progressive increase in the depth of layers with different kernel sizes. In the classification layer, we use GAP, which can integrate spatial information between features, to produce the final output result. In addition, our network applies the same data settings to any dataset, eliminating the need to switch configurations across different datasets. This demonstrates that the designed network possesses strong robustness, can adapt to different subjects across various datasets, and delivers superior recognition performance. These claims are validated on real EEG data recorded from 127 human subjects across three benchmark datasets, without any numerical simulations. The main contributions of this article are as follows.

- 1) We propose a ResDSNet for MI classification, achieving statistically significant improvements over SOTA methods across three datasets containing 127 human subjects.
- 2) We design six residual structures and reveal, through unraveled view analysis, that optimal performance is achieved when layer distributions across paths approximate a binomial distribution. This insight provides a universal architectural criterion for residual networks in MI classification, extending beyond ResDSNet.
- 3) We devise an optimized depthwise separable convolution tailored specifically for the residual baseline layer, which effectively filters noise and retains MI task features.
- 4) We develop a tailored data preprocessing module to ensure optimal compatibility with residual modules, fully exploiting the advantages of network depth on real MI-EEG data.

The remainder of this article is organized as follows. In Section II, we review the related work. In Section III, we present our proposed method, ResDSNet. In Section IV, we present the experimental results. In Section V, we present a discussion of the results. Finally, in Section VI, we describe the conclusions from our work.

II. RELATED WORK

A. MI-EEG Data Preprocessing

MI-EEG signals contain rich temporal, spatial, and spectral features that require different preprocessing strategies. In the temporal domain, MI-EEG signals are augmented through techniques such as repeated and random cropping,

overlapping time windows, and random erasing [19], [25]. In the spatial domain, EEGProgress [50] employs a horizontal topological permutation rule to optimize channel arrangement. Preprocessing in the spectral domain is the most important for MI tasks. Extracting discriminative MI signals from raw EEG requires appropriate filtering to isolate the μ (8–12 Hz) and β (18–26 Hz) bands that encompass sensorimotor event-related desynchronization/synchronization (ERD/ERS), while simultaneously attenuating low-frequency noise such as electrooculography artifacts [51], [52]. Different networks adopt different filtering strategies. For example, CSP [14], ShallowNet [34], EEGNet [18], and EEGConformer [28] employ a 4–40 Hz filtering band, DPMSNet [22] adopts 0.4–40 Hz, while DBConformer [30] applies notch filtering and bandpass filtering. However, these approaches rely on a single frequency band and cannot accommodate intersubject spectral variability. Filter bank CSP (FBCSP) decomposes MI data into multiple subbands via a filter bank. FBCNet [36], FBMSNet [37], and MFRCNet [20] extend this idea by designing parallel convolutional branches corresponding to each subband. In contrast, IFNet [19] leverages cross-frequency coupling and processes MI data with only two frequency bands. After preprocessing, MI-EEG signal features are extracted by designing different temporal, spatial, and spectral structures [22], [38]. FACTNet [53] employs frequency adapter mechanisms with temporal-periodicity inception to enhance spectral feature extraction. STGF [54] leverages graph-based representations to fuse spatial and temporal features. SSTDPN [55] jointly models spatial-spectral and temporal characteristics through a dual-prototype architecture.

B. Deep Learning-Based MI-EEG Classification

Deep learning-based MI-EEG classification models focus on capturing heterogeneous features through diverse structures. ShallowNet utilizes a neural network to achieve FBCSP-like log-variance calculations aimed at decoding band power features [14], [34]. In the EEGNet structure, depthwise separable convolution [56], [57] is incorporated to achieve a compact structure [18]. Building upon this foundation, DHCA [21] introduces dynamic hierarchical convolutional attention to dynamically capture time-frequency information. FBCNet [36] employs a strategy combining multiple frequency bands with multiscale convolutions to precisely capture MI-EEG signals across different subjects, spawning a series of variants [20], [37]. Attention mechanisms have also been integrated into MI classification structures. For instance, DMSACNN [58] utilizes deep multiscale attentional mechanisms to capture hierarchical features, MPSTANet [23] designs multiscale pooling spatial-temporal attention for improved cross-session decoding, and EISATCFusion [59] combines Inception modules with self-attention and temporal convolutional networks. EEGConformer [28] and CTNet [29] integrate convolutional layers with Transformer blocks to enhance feature extraction capability, while TMSANet [60] proposes a temporal multiscale attention module to address the limitation of Transformers in comprehensively leveraging both local and global features. Nevertheless, these models rely on shallow designs,

as simply stacking conventional convolutions on low signal-to-noise ratio MI-EEG data typically causes severe overfitting, resulting in a depth barrier. Furthermore, while depthwise separable convolutions are efficient in shallow networks [19], their potential as foundational baseline layers for deep residual structures remains unexplored.

Given the limited scale of MI datasets, lightweight models are widely utilized to achieve efficient classification while reducing computational overhead. EEGSimpleConv [35] proposes a stacked 1-D convolutional structure, and SNNforMI [61] is based on spiking neural networks. Furthermore, EEG foundation models adopt large model frameworks, aggregating EEG data from diverse tasks and increasing network depth and parameters by stacking multiscale attention layers for cross-dataset task recognition. For instance, EEGPT [62] aligns spatiotemporal representations via a hierarchical Transformer, LaBraM [63] employs channelwise patch embedding with vector-quantized neural spectrum prediction, NeuroLM [47] leverages text-aligned neural tokenization and multitask instruction tuning, and LCM [64] integrates temporal-spectral attention mechanisms. Although these foundation models increase network depth through large-scale cross-task pretraining, their MI decoding performance still falls short of SOTA task-specific models [47], [48].

C. Residual Connections

Residual connections [65] enable the training of extremely deep networks by introducing identity shortcuts. The unraveled view proposed by Veit et al. [66] reveals that residual networks essentially function as exponential ensembles of relatively shallow paths. In MI decoding, several structures have leveraged this concept. For instance, ShallowConvNet [34] investigated the impact of residual connections on MI networks. CETRNet [67] designed a temporal convolutional residual block to extract high-level temporal features. MFRCNet [20] integrates multiscale feature extraction with residual connections. Similarly, STDBNet [68] introduced learnable residual connections deeply integrated with channel attention. MSHANet [69] employed a multiscale residual block with hybrid attention for improved decoding performance. Furthermore, recent EEG foundation models, such as EEGPT [62], LaBraM [63], NeuroLM [47], and LCM [64], employ standard residual connections combined with prenormalization, a configuration commonly utilized in large language models (LLMs). However, standard residual structures from CV or NLP are not directly applicable to MI tasks, and the design of MI-adapted residual architectures lacks systematic investigation into subject structural reusability.

III. METHOD

We develop a ResDSNet for MI classification. This network is comprised of a data preprocessing module, a residual depthwise separable convolution module, and a classification module. These are illustrated in Fig. 1. The parameters of our ResDSNet model are listed in Table I.

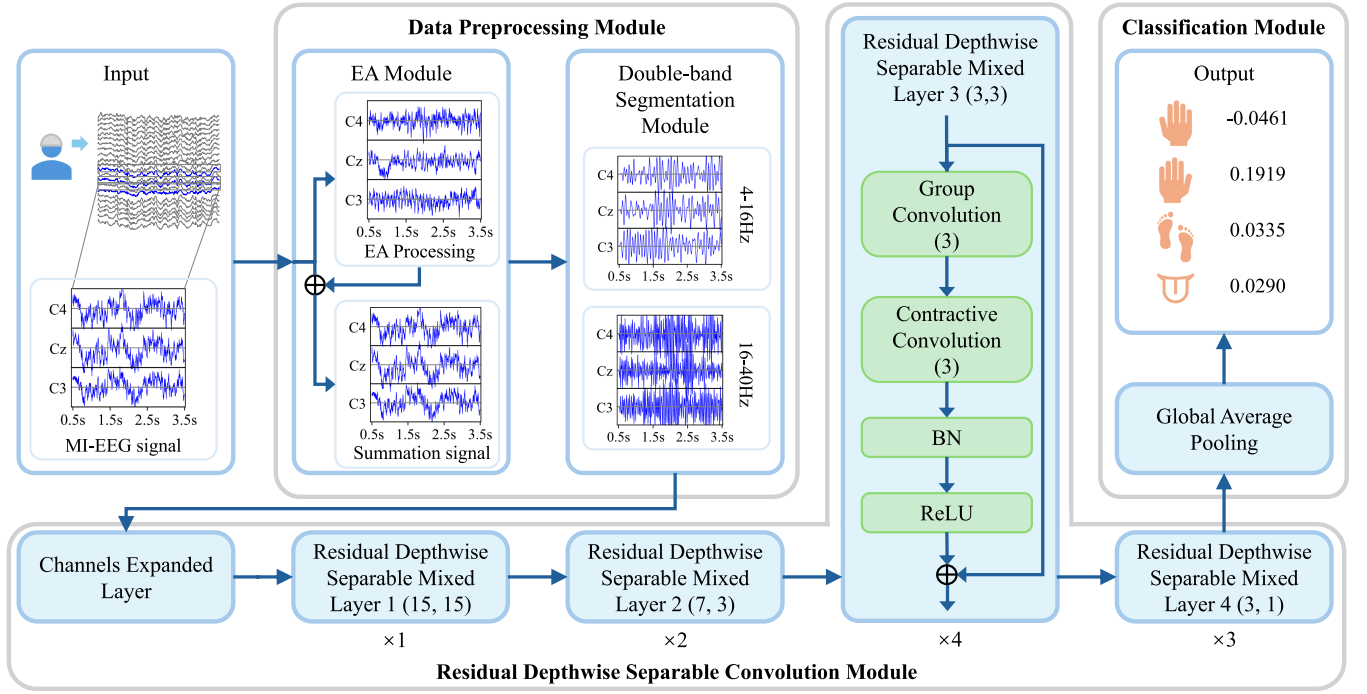


Fig. 1. Structure of the ResDSNet.

TABLE I
PARAMETERS OF OUR PROPOSED METHOD RESDSNET

Layer	Kernel Size	Output
input	–	(B, C, T)
EA	–	(B, C, T)
double-band segmentation	–	$(B, 2C, T)$
channel expanded	1	(B, C_1, T)
RDSM1 $\times 1$	15, 15	(B, C_1, T)
RDSM2 $\times 2$	7, 3	(B, C_1, T)
RDSM3 $\times 4$	3, 3	(B, C_1, T)
RDSM4 $\times 3$	3, 1	(B, C_1, T)
global average pooling	1	$(B, C_1, 1)$
dropout	–	(B, C_1)
flatten	–	$(B \times C_1)$
fully connection	–	N_C

RDSM denotes residual depthwise separable mixed layer. B denotes the batch size, C denotes the number of channels of motor imagery EEG signals, C_1 denotes the number of channels input into the Residual Depthwise Separable Mixed Layer after channel expansion, N_C denotes the number of motor imagery tasks, T denotes the number of sampling points of MI signals, and $\times X, X \in \{1, 2, \dots, N\}$ represents the number of superpositions to be added to the layer.

A. Data-Preprocessing Module

A single trial of raw MI-EEG signals is denoted as (x_i, y_i) , $i = 1, 2, \dots, n$, where $x_i \in \mathbb{R}^{C \times T}$ and $y_i \in \{1, 2, \dots, N_c\}$, with x_i representing MI-EEG data, y_i the trial labels, C the number of channels, T the number of sample points, and N_c the number of MI tasks. Therefore, (X, Y) is input n trials of MI-EEG data, where $X = [x_1, x_2, \dots, x_n]^T$ and $Y = [y_1, y_2, \dots, y_n]^T$.

First, we use the Euclidean-space data alignment (EA) module to the original MI data. The EA module first computes the average covariance matrix for each subject using n trials (x_i, y_i) , resulting in a reference matrix

$$R = \frac{1}{n} \sum_{i=1}^n x_i x_i^T. \quad (1)$$

We then perform data alignment as follows:

$$\bar{X} = R^{-1/2} X \quad (2)$$

where $\bar{X}$ denotes the aligned MI-EEG data. By computing the average covariance matrix across trials as the reference matrix, this method aligns the raw data to effectively minimize distributional discrepancies between different trials. Meanwhile, it blends the processed data with the original data in a 5:1 ratio, enhancing data robustness while preserving the characteristics of the original MI-EEG signals.

The MI-EEG signals (B , C , and T) are then transformed by the double-band segmentation (DS) module, where B represents batch trials. The DS using a filter bank reconstructs single-frequency data into various frequency bands for data augmentation, enhancing feature extraction efficiency [19]. The filter bank $F = \{f_i\}_{i=1}^{N_b}$ uses N_b fifth-order Butterworth filters, $f_i \in \mathbb{R}^{(N_b \times C) \times T}$, applied in nonoverlapping bands, spanning 4–16 and 16–40 Hz. The filtered MI data output becomes

$$X_{FB} = \mathcal{F}_{EB}(\bar{X}), \quad X_{FB} \in \mathbb{R}^{(N_b \times C) \times T} \quad (3)$$

where $\mathcal{F}_{EB}$ denotes the DS operation, with $N_b = 2$ narrow-band temporal filters. The channel dimension of the MI data increases from (B, C, T) to $(B, 2 * C, T)$.

B. Residual Depthwise-Separable Convolution Module

The residual depthwise-separable convolution module comprises the channel expanded layer and the RDSM layer.

1) *Channel Expanded Layer*: The first layer in MI recognition networks often uses spatial convolution to capture spatial characteristics across channels. However, the structure of the spatial convolution can significantly impact both the performance and complexity of CNN-based MI-EEG classification

models [50]. Therefore, we leverage the properties of 1-D convolution [19], [35], implementing a channel-expanded layer as the network's first layer. The channel-expanded layer takes the low-dimensional compressed representation of data from the preprocessing module and expands it to a higher dimension, thereby enhancing the informational richness for subsequent layers and increasing the number of residual inputs passed to the following feature maps [65]. The channel-expanded layer, denoted as $\mathcal{F}_{EX}$, is computed as

$$X_{EX}(C_{EX}) = \mathcal{F}_{EX}(X_{FB}(C_{in})) \quad (4)$$

where X_{EX} denotes the output of the channel-expanded layer. $C_{in} = 2C$ and $C_{EX} = C_1$ denote the input channels and output channels, respectively. The data are transformed by a 1×1 convolution kernel along the channel dimension, changing from $(B, 2C, T)$ to (B, C_1, T) , where C_1 represents a manually set input channel for group convolution, improving channel usability and capturing common features across datasets. Here, we set $C_1 = 80$.

2) *RDSM Layer*: After processing the channel features in the first layer, the convolution operation typically moves to the time dimension T . In developing ResDSNet, our primary focus is on establishing a suitable baseline layer structure. We analyze commonly used MI classification models to identify an appropriate baseline layer structure.

EEGNet introduced depthwise-separable convolution to create a compact structure [18], and some EEG-focused networks—such as FBCNet, FBMSNet, IFNet, and TSFCNet—have adopted it as well. Depthwise-separable convolution differs from regular convolution (RC) by processing each channel individually. This means that feature information across different branches in the depth dimension remains isolated, only merging in the separable convolution, to achieve isolated feature extraction for different bands. This characteristic minimizes noise interference between channels.

Between standard convolution and depthwise-separable convolution, there exist various kernel and channel variations [56]. To optimize depthwise-separable convolution for MI tasks, we incorporate group convolution $\mathcal{F}_{GC}$ and contractive convolution $\mathcal{F}_{CC}$. Each channel is processed independently, and the group convolution $\mathcal{F}_{GC}$ is computed as

$$X_{GC}^{(g)}\left(\frac{C_{EX}}{G}\right) = \mathcal{F}_{GC}^{(g)}\left(X_{EX}^{(g)}\left(\frac{C_{EX}}{G}\right)\right) \quad (5)$$

where g denotes the group index, $g = 0, 1, \dots, C_{EX}/G - 1$. $X_{EX}^{(g)}(C_{EX}/G)$ denotes the input channels in group g , and C_{EX} is divided into groups of size G channels per group. $\mathcal{F}_{GC}^{(g)}$ denotes the group convolution operation in group g , and $X_{GC}^{(g)}$ denotes the output of the group convolution in group g .

In the contractive convolution step $\mathcal{F}_{CC}$, unlike the original pointwise design with a kernel size of 1, we apply different kernel sizes to capture localized temporal information at various layers. The contractive convolution step $\mathcal{F}_{CC}$ is computed via

$$X_{CC}(C_{EX}) = \sum_{g=0}^{\frac{C_{EX}}{G}-1} \mathcal{F}_{CC}^{(g)}\left(X_{GC}^{(g)}\left(\frac{C_{EX}}{G}\right)\right) \quad (6)$$

where $\mathcal{F}_{CC}^{(g)}$ denotes the contraction convolution in group g , X_{CC} denotes the output of the contractive convolution, and C_{EX} denotes the output channels. Then, the depthwise-separable convolution from (5) and (6) is simply written as

$$X_{CC} = \mathcal{F}_{CC} \circ \mathcal{F}_{GC}(X_{EX}) \quad (7)$$

where $\circ$ denotes the function composition operation.

Following the depthwise-separable convolution layer, we introduce a normalization layer $\mathcal{F}_{Norm}$ and a ReLU activation function $\mathcal{F}_{ReLU}$ to obtain the output X_{DSC} . This is computed via

$$X_{DSC} = \mathcal{F}_{ReLU} \circ \mathcal{F}_{Norm}(X_{CC}). \quad (8)$$

This addition enhances the network's nonlinear expressive capability, enabling it to fit more complex functions. Thus, our depthwise-separable convolution layer comprises group convolution, contractive convolution, a normalization layer, and the ReLU activation function, creating a robust foundation for MI classification.

MI-EEG signals are characterized by a low signal-to-noise ratio, compounded by challenges such as limited samples and significant trial-to-trial variability within MI EEG datasets. As network depth increases, Acc. tends to plateau and even decline, while loss values rise, leading to training degradation as gradients fail to propagate effectively. To address these issues, we introduced a residual mechanism within the residual depthwise-separable convolution block, establishing the RDSM layer.

In the residual mechanism, the residual output mapping is defined as $\text{Concat}(\mathcal{F}_{res}(x_s), x_s)$, where $\mathcal{F}_{res}(x_s)$ represents the feature map output by the depthwise-separable mixed layer, and x_s denotes the input data passed through a shortcut connection. This approach has the potential to significantly enhance the training effect of deep networks and alleviate the training degradation problems caused by noise signals. Specifically, by positioning the ReLU activation at the end of the residual branch $\mathcal{F}_{res}$, we enforce a nonnegative sparsity constraint. Such sparsity encourages the decoupling of features and helps the network focus on significant activations [70]. For low signal-to-noise ratio MI-EEG signals, this mechanism acts to suppress weak, noise-related fluctuations while preserving discriminative patterns.

Therefore, we construct RDSM layers

$$X_{RDSC}(C_{EX}) = \text{Concat}(\mathcal{F}_{res}(X_{EX}(C_{EX})), X_{EX}(C_{EX})) \quad (9)$$

where X_{RDSC} denotes the output of the RDSM layer, and $\mathcal{F}_{res}$ denotes the function composition $\mathcal{F}_{ReLU} \circ \mathcal{F}_{Norm} \circ \mathcal{F}_{CC} \circ \mathcal{F}_{GC}$.

Guided by the path distribution analysis of residual networks detailed in Section IV-D, we design each residual block to consist of RDSM layers with identical parameters. We have designed a total of four residual blocks with varying parameters to optimize feature extraction. Inspired by the use of large convolutional kernels in the early layers of ResNet, we incorporate large kernels in the initial shallow blocks to capture broader temporal contexts and coarse-grained information in MI-EEG signals. As the network progresses, smaller kernels in the deeper layers capture fine-grained information over shorter time intervals. Our proposed ResDSNet model

supports adjustable network depth by varying the number of layers within each block, offering flexibility to adapt the network depth to different levels of complexity in MI-EEG classification tasks.

C. Classification Module

Processed MI data typically contains a high number of features along the temporal dimension. These features typically exhibit significant intraclass variance and high noise levels. To reduce the feature dimensionality, average pooling or max pooling is commonly applied. The GAP effectively reduces the dimensionality of the data and extracts essential features along the temporal axis. This is computed via

$$X_{\text{GAP}}^{(c)} = \frac{1}{T} \sum_{i=0}^{T-1} X_{\text{RDSC}}^{(c)}(i) \quad (10)$$

where $X_{\text{RDSC}}^{(c)}(i)$ denotes the value of the c th channel in the input feature map at position i . The term $X_{\text{GAP}}^{(c)}$ denotes the output of the c th channel. This operation reduces the MI signal dimensions from the original (B, C_1, T) to $(B, C_1, 1)$. A dropout layer is then applied to reduce the parameter count and mitigate overfitting. Finally, the data are flattened and passed through a fully connected layer to produce the final classification output $\hat{Y}$

$$\hat{Y} = \mathcal{F}_{\text{FC}} \circ \mathcal{F}_{\text{FL}}(X_{\text{GAP}}) \quad (11)$$

where $\mathcal{F}_{\text{FL}}$ denotes the flattening operation and $\mathcal{F}_{\text{FC}}$ denotes the fully connected layer operation.

IV. EXPERIMENTS

A. Experiments Setting

We evaluated the proposed network using three publicly available benchmark datasets. All three datasets contain real EEG signals recorded from human subjects performing MI tasks. These datasets are the BCI Competition IV Dataset IIa (IV2a) [42], the BCI Competition IV Dataset IIb (IV2b) dataset [42], and the PhysioNet dataset [71].

The IV2a dataset includes EEG recorded under four MI classes: left-hand, right-hand, feet, and tongue imagery. Real MI-EEG data from nine healthy human subjects is recorded using 22 electrodes at a sampling rate of 250 Hz. For each subject, the first session serves as the training set and the second session serves as the test set, setting up a subject-independent evaluation protocol. Each session contains 288 trials, with 72 trials per class.

The IV2b dataset includes left-hand and right-hand imagery. Real MI-EEG data are recorded from nine healthy human subjects using three electrodes and sampled at 250 Hz. For each subject, three sessions are designated as the training set, while the last two sessions form the test set, setting up a subject-independent evaluation protocol. Each session contains 120 trials. All trials in the IV2a dataset and the IV2b dataset use data from 0.5 to 3.5 s, starting from the beginning cue of the MI tasks.

The PhysioNet dataset includes EEG data recorded during motor execution, resting-state, and MI tasks. We selected

four types of MI data, including left-fist, right-fist, fists, and feet imagery. Real MI data were recorded from 109 healthy human subjects using 64 electrodes at a sampling rate of 160 Hz. PhysioNet contains six sessions, in which each session includes two out of the four MI tasks. The first 50% of trials in each session are combined for training, while the latter 50% of trials are aggregated for testing, setting up a subject-independent evaluation protocol. All trials of the PhysioNet datasets use 3 s starting from the beginning cue of the MI tasks.

We employ the cross-entropy loss together with the Adam optimizer to update all ResDSNet-related model parameters during training. The learning rate of the Adam optimizer is set to 0.003. The number of training epochs is set to 1000 for all ResDSNet-related models, and the model from the final epoch is tested. A batch size of 512 is used for all ResDSNet-related models. Increasing the batch size within the same epoch reduces training time and promotes stable gradient descent, thereby enhancing network training effectiveness. When the number of trials in the training set is less than the desired batch size, the training batch size is set to the number of trials. We use an NVIDIA RTX 4090 GPU, an Intel i9-13900KF processor, and 32 GB of RAM for training and testing.

B. Comparison With State-of-the-Art Methods

We compare our proposed ResDSNet with several SOTA models, including ShallowNet [34], DeepNet [34], EEGNet [18], FBCNet [36], EEGConformer [28], FBMSNet [37], IFNet [19], TSFCNet [38], ADFCNN [72], EISATCFusion [59], DoNet [73], EEGSimpleConv [35], CTNet [29], FACTNet [53], STGF [54], DMSACNN [58], SSTDPN [55], TMSANet [60], SNNforMI [61], DBConformer [30], on IV2a, IV2b, and PhysioNet datasets. The experimental evaluation metrics adopted include Acc., Prec., Rec., $F1$, and Cohen's kappa coefficient (Kappa) [58], [74]. The p -value (p) from the t-test is used to examine whether the proposed method achieves a statistically significant improvement in Acc. The results are shown in Table II. To ensure comparability, all models adopt their original hyperparameters except for necessary input/output adjustments. Input trials are standardized to 3 s segments, and the final fully connected layer dimensions are modified to match dataset class counts.

As shown in Table II, the proposed ResDSNet model achieves optimal results compared with SOTA methods. In terms of Acc., it attains 79.36% on the IV2a dataset, 84.95% on the IV2b dataset, and 64.13% on the PhysioNet dataset. Variations in factors such as subjects, acquisition channels, and sampling frequencies across different datasets lead to performance differences for the same network structure.

The combined work of the residual depthwise-separable convolution module and the data preprocessing module achieves substantial performance improvements over models utilizing single-channel EEG data, such as EEGNet [18] and ShallowNet [34], and over models leveraging multiband approaches, such as FBMSNet [37], TSFCNet [38], and MFR-CNet [20]. Furthermore, the synergy between the EA and dual-band features enables superior results compared with IFNet. Compared to Transformer-based models employing

TABLE II

COMPARISONS OF RESDSNET AND SOTA MODELS ON IV2a, IV2b, AND PHYSIONET DATASETS. EVALUATION METRICS INCLUDE ACCURACY (ACC., %), PRECISION (PREC., %), RECALL (REC., %), $F1$ -SCORE ($F1$, %), KAPPA, AND p -VALUE (p). ASTERISKS (*, **, ***) INDICATE RESDSNET'S SIGNIFICANT SUPERIORITY ($p < 0.05$, $p < 0.01$, $p < 0.001$)

Method	IV2a						IV2b						PhysioNet					
	Acc.	Prec.	Rec.	F1	Kappa	p	Acc.	Prec.	Rec.	F1	Kappa	p	Acc.	Prec.	Rec.	F1	Kappa	p
ShallowNet	63.04	63.09	63.04	62.68	0.5072	***	81.27	81.74	81.26	81.10	0.6253	*	41.35	43.00	41.61	38.34	0.2182	***
DeepNet	55.79	56.22	55.79	54.80	0.4105	***	81.46	82.06	81.46	81.10	0.6293	**	42.25	44.24	42.25	39.32	0.2288	***
EEGNet	60.92	61.53	60.92	60.77	0.4789	***	79.03	79.80	79.03	78.83	0.5806	0.06	51.34	52.58	51.58	50.18	0.3521	***
FBCNet	76.20	76.78	76.20	75.63	0.6826	**	80.40	80.80	80.40	80.24	0.6080	***	47.42	50.26	48.03	45.94	0.3017	***
EEGConformer	57.14	57.28	57.14	56.86	0.4285	***	81.09	81.73	81.09	80.86	0.6217	0.15	55.73	57.40	55.85	54.86	0.4098	***
FBMSNet	73.80	74.32	73.80	73.36	0.6507	**	80.35	80.46	80.35	80.30	0.6070	***	46.88	49.16	47.28	44.59	0.2924	***
IFNet	77.12	77.63	77.12	76.89	0.6950	*	81.21	81.71	81.21	81.10	0.6242	**	50.86	53.16	51.51	49.56	0.3475	***
TSFCNet	71.64	72.34	71.64	71.41	0.6219	*	80.31	80.68	80.31	80.01	0.6062	**	50.00	52.89	50.61	48.43	0.3362	***
ADFCNN	70.06	70.36	70.06	69.70	0.6008	***	83.36	84.32	83.36	83.16	0.6672	0.56	52.33	53.83	52.58	51.63	0.3651	***
EISATCFusion	63.81	64.26	63.81	63.51	0.5175	***	82.96	83.73	82.96	82.78	0.6592	0.38	54.20	55.44	54.39	53.45	0.3898	***
DoNet	69.33	69.70	69.33	68.33	0.5910	**	81.57	82.14	81.57	81.43	0.6313	***	53.72	56.44	54.09	52.51	0.3835	***
EEGSimpleConv	65.08	67.59	65.08	63.88	0.5345	***	80.21	80.44	80.21	80.16	0.6043	**	53.28	55.19	53.70	51.22	0.3788	***
CTNet	68.36	69.41	68.36	68.08	0.5782	**	83.45	83.62	83.45	83.43	0.6690	0.28	58.84	59.12	57.01	56.13	0.4522	***
MFRCNet	72.18	73.01	72.18	71.48	0.6291	***	81.75	82.22	81.75	81.64	0.6350	**	49.81	53.84	50.63	48.52	0.3346	***
FACTNet	71.18	71.63	71.18	70.27	0.6157	*	81.67	82.83	81.67	80.89	0.6334	*	50.34	51.86	50.74	48.94	0.3393	***
STGF	67.44	68.55	67.44	67.27	0.5658	**	83.04	83.59	83.04	82.89	0.6607	0.40	54.07	55.72	54.28	53.42	0.3884	***
DMSACNN	75.15	75.42	75.15	74.73	0.6687	***	80.07	80.75	80.07	79.92	0.6015	**	52.26	54.55	52.75	51.25	0.3653	***
SSTDPN	65.66	65.96	65.66	65.41	0.5422	**	81.04	82.03	81.04	80.78	0.6207	*	48.88	51.15	49.36	47.58	0.3205	***
TMSANet	68.87	68.11	68.87	67.12	0.5849	**	81.82	82.12	81.82	81.74	0.6363	*	52.64	54.99	52.81	50.61	0.3677	***
SNNforMI	68.09	69.09	68.09	66.88	0.5746	***	78.28	79.00	78.28	78.08	0.5657	**	54.46	58.27	55.30	52.16	0.3964	***
DBConformer	72.57	73.62	72.57	71.87	0.6343	*	82.91	83.15	82.91	82.77	0.6581	*	54.11	56.76	54.06	52.17	0.3868	***
ResDSNet	79.36	79.48	79.36	79.02	0.7248		84.95	85.15	84.95	84.87	0.6990		64.13	66.50	64.10	62.84	0.5205	

TABLE III

COMPARISON OF ACCURACY (% ACC.) AND KAPPA RESULTS BETWEEN DIFFERENT FREQUENCY BANDS USING RESDSNET ON THREE DATASETS

Dataset	IV2a		IV2b		PhysioNet	
Metric	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
4, 40	77.89	0.7052	82.70	0.6540	59.17	0.4565
4-20, 20-40	77.66	0.7022	83.08	0.6616	59.47	0.4585
4-8, 8-16, 16-40	75.85	0.6780	82.31	0.6461	56.62	0.4204
4-8, 8-16, 16-30, 30-40	75.39	0.6718	82.37	0.6474	56.60	0.4214
4-8, 8-12, 12-16, 16-20, 20-24, 24-28, 28-32, 32-36, 36-40	72.34	0.6312	81.94	0.6388	54.19	0.3878
16-40, 4-16	76.31	0.6842	82.96	0.6591	59.02	0.4545
36-40, 32-36, 28-32, 24-28, 20-24, 16-20, 12-16, 8-12, 4-8	72.99	0.6399	81.99	0.6398	55.22	0.4019
4-16, 16-40 (ResDSNet)	79.36	0.7248	84.95	0.6990	64.13	0.5205

attention mechanisms, such as EEGConformer [28], CTNet [29], TMSANet [60], and DBConformer [30], and as a CNN, ResDSNet incorporates structural bias presets, rendering it less susceptible to variations induced by different subjects' MI data. In comparison with models using convolutional attention mechanisms like DMSACNN [58], ResDSNet demonstrates more robust performance across different datasets. ResDSNet, utilizing a CNN structure, also maintains a significant performance advantage over alternative network structures such as graph networks STGF [54] and spiking neural networks SNNforMI [61]. Moreover, compared to MFRCNet [20], which also employs a residual mechanism, ResDSNet, incorporating its specifically designed residual layers, residual connection patterns, and convolutional kernel sizes, achieves further performance enhancements.

V. DISCUSSION

A. Different Bands

The proposed method, ResDSNet, employs different frequency bands, which can have varying impacts on the network's recognition performance. Therefore, we referenced

existing frequency band segmentation methods, including the single-band setting (4–40 Hz) used in EEGNet [18] and the filter bank methods used in FBCNet [36], which divides the band into (4–8, 8–12, 12–16, 16–20, 20–24, 24–28, 28–32, 32–36, and 36–40 Hz). We also explored the differences between (4–20 and 20–40 Hz) and the (4–16 and 16–40 Hz) used in the proposed ResDSNet. In addition, we introduced two frequency band settings mentioned in IFNet [19], including (4–8, 8–16, and 16–40 Hz) and (4–8, 8–16, 16–30, and 30–40 Hz). Finally, we considered the impact of starting the segmentation from higher frequencies on the network's performance, so we designed two reverse filter bank methods: (16–40 and 4–16 Hz) and (36–40, 32–36, 28–32, 24–28, 20–24, 16–20, 12–16, 8–12, and 4–8 Hz). The other settings of the network remain unchanged.

As shown in Table III, in the proposed network, the Acc. and Kappa exhibits a fluctuating downward trend across the three datasets. Compared with the (4–40 Hz) method, the nine-band division shows varying degrees of metric decline across different datasets, with more significant decreases observed on the IV2a and PhysioNet datasets. As distinct frequency bands of MI-EEG signals exhibit different frequency-domain

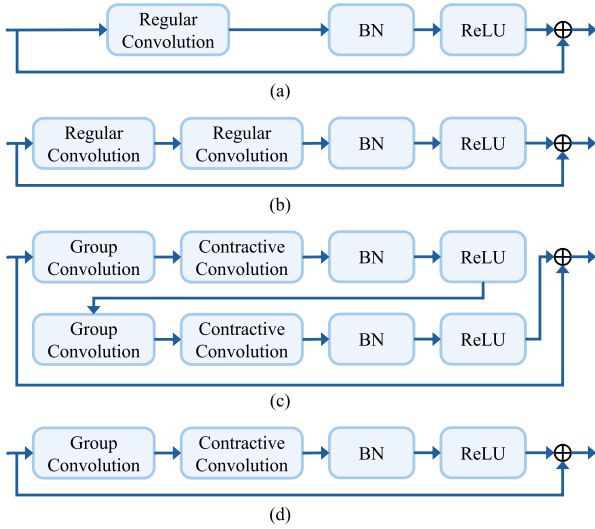


Fig. 2. Structure of the different baseline layers. (a) Residual RC layer. (b) Residual DRD layer. (c) Residual dual depthwise-separable mixed layer. (d) Residual depthwise-separable mixed layer used in the ResMIEEGNet.

TABLE IV

COMPARISON OF ACCURACY (% ACC.) AND KAPPA RESULTS BETWEEN DIFFERENT BASELINE LAYERS USING RESDSNET ON THREE DATASETS

Dataset	IV2a		IV2b		PhysioNet	
Metric	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
RC	<u>77.12</u>	0.6950	82.83	0.6565	<u>59.54</u>	0.4597
DRC	75.62	0.6749	82.94	0.6588	58.39	0.4446
DDSC	77.01	0.6934	<u>83.52</u>	0.6703	58.85	0.4509
DSC(ResDSNet)	79.36	0.7248	84.95	0.6990	64.13	0.5205

RC indicates the residual regular convolution layer, DRC indicates the residual dual regular convolution layer, DDSC indicates the residual dual depthwise separable convolution layer, DSC indicates the residual depthwise separable convolution layer of ResDSNet.

characteristics, a single frequency band in ResDSNet struggles to fully leverage these features. Limited by the impact of depthwise-separable convolution in the RDSM, the number of frequency segmentations must be consistent with the number of groups in the group convolution to achieve optimal performance; otherwise, the performance will decline or fail to meet expectations. We also observe that the results from the reversed filter bank yield similar outcomes to those without reversal. These experiments demonstrate the validity of the specific frequency band segmentation employed in our filter bank setting of the DS module.

B. Influence of Different Baseline Layers

Different baseline layers have various effects on network performance, but their specific quantitative impact on Acc. remains unclear. To explore this, we compare the effects of four types of baseline layers: an RC layer, a dual RC (DRC) layer, a dual depthwise-separable convolution (DDSC) layer, and a depthwise-separable convolution layer within our proposed method, ResDSNet. The structure of each of the different baseline layers is shown in Fig. 2.

The residual RC layer consists of a 1-D convolution layer, a normalization layer, and a ReLU activation layer in sequence, and is shown in Fig. 2(a). The residual DRC layer, shown

in Fig. 2(b), consists of two 1-D convolution layers, followed by a normalization layer and a ReLU activation layer. The residual DDSC layer, shown in Fig. 2(c), includes a depthwise-separable convolution layer, a normalization layer, a ReLU activation, another depthwise-separable convolution layer, and another normalization layer and ReLU activation, all in sequence. The remaining network layers are consistent with our proposed ResDSNet structure. The results of comparing these different baseline layers are shown in Table IV.

Table IV presents the impact of applying different baseline layers within ResDSNet on the performance of ResDSNet across the IV2a, the IV2b, and the PhysioNet datasets. We observe that using the residual RC layer with the residual mechanism achieves an Acc. of 77.12%. However, this is lower than that achieved with the residual depthwise-separable convolution layer with 79.36% Acc., which can be considered analogous to an inception module with a high number of branches, thereby enabling a broader network width [56]. In contrast, the performance of the double-layer result lags that of the single-layer result. The residual DRC layer reaches an Acc. of 75.62%, approximately 1.5% lower than the residual RC. This may be attributed to the limited branches in RC and the disruption of residual connections when using double layers. Similarly, the residual DDSC achieves only 77.01%, marking a 2.35% gap from the residual depthwise-separable convolution.

However, this gap varies across different datasets. In the IV2b dataset, the residual RC layer and the residual DRC layer yield similar results, whereas in PhysioNet, there is a 1.15% difference. This discrepancy may arise from variations in the number of channels and sampling rates used in different datasets. In the IV2b dataset, there are only three channels (C3, C4, and Cz), which are the primary feature channels for MI tasks. Under the same 250 Hz sampling rate, the gap in IV2a is significantly larger than that in the IV2b dataset. In the PhysioNet dataset, the difference between the residual dual depthwise-separable mixed layer and the RDSM layer is 5.28%, likely because more channels are better suited for the depthwise-separable structure, while increasing the number of layers leads to a rapid decline in performance.

C. Different Kernel Sizes on RDSM

We select different kernel sizes to evaluate the performance of RDSM in ResDSNet. Specifically, we test convolutional kernel sizes of 3, 7, 15, and 31. For instance, when selecting size 3, all convolutional kernels in the network are uniformly set to 3. These dimensions represent commonly used sizes in neural networks. In addition, we design a configuration where all kernel sizes start with 31, modifying the kernel sizes in Table I to ([31, 31], [15, 7], [7, 7], and [7, 3]). The results are presented in Table V.

As shown in Table V, the parameters we employ in ResDSNet demonstrate optimal performance. Our findings reveal that under identical kernel configurations, larger kernel sizes lead to faster performance degradation, particularly evident on the PhysioNet dataset with a 5% decrease. Similarly, when all convolutional kernels start with a size of 31, the performance also declines.

TABLE V

COMPARISON OF ACCURACY (% , ACC.) AND KAPPA RESULTS BETWEEN DIFFERENT KERNEL SIZES USING RESDSNET ON THREE DATASETS

Dataset	IV2a		IV2b		PhysioNet	
	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
all 3	74.46	0.6595	82.28	0.6456	61.45	0.4859
all 7	74.88	0.6651	82.72	0.6545	60.01	0.4654
all 15	73.03	0.6404	82.51	0.6502	57.77	0.4368
all 31	72.80	0.6373	81.28	0.6257	53.85	0.3831
begin 31	74.15	0.6554	82.75	0.6551	56.60	0.4209
begin 15 (ResDSNet)	79.36	0.7248	84.95	0.6990	64.13	0.5205

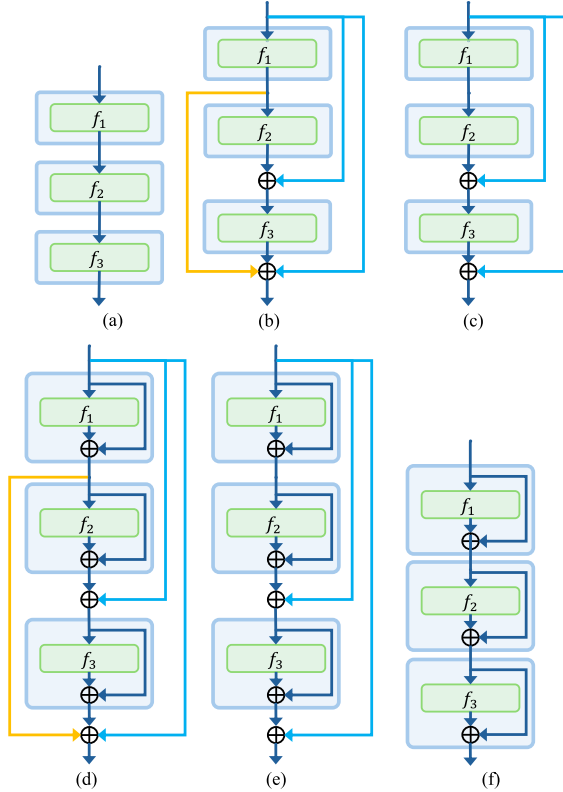


Fig. 3. Structure of different residual connection mechanisms. (a) Nonresidual. (b) Routing without IR. (c) Bypass without IR. (d) RR. (e) BR. (f) IR.

D. Influence of Different Residual Connections

The use of residual connections significantly enhances network depth and improves recognition efficiency. We evaluate six types of residual connection structures to assess their impact on network performance. Specifically, we evaluate the effect of nonresidual, interblock residual (IR), routing without IR (RIR), routing residual (RR), bypass without IR (BIR), and bypass residual (BR) connection structures, as shown in Fig. 3.

The nonresidual connection structure has no residual connections between any layers or blocks in the network. The IR connection, as we propose in ResDSNet, employs a residual mechanism with identity mapping limited to intrablock connections. In the RIR structure, residual connections are omitted within blocks, and blocks instead use DenseNet-style dense connections, where each block receives features from all preceding blocks and forward them as input features to subsequent blocks. The RR structure combines IR connections with DenseNet-style dense connections between blocks. In the BIR

TABLE VI

COMPARISON OF ACCURACY (% , ACC.) AND KAPPA RESULTS BETWEEN RESIDUAL CONNECTIONS USING RESDSNET ON THREE DATASETS

Dataset	IV2a		IV2b		PhysioNet	
	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
non-residual	67.32	0.5643	81.37	0.6275	48.46	0.3118
RIR	75.89	0.6785	82.66	0.6532	58.45	0.4454
BIR	72.22	0.6296	81.88	0.6375	47.97	0.3057
RR	77.24	0.6965	82.67	0.6534	50.42	0.3387
BR	77.58	0.7011	82.97	0.6593	49.10	0.3221
IR(ResDSNet)	79.36	0.7248	84.95	0.6990	64.13	0.5205

IR indicates inter-block residual connections of ResDSNet, RIR indicates routing without inter-block residual connections, RR indicates routing residual connections, BIR indicates bypass without inter-block residual connections, and BR indicates bypass residual connections.

structure, residual connections are excluded within blocks, and all blocks are densely connected. The BR connection structure introduces a bypass for the initial input to serve as input for all subsequent layers, creating an interblock bypass structure, while also applying residual connections within each block.

We evaluate six types of connection structures on the IV2a, IV2b, and PhysioNet datasets, as shown in Table VI. The IR connection structure in ResDSNet consistently achieves more stable Acc. than the other five connection types, reaching 79.36% Acc. and 0.7248 Kappa on the IV2a dataset, 84.95% Acc. and 0.699 Kappa on the IV2b dataset, and 64.13% Acc. and 0.5205 Kappa on the PhysioNet dataset. This improvement is due to the IR structure, which enables each layer to establish identity mappings, allowing feature maps processed in the initial channel-expansion layers to propagate through these connections. This design facilitates the seamless transmission of feature maps from the channel-expansion layer to the final RDSM layer, supporting efficient training across the network. Conversely, the network without any residual connections shows the lowest Acc., with 67.32% Acc. achieved on the IV2a dataset, 81.37% on the IV2a dataset, and 48.46% on the PhysioNet dataset.

We employ an unraveled view [66] to reconstruct different residual connections to demonstrate the differences between various mechanisms, as shown in Fig. 4. From this, it can be observed that the structural diagrams for the nonresidual, RIR, and BIR are similar to those before reconstruction. When IR connections are added, the RR, BR, and IR significantly change their structures. The IR can be viewed as a collection of 2^n independent paths (where n is the number of residual modules). The depth of each path l , representing the number of modules traversed, follows a binomial distribution, $l \sim \mathcal{B}(n, p = 0.5)$. The output is the ensemble of all paths

$$y_{\text{out}} = \sum_{b \in \{0,1\}^n} P_b(x) \quad (12)$$

where $P_b(x)$ denotes the path selected by the binary mask b ($b_i = 1$ indicates the i th module is enabled). The distribution of paths in different structures is as follows:

In Table VII, $p_s = l_0/(l_0 + l_1 + l_2 + l_3)$ represents the proportion of paths traversing only the initial layer (i.e., contributing directly with no modules). $p_m = 1/2 \times (l_1 + l_2)/(l_0 + l_1 + l_2 + l_3)$ indicates the intermediate-layer dominance, computed as half the proportion of paths traversing one or

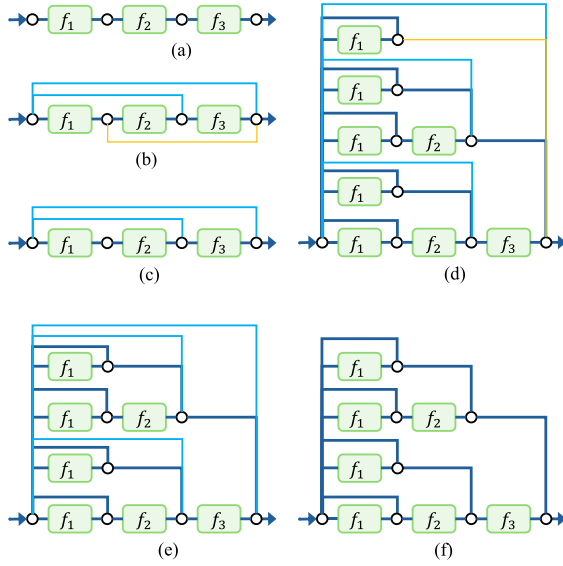


Fig. 4. Reconstruct of different residual connection using unraveled view [66]. (a) Nonresidual. (b) Routing without IR. (c) Bypass without IR. (d) RR. (e) BR. (f) IR.

TABLE VII
PATH DISTRIBUTION AND LAYER DOMINANCE ACROSS
RESIDUAL STRUCTURES

	l_0	l_1	l_2	l_3	l_{total}	$p_s(\%)$	$p_m(\%)$	$p_d(\%)$
Non-residual	0	0	0	1	1	0.00	0.00	100.00
RIR	1	2	0	1	4	25.00	25.00	25.00
BIR	1	1	0	1	3	33.33	16.67	33.33
RR	4	5	3	1	13	30.77	30.77	7.69
BR	3	4	3	1	11	27.27	31.82	9.09
IR(ResDSNet)	1	3	3	1	8	12.50	37.50	12.50

two modules. $p_d = l_3 / (l_0 + l_1 + l_2 + l_3)$ represents the deep-layer dominance, i.e., the proportion of paths traversing all three modules.

From Table VII, it can be observed that when intermediate-layer paths dominate the structure (i.e., p_m is high), such as in the proposed IR connections of ResDSNet, better performance is achieved. This has been validated across multiple datasets. Conversely, nonresidual structures yield poor performance as only deep-layer paths exist. Shallow layers extract significant features, and multilayer traversal can prevent some of these features from reaching the output, degrading performance [19].

Comparing RIR and BIR, which differ only by the presence of the yellow connection line that connects shallow features directly to the output, in Fig. 3(b), BIR has a lower p_m value. The inclusion of this connection in RIR enables shallow features to reach the final output and integrate with deeper features, resulting in a substantial performance gain for RIR over BIR on the PhysioNet dataset. While both RR and BR employ dense residual connections, they exhibit similar p_s and p_m values. This lack of intermediate-layer dominance (p_m not being predominant) likely explains their inferior performance compared to IR. Furthermore, we find significant performance differences between the RIR/BIR and the RR/BR across datasets. We speculate that this is because datasets IV2a and IV2b have fewer acquisition channels; the smaller gap in IV2b may be due to it being a binary classification task

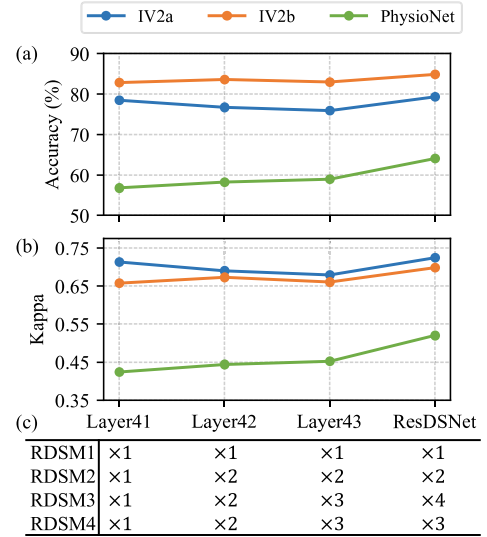


Fig. 5. Comparison of (a) Acc. (%), and (b) Kappa of our proposed network ResDSNet using (c) different depths (numbers of RDSM) on the IV2a, IV2b, and PhysioNet datasets.

with only three channels. In contrast, the PhysioNet dataset features a larger number of channels (64), a four-class task, and a larger subject number. These characteristics likely amplify the performance differences between structures in large-scale, diverse populations.

E. Influence of Different Depth on ResDSNet

To further evaluate the impact of adding or removing layers in ResDSNet with residual structures, we gradually increase the number of RDSM layers. In the “Layer41” model, each of the four RDSM layers has only one layer. In the “Layer42” model, all layers except the first are increased by one layer compared to “Layer41.” In the “Layer43” model, the first and second layers remain as in “Layer42,” while the other layers are each increased by one. This incremental addition continues until it aligns with our proposed method.

In Fig. 5, we can see that network performance gradually improves as the number of layers increases. When looking at the IV2a dataset, Layer41 achieves an Acc. of 83.42%. As network depth increases, performance first declines and then rises. Ultimately, our proposed method results in an Acc. of 78.89% and the Kappa of 0.7135. This effect occurs because subjects in the IV2a dataset tend to achieve higher Acc. with lightweight, shallow networks [19]. When the number of layers increases without sufficient network parameters, performance initially declines. However, with an increase in parameters and the implementation of residual mechanisms, the network can effectively construct robust hierarchical representations. The residual mechanism effectively prevents information loss in deeper layers, allowing the network to achieve higher performance. When looking at the IV2b dataset, Layer41 reaches an Acc. of 82.89%, whereas ResDSNet achieves a 2.06% improvement, reaching 84.95%. In the PhysioNet dataset, Acc. increases from 56.81% to 64.13%, demonstrating that network performance gradually improves under the influence of residual mechanisms. In addition, the plug-and-play layer structure offers greater flexibility for network expansion.

TABLE VIII
COMPARISON OF THE NUMBER OF PARAMETERS (K), TRAINING TIME (S), AND TEST TIME (MS) OF NETWORKS ON THREE DATASETS

Metrics	IV2a				IV2b				PhysioNet			
	Layer41	Layer42	Layer43	ResDSNet	Layer41	Layer42	Layer43	ResDSNet	Layer41	Layer42	Layer43	ResDSNet
parameters (K)	84.54	111.06	126.18	137.34	81.96	108.48	123.60	134.76	89.58	116.10	131.22	142.38
training time (s)	77	86	91	96	43	57	66	72	19	20	21	23
inference time (ms)	13.75	19.98	23.56	27.99	13.31	17.05	19.66	23.89	17.36	20.09	27.31	31.77

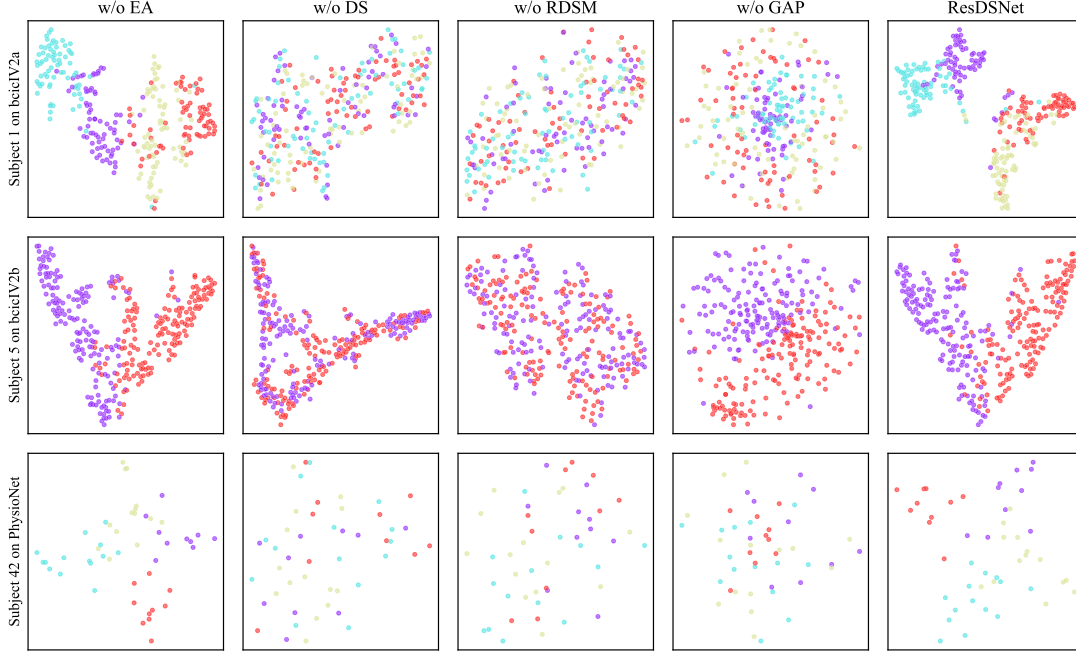


Fig. 6. t-SNE visualization illustrates trials classification of different subjects on three datasets.

We evaluate three key metrics across three datasets. The number of parameters reflects model size in thousands (K) and spatial complexity through trainable weights. Training time (seconds, s) represents the time required to train on the training set for an individual model. Inference time (milliseconds, ms) denotes the per-trial inference duration for the different depth models on the test set. As shown in Table VIII, the parameter count gradually increases with progressive layer-depth augmentation. The per-trial inference time remains under 32 ms, fully meeting conventional online experimentation requirements.

F. Different Pooling Layer

To evaluate the impact of different pooling layers on ResDSNet, we selected the commonly used average pooling layer, max pooling layer, the variance layer employed in FBCNet [36], and the GAP used in ResDSNet for assessment.

In Table IX, we found that using GAP can integrate spatial information between features [75], achieving optimal results: 79.36% Acc. and 0.7248 Kappa in the IV2a dataset, 84.95% Acc. and 0.699 Kappa in the IV2b dataset, and 64.13% Acc. and 0.5205 Kappa in the PhysioNet dataset. Similar to the results obtained in [36], the average pooling layer significantly outperforms the max pooling layer. Although the

TABLE IX
COMPARISON OF ACCURACY (% ACC.) AND KAPPA RESULTS BETWEEN DIFFERENT POOLING LAYERS USING RESDSNET ON THREE DATASETS

Dataset	IV2a		IV2b		PhysioNet	
	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
AvgPool	74.96	0.6662	83.32	0.6664	56.28	0.4159
MaxPool	60.96	0.4794	79.26	0.5852	52.70	0.3700
Var	69.10	0.5880	81.40	0.6280	55.07	0.4026
GAP(ResDSNet)	79.36	0.7248	84.95	0.6990	64.13	0.5205

variance layer performs better than the max pooling layer, it is weaker than both the average pooling layer and the proposed global pooling layer. We speculate that the batch normalization layer in RDSM normalizes the input features, reducing the occurrence of outliers, which diminishes the variance layer's ability to handle outliers and indirectly leads to a decline in performance.

G. Ablation Experiment

We validate our proposed methods and structures by conducting ablation experiments on the data preprocessing modules, the RDSM layer, and the GAP to evaluate their influence on network performance. The data preprocessing techniques include EA and DS. The results are shown in Table X.

TABLE X

COMPARISONS OF THE ABLATION STUDY WITH THE PROPOSED METHOD, RESDSNET, FOR THE METHOD'S FOUR MODULES, INCLUDING EUCLIDEAN SPACE DATA ALIGNMENT (EA), DS, THE RDSM LAYER, AND THE GAP, ON THREE DATASETS. THE *, **, AND *** INDICATE THAT THE ACCURACIES OF LIGHTMIEEGNET ARE SIGNIFICANTLY HIGHER THAN THE COMPARED METHODS WITH $p < 0.05$, $p < 0.01$, AND $p < 0.001$, RESPECTIVELY

Module	IV2a Dataset						IV2b Dataset						PhysioNet Dataset					
	Acc.	Prec.	Rec.	F1	Kappa	p	Acc.	Prec.	Rec.	F1	Kappa	p	Acc.	Prec.	Rec.	F1	Kappa	p
w/o EA	76.43	76.74	76.43	76.01	0.6857	***	83.18	83.42	83.18	83.07	0.6636	**	58.90	62.26	59.16	57.26	0.4521	***
w/o DS	76.89	76.99	76.89	76.54	0.6919	**	83.30	83.38	83.30	83.26	0.6660	***	58.94	61.51	59.17	57.25	0.4524	***
w/o RDSM	30.13	29.24	30.13	28.16	0.0684	***	53.16	53.17	53.16	47.17	0.0632	***	32.31	31.31	32.63	28.91	0.0992	***
w/o GAP	35.84	36.24	35.84	35.43	0.1445	***	75.94	76.29	75.94	75.86	0.5188	**	34.72	32.40	34.47	30.03	0.1260	***
ResDSNet	79.36	79.48	79.36	79.02	0.7248		84.95	85.15	84.95	84.87	0.6990		64.13	66.50	64.10	62.84	0.5205	

In Table X, we find that all proposed modules impact network performance. Removing the RDSM layer and the GAP layer significantly degrades the performance on the IV2a and PhysioNet datasets. For the IV2b dataset, the final GAP layer shows a smaller performance decline, while the RDSM layer still exhibits a substantial drop. This occurs primarily because unprocessed features tend to produce random predictions, converging to a 25% probability for four-class classification in the IV2a and PhysioNet datasets. Regarding the GAP layer, the three datasets have different numbers of channels, and its role is to capture global features across all channels. When the number of channels is small, the global spatial features become limited, which leads to a smaller performance decline.

We use t-SNE to visualize the ablation results of different modules, as shown in Fig. 6. As observed, ResDSNet exhibits the most distinct clustering, whereas without RDSM and GAP, features of different categories remain highly entangled. The addition of EA and DS makes the intraclass features more clustered, as seen in the comparison between the ResDSNet plot and the w/o EA plot for the IV2a dataset.

VI. CONCLUSION

In this article, we propose a ResDSNet to address the longstanding challenge of training deep structures on low signal-to-noise ratio MI-EEG data. By systematically resolving the structural incompatibilities that traditionally cause deep networks to overfit on limited EEG data, ResDSNet successfully breaks the depth barrier in MI classification. Extensive evaluations on real EEG data from 127 human subjects on three public benchmark datasets—including IV2a, IV2b, and PhysioNet—demonstrate that our deep network architecture consistently outperforms 21 SOTA models, achieving statistically significant improvements.

Our study reveals that effectively deepening MI classification networks requires the simultaneous alignment of three critical factors. First, through an unravelled-view path analysis, we demonstrate that residual mechanisms in MI networks achieve optimal performance when the layer distribution across paths approximates a binomial distribution, whereas overly dense or unsuitable connections degrade performance. Second, depthwise-separable convolution serves as a superior residual baseline layer compared to RC, functioning efficiently to filter noise while retaining task-relevant features. Finally, integrating an optimized data preprocessing module—specifically combining EA with DS—is essential to

reduce input redundancy and maintain compatibility with the subsequent deep residual modules.

Overall, ResDSNet not only provides a robust and scalable structure for MI-EEG decoding but also offers generalizable architectural design criteria for future deep BCI models. In future work, we plan to extend this deep residual structure to different BCI paradigms.

REFERENCES

- [1] D. Wu, X. Jiang, and R. Peng, "Transfer learning for motor imagery based brain-computer interfaces: A tutorial," *Neural Netw.*, vol. 153, pp. 235–253, Sep. 2022.
- [2] J. Jin, R. Xu, I. Daly, X. Zhao, X. Wang, and A. Cichocki, "MOCNN: A multiscale deep convolutional neural network for ERP-based brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 54, no. 9, pp. 5565–5576, Sep. 2024.
- [3] J. Jin, Z. Wang, R. Xu, C. Liu, X. Wang, and A. Cichocki, "Robust similarity measurement based on a novel time filter for SSVEPs detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4096–4105, Aug. 2023.
- [4] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.
- [5] Y. Yu et al., "Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and P300 potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2516–2526, Dec. 2017.
- [6] X. Xu, J. Chen, Q. He, Y. Wu, and Y. Zhou, "EEG emotion recognition with uncertainty-aware contrastive learning and frequency-aware self-attention," *IEEE Trans. Cybern.*, vol. 56, no. 4, pp. 1845–1858, Apr. 2026.
- [7] J. Phyto, W. Ko, E. Jeon, and H.-I. Suk, "TransSleep: Transitioning-aware attention-based deep neural network for sleep staging," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4500–4510, Jul. 2023.
- [8] C. L. Philip Chen, B. Chen, and T. Zhang, "AdamGraph: Adaptive attention-modulated graph network for EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2038–2051, May 2025.
- [9] X. Yu and X. Yu, "Brain-controlled wheeled mobile robots: A framework combining probabilistic brain-computer interface and model predictive control," *IEEE Trans. Cybern.*, vol. 55, no. 9, pp. 4311–4321, Sep. 2025.
- [10] J.-H. Cho, J.-H. Jeong, and S.-W. Lee, "NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13279–13292, Dec. 2022.
- [11] S. Pancholi, A. Giri, A. Jain, L. Kumar, and S. Roy, "Source aware deep learning framework for hand kinematic reconstruction using EEG signal," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4094–4106, Jul. 2023.
- [12] M.-K. Kim, H.-B. Shin, J.-H. Cho, and S.-W. Lee, "Developing brain-based bare-handed human-machine interaction via on-skin input," *IEEE Trans. Cybern.*, vol. 55, no. 4, pp. 1554–1567, Apr. 2025.
- [13] K. Barmpas, Y. Panagakis, S. Bakas, D. A. Adamos, N. Laskaris, and S. Zafeiriou, "Improving generalization of CNN-based motor-imagery EEG decoders via dynamic convolutions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1997–2005, 2023.

- [14] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.
- [15] B. Wang et al., "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 5008–5020, Oct. 2021.
- [16] F. Qi et al., "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.
- [17] J. Fumana-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernández, J. A. Sanz, and H. Bustince, "Motor-imagery-based brain-computer interface using signal derivation and aggregation functions," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7944–7955, Aug. 2022.
- [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," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [19] J. Wang, L. Yao, and Y. Wang, "IFNet: An interactive frequency convolutional neural network for enhancing motor imagery decoding from EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1900–1911, 2023.
- [20] X. Li, Z. Yang, X. Tu, J. Wang, and J. Huang, "MFRC-Net: Multi-scale feature residual convolutional neural network for motor imagery decoding," *IEEE J. Biomed. Health Informat.*, vol. 29, no. 1, pp. 224–234, Jan. 2025.
- [21] B. Lu, F. Wang, J. Chen, G. Wen, C. Hua, and R. Fu, "Dynamic hierarchical convolutional attention network for recognizing motor imagery intention," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2202–2212, May 2025.
- [22] W. Chen et al., "Enhancing the capability and accuracy of motor imagery classification: A deep neural network-powered multifaceted strategy model," *IEEE Trans. Cybern.*, early access, Apr. 6, 2026, doi: [10.1109/TCYB.2026.3678659](https://doi.org/10.1109/TCYB.2026.3678659).
- [23] W. Chen et al., "Multiscale pooling spatial-temporal attention network: Elevating cross session and small sample decoding in motor imagery brain-computer interfaces," *IEEE Trans. Syst.*, vol. 56, no. 4, pp. 2225–2239, Apr. 2026.
- [24] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*.
- [25] P. Arpaia, A. Esposito, A. Natalizio, and M. Parvis, "How to successfully classify EEG in motor imagery BCI: A metrological analysis of the state of the art," *J. Neural Eng.*, vol. 19, no. 3, Jun. 2022, Art. no. 031002.
- [26] R. Liu et al., "Dynamic graph representation learning for spatio-temporal neuroimaging analysis," *IEEE Trans. Cybern.*, vol. 55, no. 3, pp. 1121–1134, Mar. 2025.
- [27] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–18.
- [28] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2023.
- [29] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng, "CTNet: A convolutional transformer network for EEG-based motor imagery classification," *Sci. Rep.*, vol. 14, no. 1, p. 20237, Aug. 2024.
- [30] Z. Wang, H. Wang, T. Jia, X. He, S. Li, and D. Wu, "DBConformer: Dual-branch convolutional transformer for EEG decoding," 2025, *arXiv:2506.21140*.
- [31] D.-H. Shin et al., "MARS: Multiagent reinforcement learning for spatial-spectral and temporal feature selection in EEG-based BCI," *IEEE Trans. Syst.*, vol. 54, no. 5, pp. 3084–3096, May 2024.
- [32] S. Lin, C. Fan, D. Han, Z. Jia, Y. Peng, and S. Kwong, "HATNet: EEG-based hybrid attention transfer learning network for train driver state detection," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2437–2450, May 2025.
- [33] W. Ko, S. Jeong, S.-K. Song, and H.-I. Suk, "EEG-oriented self-supervised learning with triple information pathways network," *IEEE Trans. Cybern.*, vol. 54, no. 11, pp. 6495–6508, Nov. 2024.
- [34] R. T. Schirmer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [35] Y. El Ouahidi, V. Gripon, B. Pasdeloup, G. Bouallegue, N. Farrugia, and G. Lioi, "A strong and simple deep learning baseline for BCI motor imagery decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3338–3347, 2024.
- [36] R. Mane et al., "FBCNet: A multi-view convolutional neural network for brain-computer interface," 2021, *arXiv:2104.01233*.
- [37] K. Liu, M. Yang, Z. Yu, G. Wang, and W. Wu, "FBMSNet: A filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 2, pp. 436–445, Feb. 2023.
- [38] H. Zhi, Z. Yu, T. Yu, Z. Gu, and J. Yang, "A multi-domain convolutional neural network for EEG-based motor imagery decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3988–3998, 2023.
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 60, 2017, pp. 84–90.
- [40] A. Radford et al., "Improving language understanding by generative pre-training," OpenAI, San Francisco, CA, USA, Tech. Rep., 2018.
- [41] P. Dreyer, A. Roc, L. Pillette, S. Rimbart, and F. Lotte, "A large EEG database with users' profile information for motor imagery brain-computer interface research," *Sci. Data*, vol. 10, no. 1, p. 580, Sep. 2023.
- [42] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jan. 2012.
- [43] S. Liang, S. Kuang, D. Wang, Z. Yuan, H. Zhang, and L. Sun, "An auxiliary synthesis framework for enhancing EEG-based classification with limited data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2120–2131, 2023.
- [44] C.-H. Lee, H. Kim, B. C. Yoon, and D.-J. Kim, "Toward foundational model for sleep analysis using a multimodal hybrid-self-supervised learning framework," *IEEE Trans. Cybern.*, vol. 55, no. 12, pp. 5619–5632, Dec. 2025.
- [45] Z. Xie, "mHC: Manifold-constrained hyper-connections," 2025, *arXiv:2512.24880*.
- [46] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," 2021, *arxiv: 2106.04554*.
- [47] W.-B. Jiang, Y. Wang, B. Lu, and D. Li, "NeuroLM: A universal multi-task foundation model for bridging the gap between language and EEG signals," in *Proc. 13th Int. Conf. Learn. Represent.*, 2024, pp. 1–18. [Online]. Available: <https://openreview.net/forum?id=8b4add8b0aa8749d80a34ca5d941c355>
- [48] Y. Yao, H. Wang, L. Chen, Y. Peng, and J. Luo, "Foundation models for EEG decoding: Current progress and prospective research," *J. Neural Eng.*, vol. 22, no. 6, Dec. 2025, Art. no. 061002.
- [49] H. He and D. Wu, "Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.
- [50] Z. Chen, R. Yang, M. Huang, F. Li, G. Lu, and Z. Wang, "EEGProgress: A fast and lightweight progressive convolution architecture for EEG classification," *Comput. Biol. Med.*, vol. 169, Feb. 2024, Art. no. 107901.
- [51] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 14681–14722, 2021.
- [52] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, "A decoding scheme for incomplete motor imagery EEG with deep belief network," *Frontiers Neurosci.*, vol. 12, p. 42, Sep. 2018.
- [53] S. Ke, B. Yang, Y. Qin, F. Rong, J. Zhang, and Y. Zheng, "FACT-net: A frequency adapter CNN with temporal-periodicity inception for fast and accurate MI-EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 4131–4142, 2024.
- [54] X. Wang, K. Zhao, E. Shi, S. Yu, G. Chen, and S. Zhang, "ST-GF: Graph-based fusion of spatial and temporal features for EEG motor imagery decoding," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2024, pp. 3811–3816.
- [55] C. Han, C. Liu, J. Wang, Y. Wang, C. Cai, and D. Qian, "A spatial-spectral and temporal dual prototype network for motor imagery brain-computer interface," *Knowledge-Based Syst.*, vol. 315, Apr. 2025, Art. no. 113315.
- [56] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
- [57] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [58] K. Liu et al., "DMSACNN: Deep multiscale attentional convolutional neural network for EEG-based motor decoding," *IEEE J. Biomed. Health Informat.*, vol. 29, no. 7, pp. 4884–4896, Jul. 2025.

- [59] G. Liang, D. Cao, J. Wang, Z. Zhang, and Y. Wu, "EISATC-Fusion: Inception self-attention temporal convolutional network fusion for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 1535–1545, 2024.
- [60] Q. Zhao and W. Zhu, "TMSA-Net: A novel attention mechanism for improved motor imagery EEG signal processing," *Biomed. Signal Process. Control*, vol. 102, Jan. 2025, Art. no. 107189.
- [61] H. Zhang, H. Wang, J. An, S. Zheng, and D. Wu, "A lightweight spiking neural network for EEG-based motor imagery classification," *Neural Netw.*, vol. 191, Nov. 2025, Art. no. 107741.
- [62] Y. He, H. Li, W. Liu, L. Ma, G. Wang, and C. Xu, "EEGPT: Pretrained transformer for universal and reliable representation of EEG signals," in *Proc. Adv. Neural Inf. Process. Syst.*, 2024, pp. 39249–39280.
- [63] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu, "Large brain model for learning generic representations with tremendous EEG data in BCI," 2024, *arXiv:2405.18765*.
- [64] C.-S. Chen, Y.-J. Chen, and A. H.-W. Tsai, "Large cognition model: Towards pretrained EEG foundation model," 2025, *arXiv:2502.17464*.
- [65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [66] A. Veit, M. J. Wilber, and S. J. Belongie, "Residual networks are exponential ensembles of relatively shallow networks," 2016, *arxiv:1605.06431*.
- [67] H. Cai, "CETRNet: Convolutional-EMA temporal residual network for motor imagery decoding," *Academic J. Comput. Inf. Sci.*, vol. 8, no. 7, pp. 1–16, 2025.
- [68] H. Mo, "A spatiotemporal dual-branch network based on channel attention and residual connection in motor imagery EEG classification," *J. Comput. Sci. Electr. Eng.*, vol. 8, no. 2, pp. 55–62, 2026.
- [69] M. Li, J. Li, X. Zheng, J. Ge, and G. Xu, "MSHNet: A multi-scale residual network with hybrid attention for motor imagery EEG decoding," *Cognit. Neurodynamics*, vol. 18, no. 6, pp. 3463–3476, Dec. 2024.
- [70] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [71] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain–computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
- [72] T. Wei et al., "ADFCNN: Attention-based dual-scale fusion convolutional neural network for motor imagery brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 154–165, 2023.
- [73] X. Niu, N. Lu, R. Yan, and H. Luo, "Model and data dual-driven double-point observation network for ultra-short MI EEG classification," *IEEE J. Biomed. health Informat.*, vol. 28, no. 6, pp. 3434–3445, Apr. 2024.
- [74] T. Liang, X. Yu, X. Liu, H. Wang, X. Liu, and B. Dong, "EEG-CDILNet: A lightweight and accurate CNN network using circular dilated convolution for motor imagery classification," *J. Neural Eng.*, vol. 20, no. 4, 2023, Art. no. 046031.
- [75] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.