Qu et al. Brain Informatics (2026) 13:33
https://doi.org/10.1186/s40708-026-00321-1

Brain Informatics

RESEARCH Open Access

EEG-DBNet: a dual-branch framework for temporal-spectral representation learning of motor imagery electroencephalography

Youxi Qu[1], Xicheng Lou[2*], Hongying Meng[3], Zhangyong Li[1], Jianlin Wang[1], Kunpeng Mao[1] and Xinwei Li[1*]

Abstract

Purpose Motor imagery electroencephalography (MI-EEG) decoding remains challenging due to low signal-to-noise ratio and complex temporal-spectral characteristics. This study aims to develop a robust deep learning framework for effective EEG representation learning.

Methods We propose EEG-DBNet, a dual-branch neural network that jointly models temporal dynamics and spectral representations of EEG signals. The model integrates local and global convolutional modules to enable multi-scale feature extraction, complementing the dual-branch design for multi-dimensional temporal-spectral representation learning. To validate robustness, experiments are conducted on two public datasets as well as a self-collected MI-EEG dataset acquired under controlled laboratory conditions.

Results Experimental results show that EEG-DBNet achieves the best average performance on the two public benchmark datasets, BCI Competition IV-2a and IV-2b. On the self-collected CQUPT dataset, EEG-DBNet obtains competitive performance compared with representative baseline methods, suggesting its potential applicability to laboratory-acquired MI-EEG decoding. These results indicate that the proposed temporal-spectral dual-branch design is effective, while further validation on larger self-collected datasets is still needed.

Conclusion The proposed EEG-DBNet provides an effective solution for MI-EEG decoding with improved robustness. The inclusion of multiple datasets, particularly laboratory-acquired self-collected data, highlights its potential for practical brain-computer interface applications.

Keywords Electroencephalogram (EEG), Motor imagery (MI), Brain-computer interfaces (BCIs), Neural networks

*Correspondence:
Xicheng Lou
louxicheng@ctbu.edu.cn
Xinwei Li
lixinwei@cqupt.edu.cn
[1]Chongqing Engineering Research Center of Medical Electronics and Information Technology, Chongqing University of Posts and Telecommunications, 2 Chongwen Rd, Chongqing 400065, China
[2]School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, 2 Chongwen Rd, Chongqing 400065, China
[3]Department of Electronic and Electrical Engineering, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK

1 Introduction

Brain-computer interfaces (BCIs) establish a direct communication pathway between the human brain and external devices by decoding neural activity into control signals, offering a promising solution for individuals with motor impairments. Among various BCI paradigms, motor imagery electroencephalography (MI-EEG) has attracted considerable attention due to its non-invasive nature, relatively low cost, and broad applicability in rehabilitation and assistive technologies [1]. By imagining specific motor actions without actual movement, users can modulate brain activity patterns that can be detected and translated into commands, enabling applications such as robotic control, neurorehabilitation, and communication systems [2]. Despite these advantages, the practical deployment of MI-based BCIs remains challenging due to the inherent variability of EEG signals and the complexity of accurately decoding motor-related neural patterns [3].

Early studies on MI-EEG decoding primarily relied on traditional machine learning frameworks, which typically consist of three key stages: signal preprocessing, feature extraction, and classification. In particular, spatial filtering techniques such as common spatial pattern (CSP) have been widely adopted to enhance the discriminability between different motor imagery tasks by projecting EEG signals into task-relevant subspaces [4]. To further reduce data dimensionality and mitigate noise interference, feature reduction methods such as principal component analysis (PCA) have also been incorporated into the processing pipeline [5]. Subsequently, conventional classifiers, including support vector machines (SVMs), are employed to perform the final classification based on handcrafted features [6].

Despite their effectiveness, these approaches heavily depend on manual feature engineering and prior domain knowledge, making them sensitive to the intrinsic non-stationarity and low signal-to-noise ratio of EEG signals. As a result, their performance often degrades significantly in cross-session or cross-subject scenarios, limiting their robustness and practical applicability.

With the rapid development of deep learning, convolutional neural networks (CNNs) have become a mainstream solution for MI-EEG decoding due to their capability of automatic feature extraction and end-to-end learning. Compared with traditional approaches, CNN-based models can directly learn hierarchical representations from raw EEG signals, effectively capturing spatial and temporal patterns without relying on handcrafted features [7]. Representative studies have introduced compact architectures and advanced convolutional designs to improve feature representation efficiency and adapt to the characteristics of EEG signals [8–11].

Despite their success, CNN-based models may still be limited in capturing long-range temporal dependencies and global contextual information in EEG signals. To address this limitation, recent studies have explored more advanced deep learning architectures with enhanced sequence modeling capabilities. In particular, attention-based models, such as Transformer architectures, have been introduced to model global dependencies through self-attention mechanisms, enabling flexible interactions across temporal dimensions and improving the representation of complex neural dynamics [12, 13]. In addition, graph-based and multi-scale Transformer variants have been investigated to model spatial dependencies across EEG channels and hierarchical temporal features more effectively [14, 15]. More recently, emerging sequence modeling frameworks, such as state-space-based approaches, have shown promising potential as efficient alternatives for long-sequence modeling, offering strong capability in capturing long-range dependencies with improved computational efficiency [16].

However, existing approaches still suffer from several limitations in modeling MI-EEG signals. Most models focus on single-domain feature extraction, failing to jointly capture the intrinsic coupling between temporal and spectral representations, while conventional spectral transformations may disrupt feature continuity [17, 18]. Although multi-branch architectures improve feature diversity, they often lack a unified framework for multi-dimensional representation learning. In addition, current models typically rely on fixed-scale operations, limiting their ability to capture the multi-scale characteristics of MI-related neural patterns.

To overcome these limitations, we propose EEG-DBNet, a dual-branch deep network that performs multi-scale and multi-dimensional decoding for MI-EEG signals. Specifically, the proposed model learns spectral representations directly from temporal sequences while preserving feature continuity, and jointly models temporal and spectral information within a unified framework. By integrating local and global convolutional modules, EEG-DBNet effectively captures both fine-grained patterns and long-range dependencies, enabling more comprehensive representation of MI-related neural dynamics. Extensive experiments on two public datasets and a self-collected dataset demonstrate that the proposed method achieves superior performance on public benchmark datasets and competitive performance on the self-collected CQUPT dataset. Based on the above motivation and methodological design, the main contributions of this paper are summarized as follows:

1. We propose EEG-DBNet, a dual-branch network that performs parallel temporal and spectral

decoding, enabling unified multi-dimensional feature learning with multi-scale granularity.
2. We conduct comprehensive evaluations on two public MI-EEG datasets, where EEG-DBNet achieves state-of-the-art performance and consistently outperforms existing methods.
3. We design and conduct a motor imagery EEG acquisition experiment to construct a self-collected dataset. On this dataset, the proposed method achieves competitive performance, providing preliminary evidence of its applicability to laboratory-acquired MI-EEG data.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the architecture of the proposed EEG-DBNet model. Section 4 describes the experimental setup and reports the results. Finally, Sect. 5 concludes the paper and outlines potential directions for future work.

2 Related work

This section reviews existing methods for MI-EEG decoding. With the development of deep learning techniques, recent studies have mainly focused on convolutional architectures and hybrid models that integrate multiple modeling paradigms. Accordingly, related work can be broadly categorized into CNN-based methods and hybrid methods.

2.1 CNN-based methods

CNNs have been widely applied in MI-EEG decoding due to their ability to automatically extract discriminative features from raw signals through end-to-end learning. Representative models, such as EEGNet, adopt depthwise and separable convolutions to efficiently capture temporal and spatial patterns while significantly reducing model complexity [9]. Building upon this design, subsequent studies have explored more advanced convolutional architectures to enhance feature representation. For instance, EEG-Inception introduces multi-scale convolutional kernels to capture features at different temporal resolutions [10], while other works incorporate sliding window strategies and improved preprocessing schemes to further enhance classification performance [19].

In addition, efforts have been made to extend CNNs to higher-dimensional representations. For example, 3D convolutional models preserve the intrinsic spatio-temporal structure of EEG signals by jointly modeling channel and temporal dimensions, leading to improved performance and interpretability [20]. Multi-branch CNN architectures have also been proposed to learn complementary features from parallel convolutional pathways, enabling richer feature extraction across different perspectives [21, 22].

Despite these improvements, CNN-based methods still exhibit limited capability in modeling global dependencies and complex EEG representations.

2.2 Hybrid methods

Hybrid architectures that integrate convolutional operations with attention mechanisms or advanced sequence modeling have become a prominent direction for MI-EEG decoding, aiming to combine local feature extraction with global dependency modeling.

A representative line of work introduces transformer-based or attention-enhanced frameworks. For instance, EEG Conformer integrates convolutional layers with transformer encoders to jointly model local patterns and long-range dependencies [12]. Similarly, temporal-spatial transformer models explicitly employ self-attention to capture global temporal and inter-channel relationships [23]. To further enhance temporal modeling, attention-based temporal convolutional networks incorporate multi-head attention into convolutional structures for more effective temporal feature extraction [13]. In addition, channel attention combined with hierarchical transformer architectures has been explored to emphasize discriminative spatial information across EEG channels [24].

Graph-based methods have also been introduced to explicitly model structural relationships among EEG channels. The graph sequence neural network constructs dynamic adjacency matrices to capture both local and global channel dependencies for MI decoding [25]. Similarly, explainable graph neural network frameworks have been developed to identify neural biomarkers by modeling connectivity patterns across brain regions [26].

More recently, hybrid architectures integrating Transformer and alternative sequence modeling paradigms have been explored. The FA-STTM model combines Transformer and Mamba mechanisms to improve temporal dependency modeling under complex conditions [27]. In addition, the MifNet model introduces a Mamba-based interactive frequency convolutional framework to jointly model temporal dynamics and frequency-domain characteristics [28].

Although these methods have improved MI-EEG decoding performance, a unified framework for jointly modeling diverse EEG representations remains insufficiently explored.

2.3 Differences from the conference paper

This paper is an extended version of our previous conference paper [29]. Compared with the conference version, BCI Competition IV-2b and the self-collected CQUPT dataset are included for a broader evaluation. For the self-collected CQUPT dataset, details on the acquisition system, experimental paradigm, preprocessing procedure, sample construction, and evaluation protocol are provided. Moreover, the comparative experiments are extended by including representative MI-EEG decoding models, such as EEGNet, EEG-NeX, ITNet, and ATCNet. Additional ablation experiments and parameter sensitivity analyses are also provided to examine the contribution of the main components and hyperparameter choices of EEG-DBNet.

3 Method

Figure 1 illustrates the processing pipeline of the proposed EEG-DBNet. The network comprises two parallel branches, each consisting of a Local Convolutional (LC) block and a Global Convolutional (GC) block. After the input signal is processed by both branches, the resulting features are concatenated and passed through a Fully Connected (FC) layer.

EEG-DBNet is trained on the set $\{\mathrm{X}_i, y_i\}_{i=1}^{m}$, where $\mathrm{X}_i \in \mathbb{R}^{1\times E\times T}$ denotes the *i*-th trial, with *E* representing the number of EEG electrodes and *T* the number of sampled time points. The label $y_i \in \{1, 2, \dots, N_{\mathrm{cls}}\}$ denotes the class index, where N_{cls} is the total number of classes. Once trained, the model is used to classify unseen test samples.

3.1 Local convolutional block

The LC block in EEG-DBNet consists of two parallel branches—the temporal branch and the spectral branch—each designed to extract distinct features from EEG signals, as illustrated in Fig. 2.

The architecture of each LC branch is inspired by EEGNet [9], comprising a basic convolutional layer, a depthwise convolutional layer, and a separable convolutional layer, interleaved with pooling layers. Batch Normalization (BN) [30] and the Exponential Linear Unit (ELU) [31] activation function are employed to facilitate training, while a dropout rate of 0.3 is applied to mitigate overfitting.

The temporal branch illustrates the LC block in detail. The first convolutional layer applies $\hat{F}_1 = 8$ temporal filters of size (1, *K*), where $K = 2\alpha^2$. Here, α is a dataset-specific parameter, and its value varies based on the dataset. Further details on α will be provided in Sect. 4.

Next, a depthwise convolution layer with $D = 2$ and $\hat{F}_1 \times D$ filters of size (*E*, 1) is employed to extract spatial features from each temporal map, followed by average pooling with size $(1, \alpha)$ to reduce the temporal resolution.

Subsequently, a separable convolution layer with $\hat{F} = \hat{F}_1 \times D$ filters of size (1, *K*/4) models feature interdependencies while compressing the temporal range. A final pooling layer of size $(1, \alpha/2)$ further downsamples the signal. Accordingly, the LC block transforms $\mathrm{X}_i \in \mathbb{R}^{1\times E\times T}$ into $\hat{\mathrm{X}}_i \in \mathbb{R}^{\hat{F}\times 1\times \hat{T}}$, where

$$\hat{T} = \left\lceil \frac{2\times T}{\alpha^2} \right\rceil \quad (1)$$

The spectral branch shares a similar architecture but is tailored for spectral feature extraction. It employs a second pooling layer with a kernel size changed from $\alpha/2$ to α, and uses more convolutional filters, producing an output $\tilde{\mathrm{X}}_i \in \mathbb{R}^{\tilde{F}\times\tilde{T}}$, where

$$\tilde{T} = \left\lceil \frac{T}{\alpha^2} \right\rceil \quad (2)$$

This design increases spectral resolution while reducing temporal length, ensuring dimensional symmetry between branches.

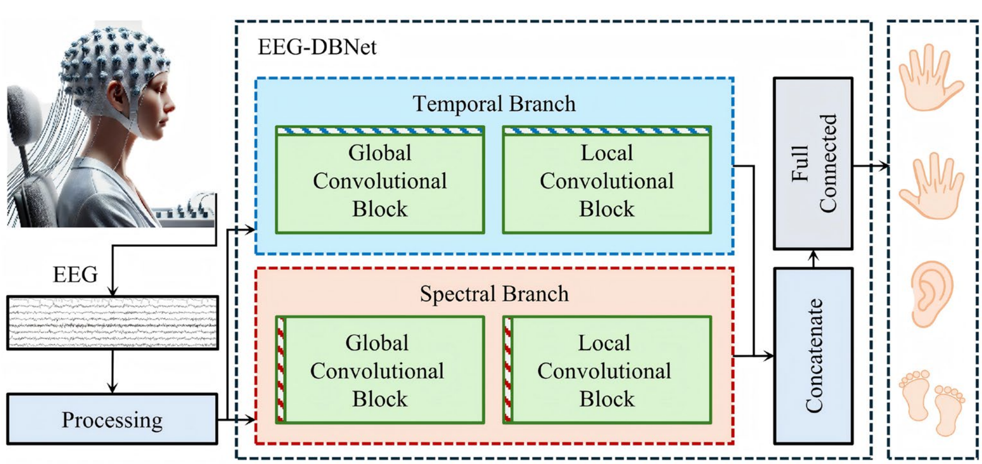

Fig. 1 Processing pipeline of the proposed EEG-DBNet model

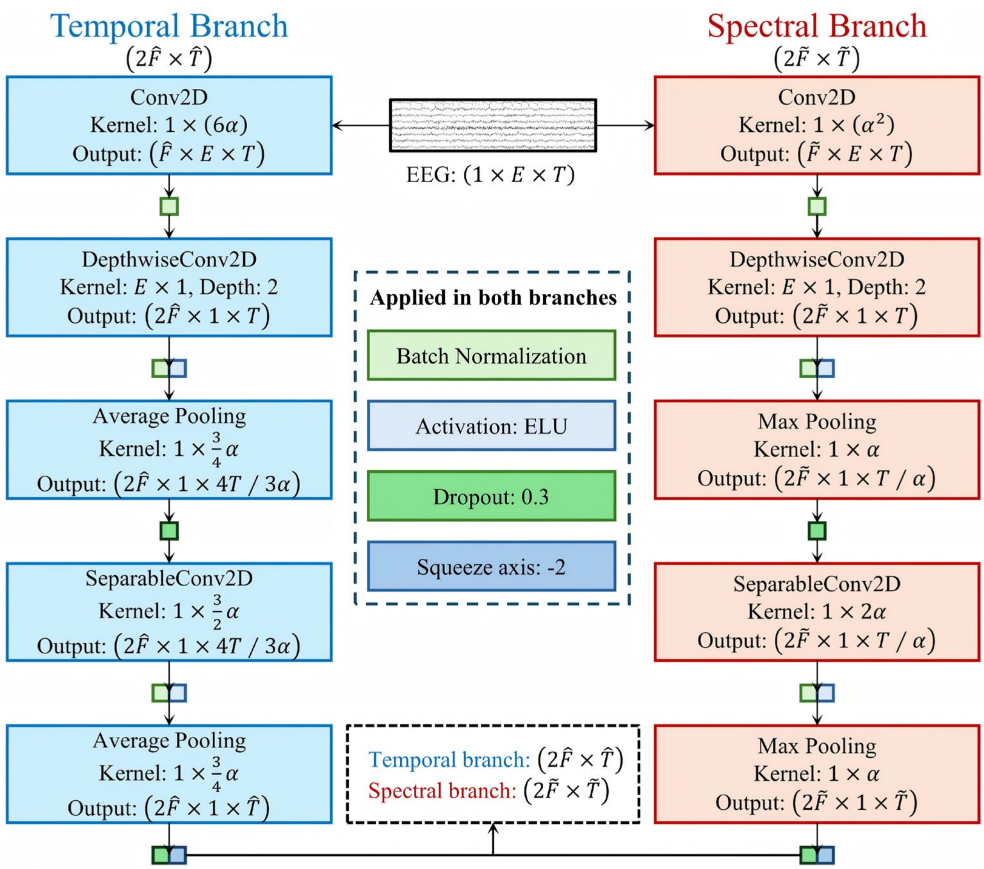

Fig. 2 Structure of the local convolutional block in EEG-DBNet

3.2 Global convolutional block

The GC block extracts global features from the LC block output by segmenting each feature sequence into N overlapping subsequences using a sliding window with stride 1. For the temporal branch output $\hat{\mathrm{X}}_i \in \mathbb{R}^{\hat{F} \times \hat{T}}$, each subsequence has length $\hat{l} = \hat{T} - N + 1$, yielding N subsequences $\hat{\mathrm{X}}_{n,i} \in \mathbb{R}^{\hat{F} \times \hat{l}}$, as illustrated in Fig. 3.

Each subsequence $\hat{\mathrm{X}}_{\mathrm{SE},n,i}$ is recalibrated via a lightweight squeeze-and-excitation (SE) module [32], which applies global average pooling followed by two FC layers with ReLU (δ) and sigmoid (σ) activations, respectively:

$$\begin{aligned}\hat{\mathrm{X}}_{\mathrm{SE},n,i} &= \hat{\mathrm{X}}_{n,i} \odot \hat{\mathrm{w}}_{n,i} \\ &= \hat{\mathrm{X}}_{n,i} \odot \sigma\left[\delta\left(\hat{\mathrm{x}}_{\mathrm{av},n,i} \cdot \hat{\mathrm{W}}_1\right) \cdot \hat{\mathrm{W}}_2\right] \\ &= \hat{\mathrm{X}}_{n,i} \odot \sigma\left[\delta\left(\hat{f}_{\mathrm{av}}\left(\hat{\mathrm{X}}_{n,i}\right) \cdot \hat{\mathrm{W}}_1\right) \cdot \hat{\mathrm{W}}_2\right] \\ &= \hat{\mathrm{X}}_{n,i} \odot \sigma\left[\delta\left(\frac{1}{\hat{F}}\sum_{j=1}^{\hat{F}} \hat{\mathrm{X}}_{n,i}(j) \cdot \hat{\mathrm{W}}_1\right) \cdot \hat{\mathrm{W}}_2\right]\end{aligned} \tag{3}$$

where $\odot$ denotes the Hadamard product, and $\hat{\mathrm{W}}_1$, $\hat{\mathrm{W}}_2$ are parameters of the fully connected layers.

The recalibrated subsequences are then processed by d dilated causal convolutional (DCC) layers, each followed by batch normalization and ELU (ζ) activation. A residual connection is used to enhance stability. Let k denote the kernel size and ℓ denote the layer index. The dilation rate of the ℓ-th DCC layer is set to $r_\ell = \ell$, where $\ell = 1, 2, \ldots, d$. The receptive field of the stacked DCC layers is computed as:

$$R_d = 1 + \sum_{\ell=1}^{d} (k-1)\ell = 1 + \frac{d(d+1)}{2}(k-1). \tag{4}$$

To avoid information loss, the receptive field must cover the entire subsequence. Therefore, the following constraint must be satisfied for the temporal branch:

$$R_d \geq \hat{T} - N + 1. \tag{5}$$

As reported in Sect. 4.5, the optimal values for d and k in both branches are 4 and 4, respectively. The output of the GC block for the n-th subsequence is given by:

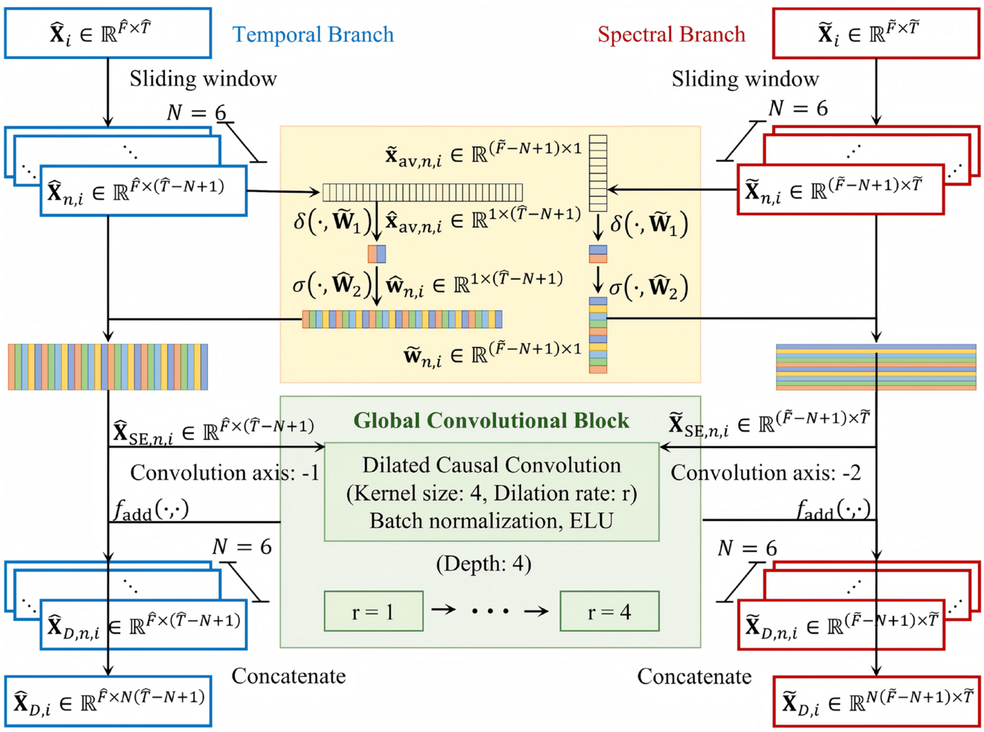

Fig. 3 Structure of the global convolutional block in EEG-DBNet

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

where the output of the ℓ-th DCC layer is computed as

$$\hat{\mathrm{X}}_{\mathrm{DCC},\ell} = \zeta\left(\mathrm{DCC}_\ell\left(\hat{\mathrm{X}}_{\mathrm{DCC},\ell-1}\right)\right), \quad \hat{\mathrm{X}}_{\mathrm{DCC},0} = \hat{\mathrm{X}}_{\mathrm{SE},n,i}, \quad \ell = 1, 2, \ldots, d. \tag{7}$$

where $\mathrm{DCC}_\ell(\cdot)$ denotes the dilated causal convolution operation in the ℓ-th DCC layer with dilation rate r_ℓ. The final output of the GC block in the temporal branch is formed by concatenating all N processed subsequences: $\hat{\mathrm{X}}_{D,i} \in \mathbb{R}^{\hat{F} \times N(\hat{T}-N+1)}$, which is then vectorized into $\hat{\mathrm{x}}_i \in \mathbb{R}^{\hat{F} \cdot N(\hat{T}-N+1)}$.

For the spectral branch output from the LC block, $\tilde{\mathrm{X}}_i \in \mathbb{R}^{\tilde{F} \times \tilde{T}}$, each subsequence has length $\tilde{l} = \tilde{F} - N + 1$, yielding N segments $\tilde{\mathrm{X}}_{n,i} \in \mathbb{R}^{\tilde{l} \times \tilde{T}}$. After SE and DCC processing, the output is similarly vectorized into $\tilde{\mathrm{x}}_i \in \mathbb{R}^{N(\tilde{F}-N+1) \cdot \tilde{T}}$.

Finally, the two feature vectors $\hat{\mathrm{x}}_i$ and $\tilde{\mathrm{x}}_i$ are concatenated and fed into a fully connected layer to generate class logits. A softmax function is then applied to obtain the predicted probability distribution over N_{cls} motor imagery classes:

$$\check{\mathrm{y}}_i = \mathrm{softmax}\left[\mathrm{W}\left(\hat{\mathrm{x}}_i ^\frown \tilde{\mathrm{x}}_i\right) + \mathrm{b}\right], \quad \check{\mathrm{y}}_i \in \mathbb{R}^{N_{\mathrm{cls}}} \tag{8}$$

where W and b denote the weights and bias of the fully connected layer, respectively, and $^\frown$ denotes vector concatenation. The model is trained using cross-entropy loss, which is consistent with the single-label classification setting.

4 Experiments

4.1 Datasets

To comprehensively evaluate the effectiveness of the proposed EEG-DBNet, extensive experiments were conducted on two publicly available motor imagery (MI) EEG datasets, namely BCI Competition IV-2a (BCI42a) [33] and BCI Competition IV-2b (BCI42b) [34], as well as a self-collected dataset (CQUPT). The use of benchmark datasets enables fair comparison with existing methods, while the CQUPT dataset further evaluates the practical applicability of the proposed model using laboratory-acquired MI-EEG data.

The BCI42a dataset consists of EEG recordings from 9 subjects, each performing four motor imagery tasks, including left hand (L), right hand (R), feet (F), and tongue (T). The EEG signals were recorded using 22 electrodes across two sessions conducted on different days. Each subject completed 72 trials per class in each session, resulting in a total of 576 trials per subject. Following the standard evaluation protocol, data from the first session were used for training, while data from the second session were used as the evaluation set. The time segment for each trial was set to [1.5, 6] seconds.

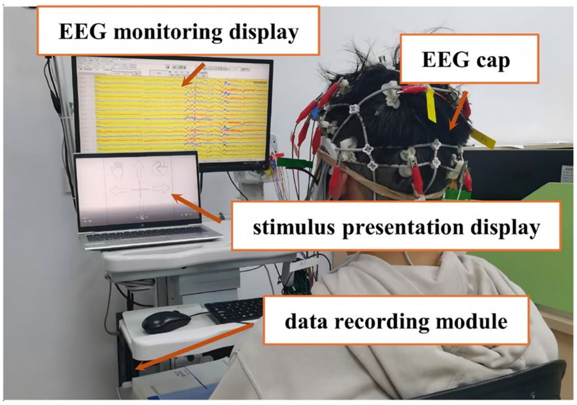

Fig. 4 Experimental setup of the EEG acquisition system for the CQUPT dataset

The BCI42b dataset includes EEG recordings from 9 subjects using 3 electrodes. Each subject performed two motor imagery tasks: left-hand (L) and right-hand (R) imagery. The dataset contains five sessions per subject. In this study, data from the first three sessions were used for training, while the remaining sessions were used as the evaluation set. Each subject completed 400 trials in the training set and 320 trials in the evaluation set, resulting in a total of 720 trials per subject. The time segment for each trial was set to [2.5, 7] seconds.

In addition to the public datasets, a self-collected motor imagery EEG dataset, referred to as the CQUPT dataset, was constructed to further evaluate the practical applicability of the proposed method. EEG signals were recorded using a Nihon Kohden Neurofax EEG-1200C system with a standard 10-20 electrode configuration. Four subjects participated in the experiment. The experimental setup of the EEG acquisition system is illustrated in Fig. 4.

The CQUPT dataset consists of three classes: resting state, left-hand motor imagery (open palm), and right-hand motor imagery (fist). The motor imagery experimental paradigm and the temporal structure of a single trial are illustrated in Fig. 5. Each trial lasted 8 s and included a resting stage [0, 2] seconds, an attention cue stage [2, 4] seconds, and a motor imagery stage [4, 8] seconds.

For CQUPT sample construction, raw EEG recordings were organized into trial-level samples according to the experimental markers. For each trial, the resting-state segment [0, 2] s was used for the resting class, and the motor imagery segment [4, 8] s was used for left-hand and right-hand MI classes. No additional sliding-window segmentation was applied. The EEG signals were band-pass filtered at 0.5–30 Hz. Each input sample was represented as $X_i \in \mathbb{R}^{1\times E\times T}$, where $E = 19$ denotes the number of EEG channels and T denotes the number of sampled time points. The sampling rate of the CQUPT dataset was 256 Hz. No additional normalization was applied.

For the CQUPT dataset, we adopted a within-subject stratified five-fold cross-validation strategy at the segmented-sample level. Specifically, segmented EEG samples from each subject were divided into five class-balanced folds, and the final results were averaged over the five folds and four subjects. A summary of the three MI-EEG datasets and their corresponding preprocessing protocols is presented in Table 1.

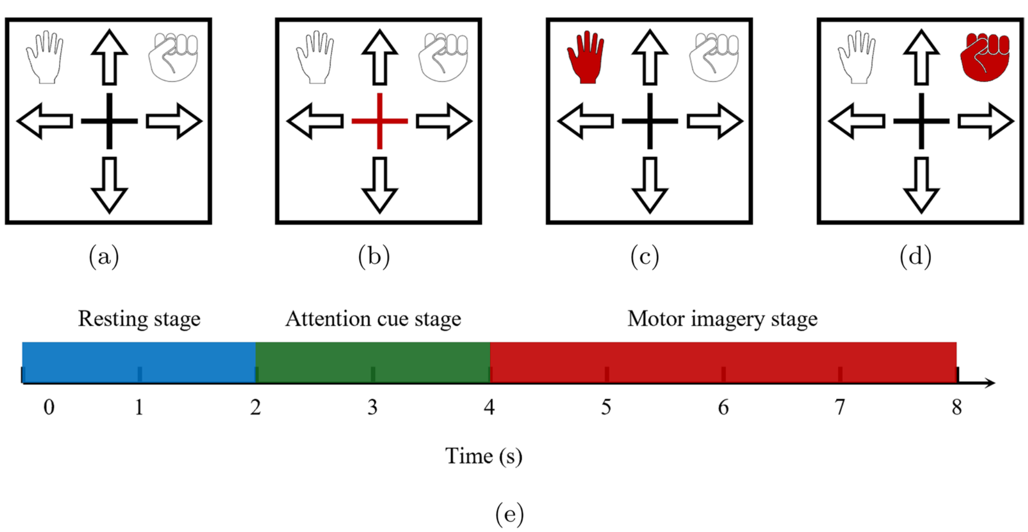

Fig. 5 Motor imagery experimental paradigm and temporal structure of a single trial. **a** Resting state interface. **b** Attention cue interface. **c** Left-hand open palm motor imagery cue. **d** Right-hand fist motor imagery cue. **e** Temporal structure of a single trial

Table 1 Summary of the three MI-EEG datasets and their split protocols

Dataset	Subjects	S.R. (Hz)	Channels	Classes	α	S.P
BCI42b	9	250	3	2	8	Session-based†
CQUPT	4	256	19	3	4	5-fold CV$^{\natural}$
BCI42a	9	250	22	4	8	Session-based‡

S.R., sampling rate; S.P., split protocol; α, scale parameter used in EEG-DBNet

† Session-based split for BCI42b: the first three sessions are used for training and the remaining sessions for evaluation

♮ Within-subject stratified five-fold cross-validation at the segmented-sample level

‡ Session-based split for BCI42a: the first session is used for training and the second session for evaluation

Table 2 Classification performance (P_a(%) and *K*) on BCI42b for the proposed EEG-DBNet with other models

Sub.	EEGNet [9]		EEG-NeX [11]		ITNet [22]		ATCNet [13]		EEG-DBNet	
	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*
1	79.82	0.5970	80.70	0.6127	82.46	0.6498	81.58	0.6320	**85.09**	**0.7019**
2	71.43	0.4282	**77.14**	**0.5428**	75.51	0.5102	73.06	0.4607	75.10	0.5018
3	90.00	0.7998	**92.61**	**0.8515**	90.00	0.7979	90.87	0.8172	91.74	0.8339
4	98.37	0.9674	**99.02**	**0.9805**	98.70	0.9739	99.02	0.9805	98.37	0.9674
5	99.63	0.9927	99.63	0.9927	98.53	0.9707	**100.00**	**1.0000**	99.27	0.9853
6	**93.63**	**0.8723**	92.03	0.8406	90.44	0.8083	90.04	0.8009	92.43	0.8483
7	95.69	0.9138	94.83	0.8966	94.40	0.8879	**96.55**	**0.9310**	95.26	0.9052
8	95.22	0.9043	96.06	0.9217	96.52	0.9303	95.22	0.9043	**96.96**	**0.9390**
9	89.80	0.7959	89.80	0.7958	88.16	0.7629	90.20	0.8038	**90.20**	**0.8039**
St.D.*	9.22	0.0721	**7.12**	**0.0556**	8.04	0.0630	8.92	0.0670	8.35	0.0661
Mean	90.40	0.8079	91.32	0.8261	90.52	0.8102	90.73	0.8145	**91.60**	**0.8318**

* Standard deviation

Bold values indicate the best result for each subject and metric

4.2 Experiment settings

The proposed EEG-DBNet and other comparative models were trained and tested using the TensorFlow framework, running on the Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz and NVIDIA A10 GPU. All models were trained using the Adam optimizer with a learning rate of 0.0009, a batch size of 64, and cross-entropy loss over 1000 epochs with early stopping set to patience of 300 epochs. Each experiment was repeated ten times, and the mean and standard deviation of the evaluation results were reported.

The evaluation metrics used were class-averaged accuracy P_a and the kappa value *K*. The class-averaged accuracy P_a is defined as:

$$P_a = \frac{1}{N_{\mathrm{cls}}} \sum_{i=1}^{N_{\mathrm{cls}}} \frac{\mathrm{TP}_i}{l_i} \tag{9}$$

where N_{cls} denotes the number of classes, TP_i denotes the number of correctly predicted samples in class *i*, and l_i is the number of samples in class *i*. The kappa value *K* is defined as:

$$K = \frac{P_a - P}{1 - P} \tag{10}$$

where P indicates the hypothetical probability of chance agreement.

4.3 Performance comparison

The proposed EEG-DBNet was compared with four representative state-of-the-art MI-EEG decoding methods, namely EEGNet [9], EEG-NeX [11], ITNet [22], and ATCNet [13]. EEGNet is regarded as a classical lightweight benchmark model for MI-EEG decoding. EEG-NeX is an enhanced variant of EEGNet with improved feature extraction capability. ITNet adopts a multi-branch temporal convolution structure, while ATCNet incorporates attention mechanisms and global temporal feature extraction modules. To ensure a fair comparison, all competing models were reproduced on the BCI42a, BCI42b, and CQUPT datasets using the hyperparameter settings reported in their original studies.

Tables 2, 3, and 4 present the experimental results for each model on the BCI42b, CQUPT, and BCI42a datasets, respectively. EEG-DBNet achieved the best results on the BCI42a dataset. ATCNet's performance was close to that of EEG-DBNet, with significantly higher accuracy than the other methods. Notably, EEG-DBNet also achieved the lowest standard deviation, indicating its superior generalization capability across subjects. Similarly, EEG-DBNet achieved the best overall results on the BCI42b dataset. Since BCI42b is a binary classification

Table 3 Classification performance ($P_a(\%)$ and *K*) on CQUPT for the proposed EEG-DBNet with other models

Sub.	EEGNet [9]		EEG-NeX [11]		ITNet [22]		ATCNet [13]		EEG-DBNet	
	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*
1	45.48	0.1770	66.05	0.4882	62.59	0.4364	64.63	0.4670	**70.13**	**0.5493**
2	41.22	0.1149	**69.75**	**0.5459**	61.32	0.4193	59.06	0.3848	63.04	0.4441
3	44.45	0.1656	**72.68**	**0.5901**	45.73	0.1819	48.18	0.2223	70.32	0.5548
4	43.39	0.1493	**71.26**	**0.5687**	61.69	0.4243	51.82	0.2758	69.45	0.5413
St.D.*	**1.82**	**0.0270**	2.86	0.0439	8.09	0.1226	7.36	0.1097	3.48	0.0525
Mean	43.63	0.1517	**69.93**	**0.5482**	57.83	0.3655	55.92	0.3375	68.23	0.5224

* Standard deviation

Bold values indicate the best result for each subject and metric

Table 4 Classification performance comparison ($P_a(\%)$ and *K*) on the BCI42a dataset

Sub.	EEGNet [9]		EEG-NeX [11]		ITNet [22]		ATCNet [13]		EEG-DBNet	
	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*	P_a	*K*
1	89.32	0.8577	85.05	0.8007	86.48	0.8197	88.61	0.8481	**90.75**	**0.8766**
2	63.25	0.5103	71.02	0.6136	67.84	0.5714	73.85	0.6516	**74.56**	**0.6610**
3	94.87	0.9316	96.34	0.9512	95.60	0.9414	97.44	0.9658	**97.44**	**0.9658**
4	67.54	0.5666	**85.09**	**0.8009**	74.12	0.6543	81.58	0.7539	82.46	0.7659
5	75.00	0.6664	81.52	0.7535	77.90	0.7050	80.80	0.7443	**82.61**	**0.7678**
6	60.93	0.4792	63.72	0.5163	64.19	0.5229	73.95	0.6525	**75.35**	**0.6713**
7	89.89	0.8654	90.61	0.8749	90.97	0.8797	**93.50**	**0.9134**	91.34	0.8844
8	87.82	0.8377	83.39	0.7786	83.03	0.7737	**90.04**	**0.8671**	88.93	0.8524
9	81.44	0.7523	84.09	0.7878	85.23	0.8030	**89.77**	**0.8635**	89.39	0.8586
St.D.*	11.93	0.1592	9.18	0.1224	9.86	0.1315	7.90	0.1053	**7.23**	**0.0963**
Mean	78.90	0.7186	82.32	0.7642	80.60	0.7412	85.50	0.8067	**85.87**	**0.8115**

* Standard deviation

Bold values indicate the best result for each subject and metric

Table 5 Ablation study of EEG-DBNet components on the BCI42a dataset

Block	$\bar{P}_a(\%)^*$	$\bar{K}^*$
Single-branch (EEGNet [9])	78.90	0.7186
Dual-branch (LC^1)	82.43	0.7656
LC^1 +GC without SE and SW	82.78	0.7703
LC^1 +GC without SE	84.39	0.7918
LC^1 +GC without SW	83.61	0.7814
LC^2 +GC	83.16	0.7754
LC^3 +GC	82.85	0.7713
LC^4 +GC	81.90	0.7587
LC^1 +GC (EEG-DBNet)	**85.87**	**0.8115**

* Average of nine subjects

[1] Average pooling for temporal branch and max pooling for spectral branch

[2] Average pooling for both branches

[3] Max pooling for both branches

[4] Max pooling for temporal branch and average pooling for spectral branch

Bold values indicate the best performance among all ablation variants for each metric

task, which is simpler than the four-class classification task in BCI42a, all methods attained relatively high classification accuracy. In this dataset, EEG-NeX achieved performance comparable to EEG-DBNet and demonstrated the best generalization capability. Nevertheless, EEG-DBNet still outperformed the other methods by approximately one percentage point in average accuracy. For the CQUPT dataset, EEG-DBNet also demonstrated competitive performance and achieved higher decoding accuracy than EEGNet, ITNet, and ATCNet.

Overall, the proposed EEG-DBNet achieved superior performance on the BCI42a and BCI42b datasets and competitive performance on the CQUPT dataset, demonstrating its effectiveness and generalization capability for MI-EEG decoding tasks.

4.4 Ablation experiments

The effectiveness of each module of EEG-DBNet was evaluated through ablation experiments on the BCI42a dataset. With 22 EEG channels, nine subjects, and a challenging four-class motor imagery task, BCI42a provides a comprehensive benchmark for analyzing module-level contributions. As shown in Table 5, the dual-branch network, which employs average pooling for temporal sequences and max pooling for spectral sequences (LC block), improves accuracy by 3.5% over the single-branch network, achieving 82.43%. This indicates that the dual-branch network has superior feature extraction capability, which can be attributed to the complementary feature extraction of the two branches. The introduction of DCCNNs further enhances accuracy by 0.35% , reaching 82.78%, demonstrating that DCCNNs serve

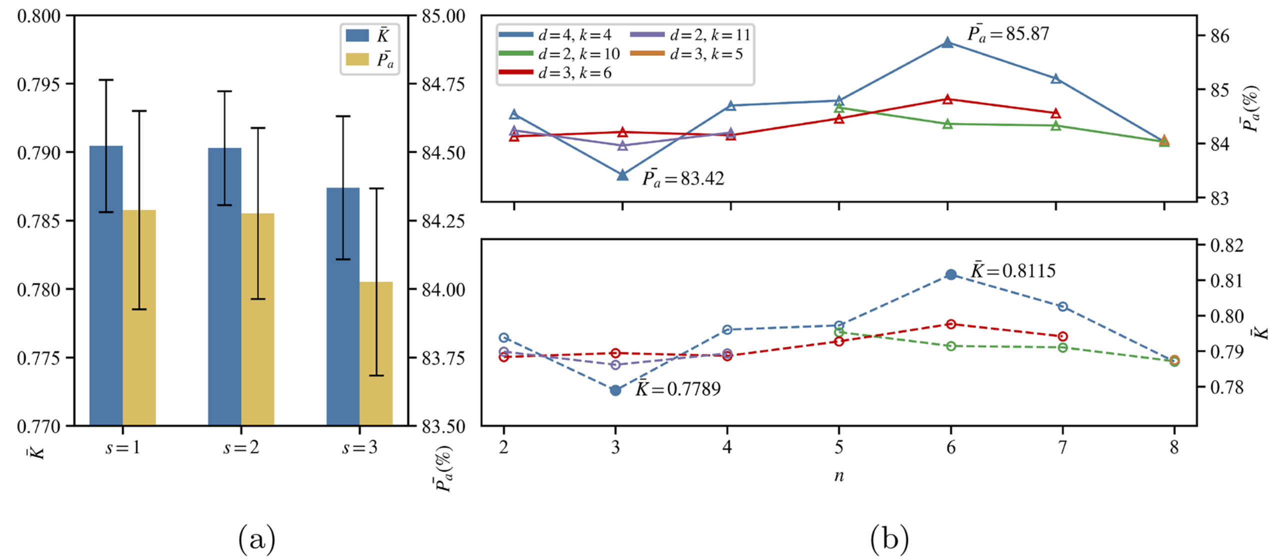

Fig. 6 Parameter sensitivity of EEG-DBNet. **a** Classification performance ($P_a(\%)$ and *K*) for different values of stride (*s*). **b** Classification performance ($P_a(\%)$ and *K*) for various hyperparameters ($s = 1$)

as a valuable supplement to CNNs. The key difference between DCCNNs and CNNs is that DCCNNs increase the receptive field by adding network depth, thereby enhancing the network's global feature extraction ability.

Splitting the sequence into six subsequences and then extracting features from these using DCCNNs improves the model's classification accuracy by 1.61%, reaching 84.39%. This suggests that sequence splitting enables DCCNNs to construct different feature extraction paradigms and increase feature diversity. By reconstructing the features of the sequence and then extracting features using DCCNNs, the model's classification accuracy improves by 0.83%, reaching 83.61%. This indicates that remodeling the relationships between sequence samples helps DCCNNs more accurately extract signal features. When both sequence splitting and feature reconstruction are used simultaneously, DCCNNs processing the reconstructed subsequences can improve the model's classification accuracy by 3.09%, surpassing the sum of their individual effects ($1.61\% + 0.83\% = 2.44\%$), reaching 85.87%. This demonstrates that reconstructed subsequences can enhance the efficiency of DCCNNs in constructing different feature extraction paradigms. The introduction of the GC block based on DCCNNs significantly improves model performance, increasing accuracy by 3.44% over the dual-branch network with only the LC block, indicating that the GC block and LC block provide substantial complementary effects.

Furthermore, Table 5 shows that the model performance is highly sensitive to the type of pooling layer used for different signal dimensions. If both network branches utilize average pooling, the accuracy of the model decreases by 2.71% compared to EEG-DBNet, reaching 83.16%. If both network branches use max pooling, the accuracy decreases by 3.02%, dropping to 82.85%. When max pooling is used in the temporal branch and average pooling in the spectral branch, accuracy drops by 3.97%, reaching 81.90%, which is even lower than the performance of the network composed solely of LC. Therefore, we deduce that for spectral sequences, specific sample points better reflect the sequence characteristics than the relationships between multiple sample points. Conversely, for temporal sequences, the relationships between sample points are more important than individual points. As a result, average pooling is more suitable for temporal sequences, while max pooling is more appropriate for spectral sequences.

4.5 Parameter sensitivity

The GC block has four hyperparameters: the stride *s* of the sliding window, the number of sliding windows *n*, the kernel size *k* of the dilated convolution, and the number of dilated convolution layers *d*. The first step of the GC block is to split the sequence, where *s* must be determined to segment the sequence into *n* parts based on its length.

We conducted 36 experiments on the BCI42a with $s \in \{1, 2, 3\}$ and $n \in \{2, 3, 4, 5\}$, matching the corresponding *k* and *d* according to the constraint of (5). We recorded the average P_a ($\bar{P_a}$) and average K ($\bar{K}$) across nine subjects to evaluate the impact of different *s* values on the GC block. The experimental results are illustrated in Fig. 6a, where the results of the 36 experiments are divided into three groups based on different *s* values. It can be observed that as *s* increases, the classification accuracy of the model decreases. This is likely due to the larger stride accentuating the differences in the features of each subsequence. Therefore, we set $s = 1$, and then chose $n \in \{2, 3, 4, 5, 6, 7, 8\}$ to select all *d* and *k* that satisfy the constraint of (5) for our experiments. The results, as shown in Fig. 6b, indicate that $d = 4$ and $k = 4$ meet the constraint conditions for all values of *n*, while $d = 3$ and $k = 5$ only satisfy the constraints when $n = 8$. The model exhibits its worst and best accuracy both at $d = 4$ and $k = 4$. Specifically, the model performs worst at $d = 4$, $k = 4$, $n = 3$ with $\bar{P}_a = 83.42\%$ and $\bar{K} = 0.7789$, and it performs best at $d = 4$, $k = 4$, $n = 6$ with $\bar{P}_a = 85.87\%$ and $\bar{K} = 0.8115$. Consequently, we set the four hyperparameters of the GC block to $s = 1$, $d = 4$, $k = 4$, and $n = 6$.

5 Conclusion and discussion

This study proposes EEG-DBNet, a dual-branch network designed to decode temporal and spectral representations of MI-EEG in parallel for accurate motor imagery classification. The temporal sequences alone provide limited signal features; thus, a multi-dimensional exploration of these features is crucial for enhancing model performance. The parallel dual-branch network is essential for this decoding approach. If temporal and spectral sequences are decoded sequentially, the model accuracy drops below 81%, highlighting the equal importance of both feature types for decoding MI-EEG.

The primary function of the LC block is to remodel sequence sample points. During this process, the LC block captures local signal features and adjusts sequence length by decoding inter-sample relationships, setting the stage for subsequent global feature extraction. Our findings indicate that using different types of pooling layers for sequences of varying dimensions significantly impacts model performance. Specifically, employing an average pooling layer in the temporal branch and a max pooling layer in the spectral branch enables optimal model performance. We infer that incorporating information from all samples within the pooling range better represents the temporal sequence characteristics, while the maximum value of the samples within the pooling range more effectively captures signal fluctuation amplitudes, thereby better representing spectral sequence characteristics.

The value of DCCNNs lies in their ability to adjust the network's receptive field, addressing the "breadth" issue in signal decoding. Sequence splitting allows DCCNNs to train multiple signal decoding models, extracting features from each subsequence and thus enhancing the network's ability to capture global features in terms of "depth." High similarity between subsequences hinders the training of diverse feature extraction models in DCCNNs. Feature reconstruction of subsequences increases their diversity while maintaining continuity, thereby improving the model's efficiency in extracting global features. Sequence splitting and feature reconstruction effectively strengthen DCCNNs, and fine-tuning DCCNN hyperparameters according to sequence length significantly improves model performance.

In summary, the design of EEG-DBNet is effective, with each module contributing to improved model performance. Experiments on the two public benchmark datasets achieved the best average results among the compared methods, while the self-collected CQUPT dataset showed competitive performance. Nevertheless, the current CQUPT dataset contains a limited number of subjects, and further validation with larger-scale self-collected data is needed. In addition, the numerous hyperparameters in EEG-DBNet make the tuning process cumbersome. Integrating the LC and GC blocks to reduce hyperparameters could improve the model's practicality. Additionally, exploring methods for extracting spatial signal features to further increase signal decoding dimensions is a key direction for future research.

Acknowledgements

The authors gratefully acknowledge the organizers and contributors of the BCI Competition IV for providing access to the public BCI Competition IV-2a and IV-2b datasets. The authors also sincerely thank all volunteers who participated in the self-collected CQUPT motor imagery EEG acquisition experiment.

Author contributions

Youxi Qu performed the experiments, implemented the model, analyzed the data, and drafted the manuscript. Xicheng Lou contributed to the conception of EEG-DBNet, methodological design, experimental design, and interpretation of the results. Xinwei Li provided overall supervision, project administration, conceptual guidance, and critical revision of the manuscript. Hongying Meng and Zhangyong Li provided methodological guidance and contributed to manuscript revision. Jianlin Wang and Kunpeng Mao contributed to data acquisition and data processing. All authors read and approved the final manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (grant number 62576066); the Key Project of Science and Technology Research Program of Chongqing Municipal Education Commission (grant number KJZD-K202400602); and the Chongqing Graduate Student Research Innovation Project (grant number CYS260478).

Data availability

The public BCI Competition IV-2a and IV-2b datasets analyzed in this study are publicly available at https://www.bbci.de/competition/iv/. The self-collected CQUPT dataset used in this study is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The use and analysis of the publicly available, fully anonymized BCI Competition IV-2a and IV-2b datasets were carried out in strict accordance with the ethical standards and guidelines of the institutional and/or national research committee. The self-collected CQUPT dataset was approved by the Ethics Committee of Chongqing University of Posts and Telecommunications under approval number NSFC202103. EEG acquisition was conducted in accordance with the approved protocol and relevant ethical requirements, and written informed consent was obtained from all participants prior to data collection.

Competing interests

The authors declare no competing interests.

Received: 18 May 2026 / Accepted: 2 July 2026
Published online: 16 July 2026

References

1. Ang KK, Guan C (2017) EEG-based strategies to detect motor imagery for control and rehabilitation. IEEE Trans Neural Syst Rehabil Eng 25(4):392–401
2. Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain-computer communication. Proc IEEE 89(7):1123–1134
3. Arpaia P, Esposito A, Natalizio A, Parvis M (2022) How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art. J Neural Eng 19(3):031002
4. Wang Q, Cao T, Liu D, Zhang M, Lu J, Bai O et al (2021) A motor-imagery channel-selection method based on SVM-CCA-CS. Meas Sci Technol 32(3):035701
5. Yu X, Chum P, Sim KB (2014) Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. Optik 125(3):1498–1502
6. Fei SW, Chu YB (2022) A novel classification strategy of motor imagery EEG signals utilizing WT-PSR-SVD-based MTSVM. Expert Syst Appl 199:116901
7. Dose H, Møller JS, Iversen HK, Puthusserypady S (2018) An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Syst Appl 114:532–542
8. Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M et al (2017) Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp 38(11):5391–5420
9. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ (2018) EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. J Neural Eng 15(5):056013
10. Zhang C, Kim YK, Eskandarian A (2021) EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. J Neural Eng 18(4):046014
11. Chen X, Teng X, Chen H, Pan Y, Geyer P (2024) Toward reliable signals decoding for electroencephalogram: a benchmark study to EEGNeX. Biomed Signal Process Control 87:105475
12. Song Y, Zheng Q, Liu B, Gao X (2023) EEG conformer: convolutional transformer for EEG decoding and visualization. IEEE Trans Neural Syst Rehabil Eng 31:710–719
13. Altaheri H, Muhammad G, Alsulaiman M (2023) Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. IEEE Trans Industr Inf 19(2):2249–2258
14. Hamidi A, Kiani K (2025) Motor imagery EEG signals classification using a Transformer-GCN approach. Appl Soft Comput 170:112686
15. Liu K, Yang T, Yu Z, Yi W, Yu H, Wang G et al (2024) MSVTNet: multi-scale vision transformer neural network for EEG-based motor imagery decoding. IEEE J Biomed Health Inform 28(12):7126–7137
16. Chen J, Pi D, Gao F, Ma C, Chen Y (2026) MI-DGHCL: motor imagery EEG domain generalization via hyperbolic contrastive learning. Expert Syst Appl 312:131477
17. Li M, Zhang N (2022) A dynamic directed transfer function for brain functional network-based feature extraction. Brain Inf 9(1):7
18. Li Z, Li M, Yang Y (2025) Motor imagery decoding network with multisubject dynamic transfer. Brain Inf 12(1):20
19. Atla KGR, Sharma R (2025) Motor imagery classification using a novel CNN in EEG-BCI with common average reference and sliding window techniques. Alex Eng J 120:532–546
20. Park D, Park H, Kim S, Choo S, Lee S, Nam CS et al (2023) Spatio-temporal explanation of 3D-EEGNet for motor imagery EEG classification using permutation and saliency. IEEE Trans Neural Syst Rehabil Eng 31:4504–4513
21. Zhao X, Zhang H, Zhu G, You F, Kuang S, Sun L (2019) A multi-branch 3D convolutional neural network for EEG-based motor imagery classification. IEEE Trans Neural Syst Rehabil Eng 27(10):2164–2177
22. Salami A, Andreu-Perez J, Gillmeister H (2022) EEG-ITNet: an explainable inception temporal convolutional network for motor imagery classification. IEEE Access 10:36672–36685
23. Hameed A, Fourati R, Ammar B, Ksibi A, Alluhaidan AS, Ayed MB et al (2024) Temporal-spatial transformer based motor imagery classification for BCI using independent component analysis. Biomed Signal Process Control 87:105359
24. Wang H, Cao L, Huang C, Jia J, Dong Y, Fan C et al (2023) A novel algorithmic structure of EEG channel attention combined with swin transformer for motor patterns classification. IEEE Trans Neural Syst Rehabil Eng 31:3132–3141
25. Cai S, Li H, Wu Q, Liu J, Zhang Y (2022) Motor imagery decoding in the presence of distraction using graph sequence neural networks. IEEE Trans Neural Syst Rehabil Eng 30:1716–1726
26. Han J, Embs A, Nardi F, Haar S, Faisal AA (2025) Finding neural biomarkers for motor learning and rehabilitation using an explainable graph neural network. IEEE Trans Neural Syst Rehabil Eng 33:554–565
27. Gao M, Gao C, Yao Z, Zhu W (2026) FA-STTM: a hybrid Transformer-Mamba network for motor imagery EEG classification. Neurocomputing 669:132481
28. Yang L, Zhu W (2025) Mifnet: a MamBa-based interactive frequency convolutional neural network for motor imagery decoding. Cogn Neurodyn 19(1):106
29. Lou X, Li X, Meng H, Li Z (2025) EEG-DBNet: a dual-branch network for temporal-spectral decoding in motor-imagery brain-computer interfaces. In: Proceedings of the 18th international conference on brain informatics. Bari, Italy. https://doi.org/10.1007/978-981-95-9575-4_28
30. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd international conference on machine learning. PMLR; p. 448–456
31. Shah A, Kadam E, Shah H, Shinde S, Shingade S (2016) Deep residual networks with exponential linear unit. In: Proceedings of the third international symposium on computer vision and the internet. ACM. p. 59–65
32. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 7132–7141
33. Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G (2008) BCI competition 2008-Graz data set A. Institute for knowledge discovery (Laboratory of Brain-Computer Interfaces). Graz University of Technology 16:1–6
34. Leeb R, Brunner C, Müller-Putz G, Schlögl A, Pfurtscheller G (2008) BCI competition 2008-Graz data set B. Graz Univ Technol Austria 16:1–6

Publisher's Note