A Collaborative Network of Bi-Mamba and CAFS for Multi-Task Classification of Unilateral Lower Limb

Zhicheng Chen[a], Junhua Li[a,b,*]

[a]School of Electronics and Information Engineering, Wuyi University, Jiangmen, 529020, China
[b]School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK

ARTICLE INFO

Keywords:
Electroencephalography
Deep Learning
Motor Imagery
Motor Execution
Motor Observation
Bidirectional Mamba

ABSTRACT

Accurate decoding of motor-related electroencephalography signals remains challenging in brain-computer interface research. To date, motor imagery of the upper limbs has been extensively investigated, while research on the lower limbs remains limited. Meanwhile, multiple tasks involving the unilateral lower limb (including motor imagery, motor execution, and motor observation for both the left and right lower limb) pose decoding challenges because these tasks activate similar brain regions (i.e., sensorimotor cortex), and contralateral cortex activation is closer to the brain's midline compared to the upper limbs. These challenges make classification challenging for existing deep learning models. To address the above problem, we proposed a collaborative network of bidirectional Mamba and convolutional attention feature selection (CN-BMCAFS) to classify multiple tasks of the unilateral lower limb. In CN-BMCAFS, a global temporal bidirectional Mamba module (GT Bi-Mamba) is employed to capture globally discriminative temporal features. The convolutional attention feature selection module (CAFS) is designed to capture informative local temporal features. The GT Bi-Mamba and CAFS are cascaded in a recurrent manner, enabling the combined advantage of GT Bi-Mamba capturing global temporal features while considering local temporal correlations and CAFS capturing local temporal features while considering global temporal correlations. The results demonstrated that CN-BMCAFS achieved the best performance in classification accuracy, with an accuracy improvement ranging from 1.79% to 14.5% compared to other models. Visualization results show that GT Bi-Mamba enhances the model's ability to accurately distinguish features among MI, ME, and MO, thereby improving decoding performance.

1. Introduction

Brain-computer interfaces (BCIs) enable direct communication between the brain and external devices without relying on the involvement of the human body's peripheral nerves and muscles (Nicolas-Alonso and Gomez-Gil, 2012; Wang et al., 2023a). Electroencephalography (EEG) is a non-invasive approach of recording brain activity, and has been extensively employed in the BCI field due to its low cost, portability, and high temporal resolution (Bousseta et al., 2018; Li et al., 2023, 2026). Motor imagery (MI) is one of the classical paradigms in BCIs, which activates the sensorimotor cortex during imagining limb movements (Decety, 1996). Compared to motor imagery, motor execution (ME) easily and significantly activates the aforementioned brain region (Kraeutner et al., 2014). This brain region could also be activated by motor observation (MO) (Kim and Lee, 2021). MI, ME, and MO activate similar brain regions (i.e., sensorimotor regions), but the features of their EEG signals are distinguishable. Currently, motor imagery of the upper limbs has been extensively investigated (Vuckovic and Sepulveda, 2008; Mammone et al., 2020; Zhang et al., 2023; Bi et al., 2023), while research on the lower limbs remains limited (Asanza et al., 2022). Meanwhile, multiple tasks involving the unilateral lower limb (including MI, ME, and MO for both the left and right lower limb) pose decoding challenges because these tasks activate similar brain regions, and contralateral cortex activation is closer to the brain's midline compared to the upper limbs (Li et al., 2018). These challenges make classification difficult for existing deep learning models.

The process of decoding EEG signals through traditional machine learning can be divided into two stages: feature extraction and feature classification. The feature extraction stage involves various techniques, such as Common Spatial Pattern (CSP) (Ang et al., 2012) for spatial feature extraction, Continuous Wavelet Transform (Chaudhary et al., 2020) and Empirical Wavelet Transform (Bhattacharyya and Pachori, 2017) for time-frequency feature extraction. These extracted features are then classified using supervised algorithms, such as support vector machine (Siuly and Li, 2012), Linear Discriminant Analysis (Garrett et al., 2003), and K-Nearest Neighbors (Yazdani et al., 2009). However, these traditional methods require handcrafted feature engineering, which complicates the decoding of EEG signals (Liu et al., 2020).

Deep learning provides a solution for the above limitation, and merges feature extraction and feature classification together. Schirrmeister et al. (Schirrmeister et al., 2017) designed two convolutional neural network (CNN) models, DeepConvNet and ShallowConvNet, for end-to-end EEG signal decoding and demonstrated promising decoding performance. Lawhern et al. (Lawhern et al., 2018) proposed EEGNet, which employs depth-wise separable convolutions

This work was supported in part by the Key Discipline Research Capability Enhancement Project in Guangdong Province under Grant 2024ZDJS033, and in part by the Guangdong University Innovation Team Program under Grant 2025KCXTD048.

*Corresponding author.
E-mail address: junhua.li@essex.ac.uk (J. Li).

to not only maintain high decoding performance but also reduce the number of model parameters, making it suitable for decoding tasks across various BCI paradigms. Santamaria-Vazquez et al. (Santamaría-Vázquez et al., 2020) designed EEG-inception, which employs convolutional kernels of different sizes to enhance decoding performance. To further improve the classification performance of CNN-based models, an attention mechanism is used to allocate higher weights to the critical information contained in EEG signals. Ye et al. (Ye et al., 2023) proposed AB-DPCGRU, an attention-based convolutional recurrent network that integrates spatial, spectral, and temporal EEG features for emotion recognition. Miao et al. (Miao et al., 2023) proposed LMDA-Net, which integrates deep attention, channel attention, and CNN. Altaheri et al. (Altaheri et al., 2023) introduced ATC-Net, which combines multi-head self-attention with temporal convolutional neural networks. However, these models have limited ability to establish reliable long-term dependencies. To address this issue, a better option might be the Transformer architecture. Many studies combine the advantages of CNN and Transformer to extract local and global feature representations (Ma et al., 2022; Song et al., 2023; Ding et al., 2025). Despite their advancements, these models remain constrained by specific limitations, such as the quadratic complexity of Transformer in long sequence modelling, which leads to low computational efficiency.

Lately, Gu et al. (Gu and Dao, 2024) proposed Mamba, a new architecture based on the State Space Model (SSM), which introduces a selective scanning mechanism that allows the model to dynamically choose valuable information based on the input. Mamba has shown superior performance across various domains, proving its efficiency in continuous long-sequence modelling (Yue and Li, 2024; He et al., 2025; Jiang et al., 2024; Wang et al., 2024; Zhu et al., 2024). Several studies have applied Mamba to EEG decoding. For example, Hu et al. (Hu et al., 2025) integrated Mamba with dynamic graph learning to capture long-range temporal dependencies and dynamic inter-channel relationships in EEG and ECG signals. Zhou et al. (Zhou et al., 2025) proposed miMamba to learn multi-scale spatio-temporal features for EEG-based emotion recognition. However, the application of Mamba in motor imagery EEG decoding remains relatively limited and requires further exploration.

To tackle the above issues, we proposed a collaborative network of bidirectional Mamba and convolutional attention feature selection (CN-BMCAFS) to classify multiple tasks of the unilateral lower limb. First, the shallow spatio-temporal feature encoder is employed to extract basic spatio-temporal features. The encoded features are fed into the global temporal bidirectional Mamba module (GT Bi-Mamba) to capture globally discriminative temporal features, followed by the convolutional attention feature selection module (CAFS) for capturing informative local temporal features. The GT Bi-Mamba module and the CAFS module are cascaded in a recurrent manner, enabling the combined advantage of GT Bi-Mamba capturing global temporal features while considering local temporal correlations and CAFS capturing local temporal features while considering global temporal correlations. Finally, the features are fed into the classifier module to obtain the final classification results. The main contributions of this study are as follows.

- We proposed CN-BMCAFS for decoding EEG signals of multiple tasks of the unilateral lower limb. In this model, the GT Bi-Mamba module and CAFS module are cascaded in a recurrent manner, enabling the model to recurrently couple global temporal features and local temporal features, thereby generating more robust EEG feature representations.
- The GT Bi-Mamba module is designed to extract globally discriminative temporal features, enhancing the model's ability to accurately distinguish features among MI, ME, and MO.
- We developed a CAFS module to capture informative local temporal features, complementing the shortcomings of the GT Bi-Mamba module in local modelling capability.

The remainder of this paper is structured as follows. A detailed description of the CN-BMCAFS is given in Section 2. The experiment details are presented in Section 3. Section 4 provides the results and discussions. Finally, a conclusion and future work are given in Section 5.

2. Proposed Model

The architecture of CN-BMCAFS for classifying multiple tasks of the unilateral lower limb is shown in Fig. 1. CN-BMCAFS is capable of recurrently coupling global temporal features and local temporal features, thereby generating more robust EEG feature representations. The model consists of four components: shallow spatio-temporal feature encoder, global temporal bidirectional Mamba (GT Bi-Mamba) module, convolutional attention feature selection (CAFS) module and classifier module. We presented each module in detail below.

2.1. Shallow Spatio-Temporal Feature Encoder

We design a shallow spatio-temporal feature encoder based on CNNs to extract the basic spatio-temporal features of EEG signals. The first layer employs $F(F = 16)$ kernels with the size of $(1, 64)$ and a stride of 1 to capture temporal-dependent features. The second layer includes F kernels with the size of $(E_l, 1)$ and a stride of 1, where E_l is the number of EEG electrodes, to capture interaction information among EEG electrodes (Wang et al., 2022a). Subsequently, batch normalization is applied to improve training stability and the generalization of the model. The exponential linear unit (ELU) is used as the activation function for nonlinearity. The third layer is an average pooling along the time dimension with the kernel size of $(1, 85)$ and a stride of 15. This average pooling layer is employed to enhance feature smoothing and decrease computational complexity. Finally, we squeeze the electrode dimension of 1 and swap between

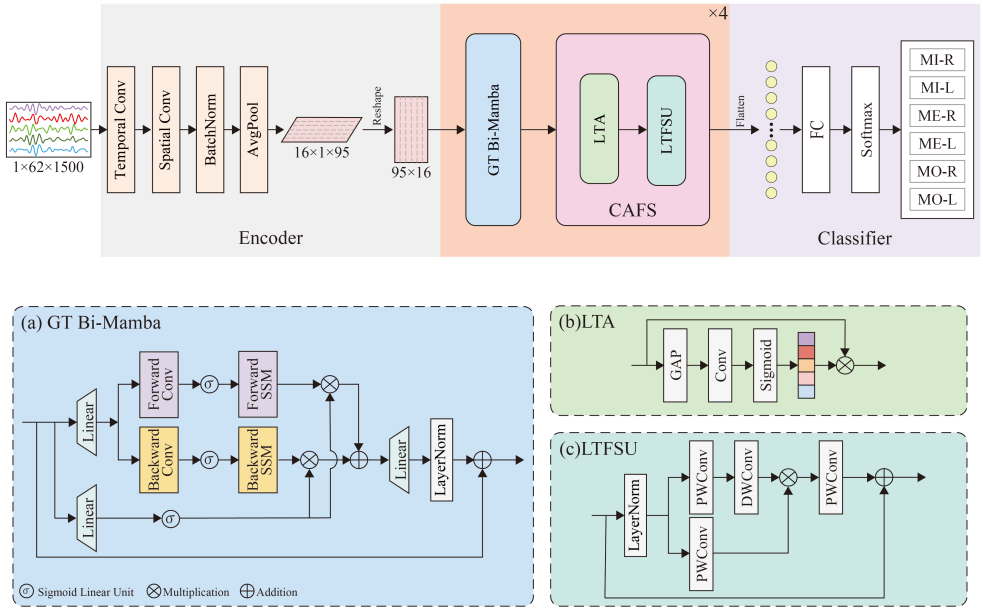


Fig. 1. Architecture of CN-BMCAFS. CN-BMCAFS comprises a shallow spatiotemporal feature encoder, a global temporal bidirectional Mamba (GT Bi-Mamba) module, a convolutional attention feature selection (CAFS) module, and a classifier. CAFS comprises two components: local temporal attention (LTA) and a local temporal feature selection unit (LTFSU). The detailed architectures of GT Bi-Mamba, LTA, and LTFSU are shown in panels (a), (b), and (c), respectively.

the filter dimension and the time dimension. The parameter settings are summarized in Table 1.

Table 1
Parameter settings

Layer	Input	Parameter	Output
Temporal Conv	$(1, 62, 1500)$	Kernel: $(1, 64)$ Stride: 1 Padding: same	$(F, 62, 1500)$
Spatial Conv	$(F, 62, 1500)$	Kernel: $(E_l, 1)$ Stride: 1 Padding: 0	$(F, 1, 1500)$
Average Pooling	$(F, 1, 1500)$	Kernel: $(1, 85)$ Stride: 15 Padding: 0	$(F, 1, 95)$

2.2. Global Temporal Bidirectional Mamba Module

We employ GT Bi-Mamba, which integrates forward and backward Mambas, to capture globally discriminative temporal features. As in Mamba, SSM, which is inspired by control systems, can map a one-dimensional sequence $x(t) \in \mathbb{R}$ to $y(t) \in \mathbb{R}$ through a hidden state $h(t) \in \mathbb{R}^N$. The above process can be expressed as linear ordinary differential equations:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t) \tag{1}$$
$$y(t) = \mathbf{C}h(t) \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$, $\mathbf{B} \in \mathbb{R}^{N\times 1}$ and $\mathbf{C} \in \mathbb{R}^{1\times N}$ are state transition matrices, with a state size of N.

Mamba uses discretization methods to embed continuous systems into deep learning. Specifically, a time scale parameter Δ is employed to transform continuous parameters $\mathbf{A}$ and $\mathbf{B}$ from the continuous systems into the discrete parameters $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ using zero-order hold method, defined as follows:

$$\overline{\mathbf{A}} = \exp(\Delta\mathbf{A}) \tag{3}$$
$$\overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B} \tag{4}$$

After discretizing $\mathbf{A}$ and $\mathbf{B}$, Eq. (1) and Eq. (2) using a time scale parameter Δ can be redefined as follows:

$$h_t = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t \tag{5}$$
$$y_t = \mathbf{C}h_t \tag{6}$$

Finally, the model's output is obtained through a global convolution, which can be represented as:

$$\overline{\mathbf{K}} = (\mathbf{C}\overline{\mathbf{B}}, \mathbf{C}\overline{\mathbf{A}\mathbf{B}}, \cdots, \mathbf{C}\overline{\mathbf{A}}^{L-1}\overline{\mathbf{B}}) \tag{7}$$
$$y = x \star \overline{\mathbf{K}} \tag{8}$$

where L is the length of the input sequence x, $\overline{\mathbf{K}} \in \mathbb{R}^L$ is a structured convolution kernel, and $\star$ represents the convolution operation.

The pseudocode of GT Bi-Mamba module is presented in Algorithm 1. The output $x_t \in \mathbb{R}^{95\times 16}$ from shallow spatiotemporal feature encoder is fed to GT Bi-Mamba module.

Algorithm 1 GT Bi-Mamba

Require: $x_t : (S, T, F)$; Batch Size $S = 16$; Number of EEG Time Points $T = 95$; Number of Filters $F = 16$.
Ensure: $x_m : (S, T, F)$
1: $o : (S, T, E) \leftarrow \text{Linear}(x_t)$
2: $z : (S, T, E) \leftarrow \text{Linear}(x_t)$
3: **Process GT Bi-Mamba**
4: **for** $i \in \{\text{forward}, \text{backward}\}$ **do**
5: $o' : (S, T, E) \leftarrow \text{SiLU}(\text{Conv1d}_i(o))$
6: $B_i : (S, T, N) \leftarrow \text{Linear}_i(o')$
7: $C_i : (S, T, N) \leftarrow \text{Linear}_i(o')$
8: *Softplus ensures positive* Δ_i
9: $\Delta_i : (S, T, E) \leftarrow \log(1 + \exp(\text{Linear}_i(o') + \text{Parameter}_i^{\Delta}))$
10: *Shape of* Parameter_i^A *is* (E, N)
11: $\overline{A}_i : (S, T, E, N) \leftarrow \Delta_i \otimes \text{Parameter}_i^A$
12: $\overline{B}_i : (S, T, E, N) \leftarrow \Delta_i \otimes B_i$
13: $y_i : (S, T, E) \leftarrow \text{SSM}(\overline{A}_i, \overline{B}_i, C_i)(o')$
14: **end for**
15: $x_m : (S, T, F) \leftarrow x_t + \text{Norm}(\text{Linear}(\sum_{i \in \{\text{forward,backward}\}} y_i \odot \text{SiLU}(z)))$
16: **return** x_m

First, x_t is linearly projected into o and z, expanding the filter dimension to E (Lines 1∼2 in Algorithm 1). Subsequently, o undergoes further processing through 1D convolution and SILU activation to obtain o' (Line 5 in Algorithm 1). o' generates state representations y_i by discretized SSM (Lines 6∼13 in Algorithm 1). Finally, y_i is gated by z (activated by SiLU), followed by a linear layer and a normalization layer to produce the output $x_m \in \mathbb{R}^{95\times16}$ (Line 15 in Algorithm 1).

In addition, the hyperparameter settings of GT Bi-Mamba are summarized in Table 2. The state size and expansion dimension are set to 16 and 32, respectively; the convolution kernel size is 4; Layer Normalization is adopted; and four GT Bi-Mamba blocks are used.

Table 2
The hyperparameter settings of GT Bi-Mamba

State Size	Expansion Dimension	Convolution Kernel	Normalization	Number of Blocks
16	32	4	Layer Norm	4

2.3. Convolutional Attention Feature Selection Module

The CAFS module is utilized to capture informative local temporal features, supplementing the deficiency of GT Bi-Mamba module in local modeling capability. As shown in Fig. 1, the CAFS consists of two components: local temporal attention (LTA) and the local temporal feature selection unit (LTFSU). The output $x_m \in \mathbb{R}^{95\times16}$ from GT Bi-Mamba is fed to CAFS. The CAFS captures local temporal features while considering global temporal correlations. The whole process of CAFS can be represented as follows:

$$x_{ma} = \text{LTA}(x_m) \tag{9}$$
$$x_{mc} = \text{LTFSU}(x_{ma}) \tag{10}$$

In LTA, x_m is first processed by average pooling with a convolution kernel of size 16 and a stride of 1 along the filter dimension, compressing it into a temporal feature map (i.e., $P_{avg}^t \in \mathbb{R}^{95\times1}$). Subsequently, we apply a convolution along the temporal dimension with a kernel size of 7 and a stride of 1, followed by a Sigmoid activation function, producing a temporal attention map $M_t \in \mathbb{R}^{95\times1}$. Finally, we obtain refined EEG features $x_{ma} \in \mathbb{R}^{95\times16}$ by element-wise multiplying x_m with M_t. The process can be mathematically expressed as:

$$M_t = \text{Sigmoid}(f(P_{avg}^t)) \tag{11}$$
$$x_{ma} = x_m \odot M_t \tag{12}$$

where $f(\cdot)$ denotes convolution and $\odot$ denotes element-wise multiplication.

LTFSU integrates gated linear unit and depth-wise separable convolutions to enable an adaptive gating mechanism and reduce the computational cost by reducing the number of parameters. The output $x_{ma} \in \mathbb{R}^{95\times16}$ from LTA is fed into LTFSU. First, x_{ma} is normalized using the layer normalization function, followed by linear projections into q_i and q_j. Subsequently, q_i is processed by a depth-wise convolution with a kernel size of 5 and a stride of 1 to capture local temporal features, and is then gated by q_j to obtain q_k. Finally, q_k is processed by a point-wise convolution to integrate the feature maps, and is then integrated into x_{ma} to produce the final output $x_{mc} \in \mathbb{R}^{95\times16}$. The above process in LTFSU can be expressed:

$$q_k = f_{DW}(q_i) \odot q_j \tag{13}$$

$$x_{mc} = x_{ma} + f_{PW}(q_k) \tag{14}$$

where $f_{PW}(\cdot)$ denotes point-wise convolution and $f_{DW}(\cdot)$ denotes depth-wise convolution.

2.4. Recurrent Cascade Mechanism

GT Bi-Mamba and CAFS are cascaded in a recurrent manner over four stages, enabling the model to repeatedly couple global and local temporal EEG features and thereby generate more robust EEG feature representations. Specifically, in the first stage, the output $x_t \in \mathbb{R}^{95\times16}$ of the shallow spatio-temporal feature encoder is fed into the first GT Bi-Mamba to obtain $x_m^1 \in \mathbb{R}^{95\times16}$. Subsequently, x_m^1 is processed by the first CAFS to generate $x_{mc}^1 \in \mathbb{R}^{95\times16}$. The first stage is expressed as:

$$x_m^1 = \text{GT Bi-Mamba}_1(x_t) \tag{15}$$

$$x_{mc}^1 = \text{CAFS}_1(x_m^1) \tag{16}$$

In the remaining three stages, the CAFS output of the preceding stage is used as the input to the GT Bi-Mamba module of the subsequent stage. This process is formulated as:

$$x_m^i = \text{GT Bi-Mamba}_i(x_{mc}^{i-1}), \quad i = 2, 3, 4 \tag{17}$$

$$x_{mc}^i = \text{CAFS}_i(x_m^i), \quad i = 2, 3, 4 \tag{18}$$

During these stage-wise interactions, GT Bi-Mamba considers the local temporal information captured by the CAFS in the preceding stage while extracting global temporal features. In turn, CAFS considers the global temporal information captured by the corresponding GT Bi-Mamba while extracting local temporal features. After four stages, x_{mc}^4 is fed into the classifier module.

2.5. Classifier

The Classifier Module consists of a fully connected layer and a softmax layer. The final classification results of CNBMCAFS are generated in this module.

3. Experiment Details

3.1. Data Description

The dataset used in this study consists of 28 healthy subjects, who performed six actions according to cues displayed on a computer monitor. The actions are MI of left lower limb (MI-L), MI of right lower limb (MI-R), ME of left lower limb (ME-L), ME of right lower limb (ME-R), MO of left lower limb (MO-L), and MO of right lower limb (MO-R). There are 72 trials per action, resulting in a total of 432 trials for six actions. Sixty-two electrodes were used to record EEG signals, and two pairs of electrodes were used to record electrooculography signals. All signals were recorded at a sampling rate of 500 Hz and downsampled to 250 Hz. The study was reviewed and approved by the Humanities, Science and Health, or Social Science Ethics Sub-Committee at the University of Essex. More details about the dataset can be found in (Wang et al., 2022b, 2023b). In this study, the EEG segments of 6 seconds after the onset of cues were extracted. Meanwhile, a band-pass filter with a frequency range of 8 $\sim$ 30 Hz was applied to them.

3.2. Compared Models

We compare the proposed model with seven models, which represent either classical or cutting-edge methods in EEG decoding.

- **FBCSP-SVM:** FBCSP-SVM is a classical MI decoding method. It extracts the most discriminative spatial features by selecting the optimal CSPs on multiple sub-bands, then uses a support vector machine for classification.
- **Deep ConvNet:** Deep ConvNet consists of a convolutional-pooling layer with spatio-temporal filtering, followed by three standard convolutional-max-pooling blocks, and a final dense softmax classification layer.
- **EEGNet:** EEGNet is a compact CNN-based network that learns spatio-temporal features using depth-separable convolution and point-wise convolution. It is fast to train, performs well, and is applicable to different BCI paradigms.
- **EEG-Inception:** EEG-Inception uses Inception modules to extract multi-scale features, achieving superior performance while maintaining a lightweight architecture.
- **LMDA-Net:** LMDA-Net consists of a convolutional neural network and two attention modules. The model enhances classification performance by integrating channel attention and deep attention.
- **ATCNet:** ATCNet consists of three main modules. The convolution module encodes the EEG low-level spatio-temporal information. The attention module highlights the most important features in the time series using multi-head self-attention. The temporal convolution module extracts high-level temporal features.
- **EEG-Deformer:** EEG-Deformer captures coarse-to-fine temporal dynamics in EEG signals through a hierarchical coarse-to-fine Transformer module. A dense information purification module is designed to adaptively fuse the features generated by each Transformer module.

3.3. Performance Evaluation Metrics

In our study, the Wilcoxon paired signed rank test (Woolson, 2005) is used to assess whether the performance differences between the CN-BMCAFS and the compared models are statistically significant, with false discovery rate correction applied to the p-values. Classification accuracy, Cohen's kappa coefficient, and macro-F1 are used as evaluation metrics and are calculated as follows:

$$\text{Accuracy} = \frac{\sum_{i=1}^{6} M_{ii}}{\sum_{i=1}^{6}\sum_{j=1}^{6} M_{ij}} \tag{19}$$

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \tag{20}$$

$$\text{F1}_{macro} = \frac{1}{6}\sum_{i=1}^{6} \frac{2TP_i}{2TP_i + FP_i + FN_i} \tag{21}$$

where M_{ij} denotes the number of samples with true class i and predicted class j, while M_{ii} denotes the number of correctly classified samples in class i. P_o is the actual agreement between the predicted and true labels, and P_e is the expected agreement based on the marginal class distributions. TP_i, FP_i, and FN_i denote the numbers of true-positive, false-positive, and false-negative samples for class i, respectively.

3.4. Training Setting

In our proposed model, cross-entropy loss is used for training. We use an Adam optimizer with a learning rate of 0.0001, and optimization parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The batch size is set to 16.

The compared models follow the training settings as described in their respective papers. EEG-Inception and EEGNet models are proposed for data with a sampling rate of 128 Hz. Therefore, we adjust the lengths of temporal convolution kernels and temporal pooling layers in both architectures by a factor of 2, consistent with previous work (Mane et al., 2021), to approximately match with the sampling rate of 250 Hz.

All models are evaluated using subject-specific five-fold cross-validation. For each subject, the trials are randomly divided into five folds. In each iteration, four folds are used for training and the remaining fold for testing, such that each fold serves as the testing set once. The data splitting is identical for all models. Before training, Z-score normalization is performed using only the training set, and the obtained normalization parameters are applied to the testing set. The number of training epochs is set to 200. All experiments are conducted using the PyTorch framework on a single NVIDIA GeForce RTX 4060 GPU.

4. Results and Discussions

4.1. Performance Evaluation

The training loss and accuracy curves of CN-BMCAFS are shown in Fig. 2. As training progresses, the loss gradually decreases, while the accuracy steadily increases. Both curves eventually stabilize, indicating that CN-BMCAFS with a recurrent cascade of GT Bi-Mamba and CAFS converges stably during training.

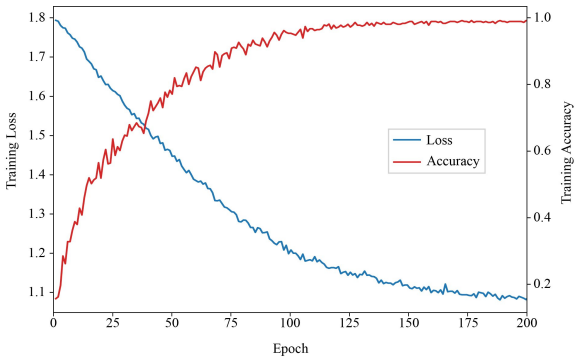


Fig. 2. Loss and accuracy during training of CN-BMCAFS.

Fig. 3 shows the classification accuracies of different models. The classification accuracies are obtained by averaging across all subjects. It can be seen that CN-BMCAFS achieves an accuracy of 55.4%, outperforming all compared models. The performance differences between CN-BMCAFS and all compared models are statistically significant. Except for EEG-Deformer, deep learning-based models surpass FBCSP-SVM in classification accuracy. Compared to the second-best ATCNet, CN-BMCAFS improves classification accuracy by 1.79%. Compared to FBCSP-SVM, CN-BMCAFS improves classification accuracy by 13.67%. Compared to the CNN-based methods such as DeepConvNet, EEGNet and EEG-Inception, CN-BMCAFS improves classification accuracy by 11.09%, 3.28%, and 4.68%, respectively. Compared to the transformer-based EEG-Deformer, CN-BMCAFS improves classification accuracy by 14.5%. The detailed classification accuracies for each subject are listed in Table 3. From Table 3, it can be seen that CN-BMCAFS outperforms all compared models across 18 subjects, demonstrating particularly significant improvements in Subjects 15 and 16, with classification gains of at least 9.5% and 9.99%, respectively. The Cohen's kappa and macro-F1 results of all compared models are presented in Fig. 4. CN-BMCAFS achieves the highest Cohen's kappa and macro-F1 values of 0.4648 and 0.5492, respectively, outperforming all compared models. The superior performance of CN-BMCAFS is attributed to the effective capture of global and local temporal features by GT Bi-Mamba and CAFS, respectively.

To further evaluate the computational efficiency of the proposed model, we compare the computational complexity among different models. The results are presented in Table 4. As shown in Table 4, compared with ATCNet, which achieves the second-best classification performance, CN-BMCAFS requires only approximately 36.9% of the parameters, 29.2% of the inference time, and 16.9% of the

training time. Furthermore, compared with the Transformer-based EEG-Deformer, CN-BMCAFS requires only approximately 2.9% of the parameters, 23.2% of the training time, and 23.9% of the memory usage. These results demonstrate that CN-BMCAFS maintains high classification performance while preserving favorable computational efficiency.

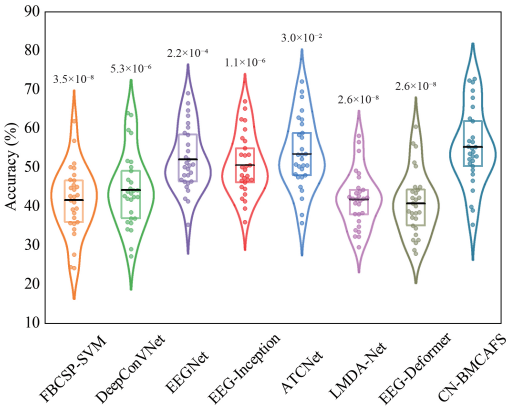


Fig. 3. Classification accuracies of CN-BMCAFS and the compared models. Colored dots denote subject-wise accuracies, boxes represent the interquartile ranges, and black lines indicate the mean values. P-values are annotated above the corresponding violins.

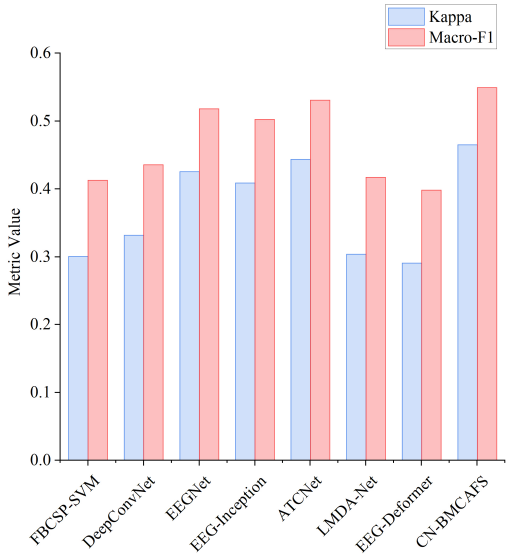


Fig. 4. The kappa and macro-F1 of the CN-BMCAFS and the compared models.

4.2. Analysis of MI, ME, and MO Categories

The confusion matrices of all models are presented in Fig. 5. It can be observed that CN-BMCAFS excels in classifying both MI and ME categories. Compared to other models, it achieves the highest classification accuracy for MI of left lower limb, MI of right lower limb, ME of left lower limb, and ME of right lower limb. For ME of left lower limb, CN-BMCAFS shows the most significant improvement in classification accuracy, achieving at least a 4.17% increase compared to the other models. For MO of left lower limb, CN-BMCAFS achieves lower classification accuracy than ATCNet and DeepConvNet, but outperforms FBCSP-SVM, EEGNet, EEG-Inception, LMDA-Net, and EEG-Deformer. For MO of right lower limb, CN-BMCAFS achieves slightly lower accuracy than ATCNet, but outperforms FBCSP-SVM, DeepConvNet, EEGNet, EEG-Inception, LMDA-Net, and EEG-Deformer. Fig. 6 presents the number of subjects for whom one category demonstrates superior classification performance over the other. For instance, if ME outperforms MI for a given subject, the count for ME is incremented by one. The results indicate that the classification accuracy of ME consistently exceeds that of MI across all subjects. It also outperforms MO for the majority of subjects. It is not surprising to observe the highest performance in ME because event related desynchronization induced by ME is usually greater in magnitude and longer in duration (Duann and Chiou, 2016). MO outperforms MI in classification accuracy among 23 subjects. This is likely because MO involves the observation of the actual action execution, which triggers the mirror neuron system in a more natural manner compared to MI (Gatti et al., 2013).

4.3. Ablation Study

The GT Bi-Mamba and CAFS modules constitute the core components of CN-BMCAFS. To investigate their individual contributions to classification accuracy, we conduct an ablation study. The results are presented in Table 5. Compared with the configuration containing neither module, introducing GT Bi-Mamba improves the classification accuracy by 1.82%. Further incorporating CAFS into GT Bi-Mamba yields an additional improvement of 1.75%. When only CAFS is included, the model may have difficulty learning more discriminative feature representations by repeatedly extracting local temporal features. These results indicate that GT Bi-Mamba captures globally discriminative temporal features through forward and backward Mambas, whereas CAFS compensates for the limitation of GT Bi-Mamba in extracting local temporal features. The collaboration between the two modules enables more effective integration of global and local temporal information, thereby further improving classification performance.

4.4. Visualization

In order to visually assess the effect of GT Bi-Mamba on the extraction of globally discriminative temporal features from EEG signals, we use t-distributed stochastic neighbour embedding (Van Der Maaten, 2014) to map the captured high-dimensional features onto a two-dimensional space. The visualization of the features extracted by GT Bi-Mamba for an example subject (subject 27) is presented in Fig. 7. It can be observed that there is larger overlap among tasks when GT Bi-Mamba is not used. When GT Bi-Mamba is used, distinct boundaries are observed among MI, ME, and

Table 3
The classification accuracy of each subject for the CN-BMCAFS and the compared models. Bold text emphasizes the best results.

Subjects \ Models	FBCSP-SVM	Deep ConvNet	EEGNet	EEG-Inception	ATCNet	LMDA-Net	EEG-Deformer	CN-BMCAFS
01	56.99	59.77	66.68	65.29	69.67	54.41	55.57	**69.91**
02	46.78	37.05	47.01	46.33	47.68	38.21	31.49	**55.56**
03	38.42	42.58	46.07	49.54	50.45	38.17	38.41	**51.86**
04	44.45	63.65	63.44	56.93	62.98	58.32	60.64	**65.76**
05	45.15	34.27	44.22	41.45	**50.69**	32.41	31.49	44.21
06	37.25	42.11	48.14	47.44	50.68	38.42	38.40	**52.29**
07	33.09	42.82	45.16	44.67	48.39	40.96	37.26	**49.53**
08	42.60	43.51	50.69	48.16	49.78	43.54	42.37	**53.69**
09	35.66	43.97	51.37	53.23	53.46	44.42	40.26	**54.16**
10	45.62	41.89	52.78	50.48	52.31	48.62	38.88	**53.02**
11	39.62	42.58	50.23	50.22	**53.47**	44.90	41.87	51.16
12	50.01	51.61	54.38	53.45	**57.86**	40.74	42.34	56.92
13	24.30	34.03	46.54	46.76	48.15	34.04	30.79	**52.99**
14	44.90	51.86	61.56	62.52	**68.28**	47.69	45.15	62.71
15	43.07	49.28	58.54	53.46	58.55	43.97	44.43	**68.05**
16	47.00	59.06	56.01	59.95	62.49	41.67	43.95	**72.48**
17	35.88	35.89	46.52	45.14	45.14	42.13	37.04	**46.77**
18	24.55	29.17	**42.13**	39.60	37.96	33.58	28.93	39.13
19	27.77	27.31	35.43	**36.11**	35.86	29.64	28.01	35.43
20	48.19	44.69	49.75	47.95	**50.92**	42.37	43.52	50.48
21	34.24	37.03	41.43	**42.13**	**42.13**	32.39	34.50	40.04
22	51.16	50.22	**59.02**	55.12	59.01	49.31	51.40	57.87
23	42.83	38.66	48.16	46.99	44.23	34.97	32.88	**48.38**
24	39.36	44.20	50.92	46.98	54.84	36.33	38.64	**55.09**
25	62.06	64.14	69.24	62.28	**72.25**	56.71	56.31	72.03
26	41.20	46.98	60.42	57.65	61.33	42.81	45.13	**62.03**
27	50.23	46.31	64.80	67.12	64.80	42.15	50.48	**72.93**
28	36.12	36.09	48.84	43.27	47.69	42.57	35.17	**56.71**

Table 4
Comparison of computational efficiency among different models.

Model	Parameters(k)	Inference Time(ms/trial)	Train Time (s/subject)	Memory Usage(MB)
FBCSP-SVM	5.46(CPU)	1.34(CPU)	196.43(CPU)	-
DeepConvNet	319.1	0.96	204.87	10.60
EEGNet	2.52	0.61	452.24	6.05
EEG-Inception	27.99	1.43	1188.66	8.62
ATCNet	115.34	19.31	4881.6	12.22
LMDA-Net	18.88	1.35	1649.91	20.66
EEG-Deformer	1476.51	3.17	3555.97	53.08
CN-BMCAFS	42.55	5.64	825.35	12.66

MO, though a certain overlap remains between the left lower limb task and the right lower limb task within each category.

To quantitatively assess feature discriminability, we calculate the distances between class centers from the learned features before the Classifier and average them across all 28 subjects, as shown in Fig. 8. Without GT Bi-Mamba, the class centers of the six tasks are relatively close to each other. After introducing GT Bi-Mamba, the distances among them increase substantially, indicating enhanced separability among the six tasks.

Furthermore, Fig. 8 shows that MI is positioned between ME and MO, whereas ME and MO exhibit greater separation. ME involves actual motor commands together with proprioceptive and somatosensory feedback generated during movement. In contrast, MO mainly involves visual action processing and mirror neuron system activity without actual motor output or movement-related sensory feedback. MI involves the internal simulation and planning of an action and thus shares motor-planning processes with ME. Meanwhile, similar to MO, MI does not involve overt movement or the associated peripheral sensorimotor feedback. These shared

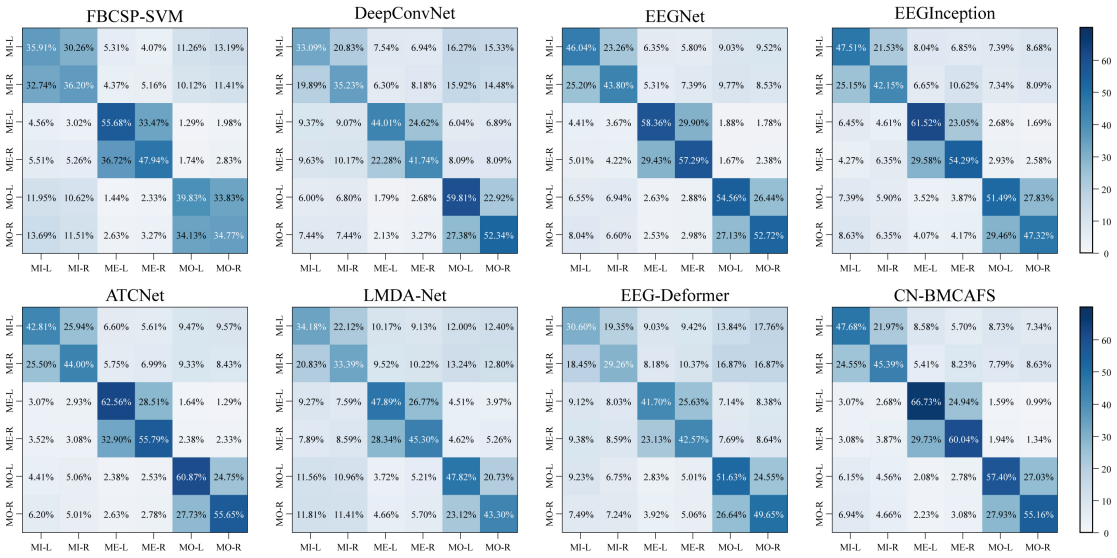


Fig. 5. Confusion matrices for different models. The numbers are the accuracies averaged across all subjects.

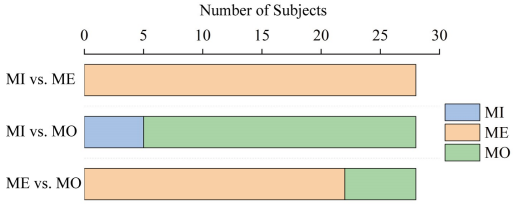


Fig. 6. The paired comparison results in terms of the number of subjects for whom one category demonstrates superior classification performance over the other. In the comparison between MI and ME, ME outperforms MI for all subjects. While, MO and ME outperform for the majority of subjects in the comparisons of MI vs. MO and ME vs. MO, respectively.

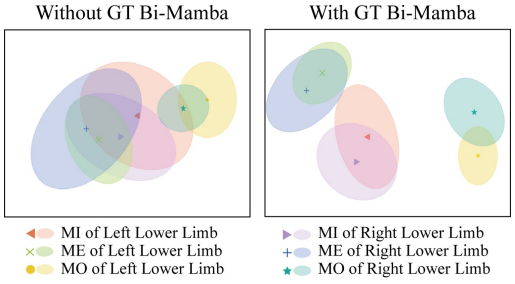


Fig. 7. Visualization of the features extracted by the GT Bi-Mamba. The symbols indicate the centers of each task. The color-coded areas delineate the extent of each task.

Table 5
Results of ablation study

GT Bi-Mamba	CAFS	Accuracy	Kappa	Macro-F1
×	×	51.83%	0.4220	0.5077
✓	×	53.65%	0.4437	0.5313
×	✓	50.82%	0.4054	0.4970
✓	✓	55.40%	0.4648	0.5492

characteristics may explain the intermediate position of MI, whereas the different physiological processes involved in ME and MO may account for their greater separation. These results indicate that GT Bi-Mamba improves the separability of the learned feature representations, thereby enhancing classification performance.

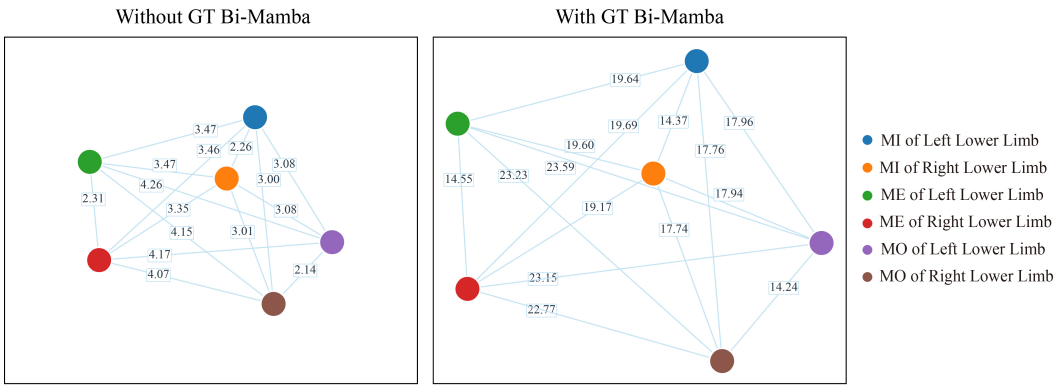
5. Conclusion and Future Work

5.1. Conclusion

In this paper, we proposed CN-BMCAFS for classifying multiple tasks of the unilateral lower limb. In CN-BMCAFS, the GT Bi-Mamba is employed to capture globally discriminative temporal features. The CAFS is designed to capture informative local temporal features. The GT Bi-Mamba and the CAFS collaborate to maximize the model performance. The results demonstrate that CN-BMCAFS achieves the highest classification accuracy and kappa, with values of 55.4% and 0.4648, respectively. Ablation study indicates that GT Bi-Mamba is a crucial component of the proposed model, improving the classification accuracy by 1.82%. Visualization result shows that GT Bi-Mamba enhances the model's ability to accurately distinguish features among MI, ME, and MO, verifying the effective role of bidirectional Mamba in EEG decoding.

5.2. Future Work

In future work, we will expand the dataset and conduct cross-subject studies. We will also investigate the integration of Mamba with other EEG decoding methods. For example, adaptive cognitive graphs and multi-domain constraint learning could be introduced to model dynamic relationships among brain regions (Gao et al., 2025). In addition, multilevel prototype constraints in hyperbolic space could be combined with bidirectional Mamba to learn hierarchical



Not to scale, see indicated distances.

Fig. 8. Visualization of the distances between class centers without and with the GT Bi-Mamba module, averaged across 28 subjects.

EEG representations and improve feature discriminability (Gao et al., 2026). These research directions will support further investigation into the classification of multiple tasks involving the unilateral lower limb.

References

Altaheri, H., Muhammad, G., Alsulaiman, M., 2023. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. IEEE Transactions on Industrial Informatics 19, 2249–2258. doi:10.1109/TII.2022.3197419.

Ang, K.K., Chin, Z.Y., Wang, C., Guan, C., Zhang, H., 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Frontiers in Neuroscience 6, 39. doi:10.3389/fnins.2012.00039.

Asanza, V., Peláez, E., Loayza, F., Lorente-Leyva, L.L., Peluffo-Ordóñez, D.H., 2022. Identification of lower-limb motor tasks via brain–computer interfaces: A topical overview. Sensors 22, 2028. doi:10.3390/s22052028.

Bhattacharyya, A., Pachori, R.B., 2017. A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. IEEE Transactions on Biomedical Engineering 64, 2003–2015. doi:10.1109/TBME.2017.2650259.

Bi, J., Chu, M., Wang, G., Gao, X., 2023. TSPNet: A time-spatial parallel network for classification of EEG-based multiclass upper limb motor imagery BCI. Frontiers in Neuroscience 17, 1303242. doi:10.3389/fnins.2023.1303242.

Bousseta, R., El Ouakouak, I., Gharbi, M., Regragui, F., 2018. EEG based brain computer interface for controlling a robot arm movement through thought. IRBM 39, 129–135. doi:10.1016/j.irbm.2018.02.001.

Chaudhary, S., Taran, S., Bajaj, V., Siuly, S., 2020. A flexible analytic wavelet transform based approach for motor-imagery tasks classification in BCI applications. Computer Methods and Programs in Biomedicine 187, 105325. doi:10.1016/j.cmpb.2020.105325.

Decety, J., 1996. The neurophysiological basis of motor imagery. Behavioural Brain Research 77, 45–52. doi:10.1016/0166-4328(95)00225-1.

Ding, Y., Li, Y., Sun, H., Liu, R., Tong, C., Liu, C., Zhou, X., Guan, C., 2025. EEG-Deformer: A dense convolutional transformer for brain-computer interfaces. IEEE Journal of Biomedical and Health Informatics 29, 1909–1918. doi:10.1109/JBHI.2024.3504604.

Duann, J.R., Chiou, J.C., 2016. A comparison of independent event-related desynchronization responses in motor-related brain areas to movement execution, movement imagery, and movement observation. PLOS ONE 11, e0162546. doi:10.1371/journal.pone.0162546.

Gao, D., Liu, M., Zhang, H., Wang, M., Chang, H., Ouyang, G., Liu, S., Li, P., 2025. A multi-domain constraint learning system inspired by adaptive cognitive graphs for emotion recognition. Neural Networks 188, 107457. doi:10.1016/j.neunet.2025.107457.

Gao, D., Ning, J., Peng, Z., Tang, M., Deng, A., Liu, S., Ding, X., Wang, M., Wang, L., Xie, L., Li, P., 2026. Multilevel prototype constraints based on hyperbolic space for EEG auditory attention decoding. Pattern Recognition 179, 113741. doi:10.1016/j.patcog.2026.113741.

Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H., 2003. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 141–144. doi:10.1109/TNSRE.2003.814441.

Gatti, R., Tettamanti, A., Gough, P.M., Riboldi, E., Marinoni, L., Buccino, G., 2013. Action observation versus motor imagery in learning a complex motor task: A short review of literature and a kinematics study. Neuroscience Letters 540, 37–42. doi:10.1016/j.neulet.2012.11.039.

Gu, A., Dao, T., 2024. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 doi:10.48550/arXiv.2312.00752.

He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L., 2025. MambaAD: Exploring state space models for multi-class unsupervised anomaly detection. arXiv preprint arXiv:2404.06564 doi:10.48550/arXiv.2404.06564.

Hu, J., Rahman, M.M.U., Laleg-Kirati, T.M., 2025. Adaptive long-range modeling of EEG and ECG with Mamba and dynamic graph learning. Scientific Reports 15, 38762. doi:10.1038/s41598-025-22684-x.

Jiang, X., Li, Y.A., Florea, A.N., Han, C., Mesgarani, N., 2024. Speech slytherin: Examining the performance and efficiency of Mamba for speech separation, recognition, and synthesis. arXiv preprint arXiv:2407.09732 doi:10.48550/arXiv.2407.09732.

Kim, J.C., Lee, H.M., 2021. EEG-based evidence of mirror neuron activity from app-mediated stroke patient observation. Medicina 57, 979. doi:10.3390/medicina57090979.

Kraeutner, S., Gionfriddo, A., Bardouille, T., Boe, S., 2014. Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations. Brain Research 1588, 81–91. doi:10.1016/j.brainres.2014.09.001.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: A compact convolutional network for EEG-based brain-computer interfaces. Journal of Neural Engineering 15, 056013. doi:10.1088/1741-2552/aace8c.

Li, J., Thakor, N., Bezerianos, A., 2018. Unilateral exoskeleton imposes significantly different hemispherical effect in parietooccipital region, but not in other regions. Scientific Reports 8, 13470. doi:10.1038/s41598-018-31828-1.

Li, P., Peng, M., Zhang, H., Liu, S., Gao, D., Qin, Y., Wu, D., Liu, T., 2026. An adaptive decoupling learning system informed by the brain functional structure for EEG decoding. Neural Networks 195, 108228. doi:10.1016/j.neunet.2025.108228.

Li, P., Zhang, Y., Liu, S., Lin, L., Zhang, H., Tang, T., Gao, D., 2023. An EEG-based brain cognitive dynamic recognition network for representations of brain fatigue. Applied Soft Computing 146, 110613.

doi:10.1016/j.asoc.2023.110613.

Liu, X., Shen, Y., Liu, J., Yang, J., Xiong, P., Lin, F., 2020. Parallel spatial–temporal self-attention CNN-based motor imagery classification for BCI. Frontiers in Neuroscience 14, 587520. doi:10.3389/fnins.2020.587520.

Ma, Y., Song, Y., Gao, F., 2022. A novel hybrid CNN-Transformer model for EEG motor imagery classification, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8. doi:10.1109/IJCNN55064.2022.9892821.

Mammone, N., Ieracitano, C., Morabito, F.C., 2020. A deep CNN approach to decode motor preparation of upper limbs from time–frequency maps of EEG signals at source level. Neural Networks 124, 357–372. doi:10.1016/j.neunet.2020.01.027.

Mane, R., Chew, E., Chua, K., Ang, K.K., Robinson, N., Vinod, A.P., Lee, S.W., Guan, C., 2021. FBCNet: A multi-view convolutional neural network for brain-computer interface. arXiv preprint arXiv:2104.01233 doi:10.48550/arXiv.2104.01233.

Miao, Z., Zhao, M., Zhang, X., Ming, D., 2023. LMDA-Net:a lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability. NeuroImage 276, 120209. doi:10.1016/j.neuroimage.2023.120209.

Nicolas-Alonso, L.F., Gomez-Gil, J., 2012. Brain computer interfaces, a review. Sensors 12, 1211–1279. doi:10.3390/s120201211.

Santamaría-Vázquez, E., Martínez-Cagigal, V., Vaquerizo-Villar, F., Hornero, R., 2020. EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering 28, 2773–2782. doi:10.1109/TNSRE.2020.3048106.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping 38, 5391–5420. doi:10.1002/hbm.23730.

Siuly, S., Li, Y., 2012. Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering 20, 526–538. doi:10.1109/TNSRE.2012.2184838.

Song, Y., Zheng, Q., Liu, B., Gao, X., 2023. EEG Conformer: Convolutional transformer for EEG decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering 31, 710–719. doi:10.1109/TNSRE.2022.3230250.

Van Der Maaten, L., 2014. Accelerating t-SNE using tree-based algorithms. The Journal of Machine Learning Research 15, 3221–3245.

Vuckovic, A., Sepulveda, F., 2008. Delta band contribution in cue based single trial classification of real and imaginary wrist movements. Medical & Biological Engineering & Computing 46, 529–539. doi:10.1007/s11517-008-0345-8.

Wang, T., Bezerianos, A., Cichocki, A., Li, J., 2022a. Multikernel capsule network for schizophrenia identification. IEEE Transactions on Cybernetics 52, 4741–4750. doi:10.1109/TCYB.2020.3035282.

Wang, Z., Chen, C., Li, J., Wan, F., Sun, Y., Wang, H., 2023a. ST-CapsNet: Linking spatial and temporal attention with capsule network for P300 detection improvement. IEEE Transactions on Neural Systems and Rehabilitation Engineering 31, 991–1000. doi:10.1109/TNSRE.2023.3237319.

Wang, Z., Daly, I., Li, J., 2023b. An evaluation of hybrid deep learning models for classifying multiple lower limb actions, in: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE. pp. 1–4. doi:10.1109/EMBC40787.2023.10340894.

Wang, Z., Kong, F., Feng, S., Wang, M., Yang, X., Zhao, H., Wang, D., Zhang, Y., 2024. Is Mamba effective for time series forecasting? arXiv preprint arXiv:2403.11144 doi:10.48550/arXiv.2403.11144.

Wang, Z., Li, J., Daly, I., Li, J., 2022b. Machine learning for multi-action classification of lower limbs for BCI, in: 2022 International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE. pp. 84–89. doi:10.1109/iCCECE55162.2022.9875092.

Woolson, R.F., 2005. Wilcoxon signed-rank test, in: Encyclopedia of Biostatistics. Wiley. volume 8. doi:10.1002/0470011815.b2a15177.

Yazdani, A., Ebrahimi, T., Hoffmann, U., 2009. Classification of EEG signals using Dempster Shafer theory and a k-nearest neighbor classifier, in: 2009 4th International IEEE/EMBS Conference on Neural Engineering, IEEE. pp. 327–330. doi:10.1109/NER.2009.5109299.

Ye, Z., Jing, Y., Wang, Q., Li, P., Liu, Z., Yan, M., Zhang, Y., Gao, D., 2023. Emotion recognition based on convolutional gated recurrent units with attention. Connection Science 35, 2289833. doi:10.1080/09540091.2023.2289833.

Yue, Y., Li, Z., 2024. MedMamba: Vision mamba for medical image classification. arXiv preprint arXiv:2403.03849 doi:10.48550/arXiv.2403.03849.

Zhang, R., Chen, Y., Xu, Z., Zhang, L., Hu, Y., Chen, M., 2023. Recognition of single upper limb motor imagery tasks from EEG using multi-branch fusion convolutional neural network. Frontiers in Neuroscience 17, 1129049. doi:10.3389/fnins.2023.1129049.

Zhou, X., Huang, D., Peng, X., Yin, L., 2025. miMamba: EEG-based emotion recognition with multi-scale inverted Mamba models. IEEE Transactions on Affective Computing 16, 3266–3278. doi:10.1109/TAFFC.2025.3587443.

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X., 2024. Vision Mamba: Efficient visual representation learning with bidirectional state space model, in: Proceedings of the 41st International Conference on Machine Learning, PMLR. pp. 62429–62442.