

Spatiotemporal Feature-Enhanced Bi-directional Mamba Network for Motor Imagery EEG Decoding

Ruiqing Li

Fujian Agriculture and Forestry University

Tingting Zhang

Fujian Agriculture and Forestry University

Zijian Li

Fujian Agriculture and Forestry University

Qi Deng

Fujian Agriculture and Forestry University

Dianyi Sun

Fujian Agriculture and Forestry University

Haorong Liao

Fujian Agriculture and Forestry University

Xiaohan Li

Fujian Agriculture and Forestry University

Jinghu Li

Fujian Agriculture and Forestry University

Junfeng Han

Fujian Agriculture and Forestry University

Fan Wang

Fujian Normal University

Guoyu Hu

Fujian Agriculture and Forestry University

Xiangzeng Kong

xzkongfafu@126.com

Fujian Agriculture and Forestry University

Article

Keywords: Brain-computer interface, Motor imagery, Electroencephalography, Mamba, Bidirectional state space model, Three-view statistical pooling

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10019404/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Spatiotemporal Feature-Enhanced Bi-directional Mamba Network for Motor Imagery EEG Decoding

Ruiqing Li^{1,+}, Tingting Zhang^{1,2,+}, Zijian Li³, Qi Deng¹, Diany Sun³, Haorong Liao¹, Xiaohan Li¹, Jinghu Li¹, Junfeng Han³, Fan Wang^{4,*}, Guoyu Hu^{1,5,*}, and Xiangzeng Kong^{1,*}

¹Institute of Artificial Intelligence in Agriculture and Forestry, College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China

²School of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou 350002, China

³College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China

⁴College of Physical Education, Fujian Normal University 350117 China

⁵College of Animal Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

*fanwang@fjnu.edu.cn (F.W.); hgy202603@163.com (G.H.); xzkongfafu@126.com (X.K.).

⁺These authors contributed equally to this work.

ABSTRACT

The accurate decoding of motor imagery electroencephalography (MI-EEG) signals is critical for brain-computer interface (BCI) applications in motor rehabilitation. However, existing methods often suffer from insufficient spatiotemporal feature fusion, high computational complexity, and limited generalization capability. To address these limitations, this paper proposes STE-BiMamba, a spatiotemporal feature-enhanced bidirectional Mamba model for efficient MI-EEG decoding. The model incorporates three core components: a multi-scale temporal feature extraction module that employs parallel convolutions with different kernel sizes to simultaneously capture transient event-related desynchronization and long-term rhythmic patterns, a three-view statistical pooling module that extracts highly discriminative spatiotemporal representations from mean, variance, and maximum perspectives, and a bidirectional Mamba global sequence modeling module that efficiently captures bidirectional long-range dependencies with linear time complexity. Extensive experiments on our self-collected VR-MI dataset and three public datasets (BCIC-IV-2a, BCIC-IV-2b, HGD) under both within-subject cross-session and cross-subject settings validate the effectiveness of STE-BiMamba. On VR-MI, the model achieved 97.50% accuracy (cross-session) and 85.05% (cross-subject), while attaining state-of-the-art results on all public datasets. Compared with Transformer-based models, STE-BiMamba doubles the inference speed while reducing the number of parameters by a factor of 3.5. STE-BiMamba offers a high-precision and lightweight solution for practical motor imagery BCI applications.

Keywords: Brain-computer interface; Motor imagery; Electroencephalography; Mamba; Bidirectional state space model; Three-view statistical pooling

Introduction

Brain-Computer Interface (BCI) is an emerging technology that enables direct interaction between the human brain and external devices via brain signals^{1,2}. Non-invasive electroencephalography (EEG), due to its relatively convenient acquisition and high safety, has become the most commonly used signal acquisition method in BCI research³. Currently, EEG-driven BCI systems have been applied to various external devices, including robotic arms⁴, exoskeletons⁵, and spelling systems⁶. Among different brain activity paradigms, Motor Imagery (MI) is of particular interest because it can generate distinguishable signals in the motor cortex without external stimulation⁷. Studies have shown that MI-based BCIs can effectively assist in the motor rehabilitation of paralyzed patients^{8,9}, yet accurate decoding of MI-EEG signals remains a major challenge due to physiological factors such as impedance¹⁰.

Early studies primarily relied on traditional machine learning methods, such as Linear Discriminant Analysis (LDA)¹¹ and Support Vector Machine (SVM)¹², for EEG signal classification. Due to the low signal-to-noise ratio of EEG, preprocessing and feature extraction are generally required to improve classification performance. Among these methods, Common Spatial Patterns (CSP)¹³ is the most widely used feature extraction algorithm, which highlights spatial differences between different signal classes. Further improvements, such as Filter Bank CSP (FBCSP)¹⁴, employ feature selection strategies to significantly enhance classification performance. However, these approaches heavily depend on manually designed features and face

limitations when handling large-scale datasets.

With the development of deep learning, along with the support of GPUs and large-scale datasets, EEG decoding research has rapidly advanced in recent years. Convolutional Neural Networks (CNNs)¹⁵ have become the mainstream method due to their excellent local feature extraction capability. Models such as EEGNet¹⁶, Shallow ConvNet¹⁷, and Deep ConvNet¹⁸ have demonstrated strong performance across multiple brain activity paradigms. However, CNNs are limited by convolution kernel sizes, restricting their ability to extract temporal features, and existing methods often lack effective spatiotemporal feature fusion, making it difficult to fully capture the complex spatiotemporal dependencies of EEG signals^{19,20}. To address these issues, researchers have introduced Recurrent Neural Networks (RNNs)²¹, particularly Long Short-Term Memory networks (LSTMs)²², which can effectively model long-term dependencies in sequences and have shown significant advantages in EEG classification tasks. Nevertheless, RNNs still suffer from limited parallel computing efficiency²³. In contrast, Transformers²⁴ utilize self-attention mechanisms to capture global dependencies in sequential signals and have shown good performance in EEG modeling. However, their quadratic time complexity leads to high computational costs²⁵, limiting practical applications in real-time BCI systems.

Recently, the state-space model (SSM)-based Mamba architecture²⁶ introduced a selective mechanism similar to attention, which can effectively identify important information in sequences. Compared with Transformers, Mamba demonstrates superior performance and computational efficiency in various tasks, especially for long-sequence modeling, and shows significant advantages in time series prediction²⁷.

Considering all these challenges, this paper proposes a novel end-to-end deep learning network, named the Spatiotemporal Feature Enhanced Bi-directional Mamba model (STE-Mamba). The model is designed to fully leverage multi-scale temporal information and global spatial features and to improve MI-EEG decoding performance through efficient global sequence modeling while maintaining linear-time computational complexity. The network consists of four main modules: a multi-scale spatiotemporal feature extraction module, a tri-view statistical pooling module, a bi-directional Mamba global sequence modeling module, and a classification module. In the multi-scale spatiotemporal feature extraction module, convolutional operations are first employed to jointly extract multi-scale temporal features and global spatial features. Next, the tri-view statistical pooling strategy extracts robust spatiotemporal representations from the fused features, producing more discriminative multi-view representations. These representations are then fed into the bi-directional Mamba module to capture long-range global dependencies and enable efficient sequence modeling, while ensuring the module's computation remains linear. Finally, the classifier outputs the predicted motor imagery class. To further validate the effectiveness of the proposed method, detailed comparative experiments were conducted on two public MI-EEG datasets, demonstrating that STE-Mamba significantly outperforms existing approaches while maintaining computational efficiency.

The main contributions of this paper are summarized as follows:

1. We propose STE-BiMamba, a novel end-to-end spatiotemporal feature-enhanced bidirectional Mamba network for motor imagery EEG decoding. The model integrates multi-scale temporal feature extraction, tri-view statistical pooling, and bidirectional Mamba modeling to achieve efficient and accurate spatiotemporal dependency modeling in EEG signals;
2. To comprehensively capture the multi-dimensional statistical properties of EEG signals and improve feature discriminability and robustness, we design a tri-view statistical pooling strategy, extracting fused spatiotemporal features from mean, variance, and maximum perspectives;
3. We introduce a bi-directional Mamba global sequence modeling module, which efficiently captures long-range dependencies with linear computational complexity ($O(n)$), offering a strong advantage in both performance and efficiency for real-time BCI applications;
4. We present VR-MI, a high-quality self-collected MI-EEG dataset recorded in a virtual reality environment, to facilitate further research in the BCI field.

Related Work

In the field of EEG research, early analyses primarily relied on traditional machine learning methods to uncover information related to neural classification and neuroimaging. With the growing effectiveness of deep learning (DL) in modeling time series, recent studies have increasingly adopted DL-based models, achieving notable results in motor imagery (MI) EEG decoding tasks.

Convolutional Neural Networks (CNNs) are among the most widely used deep learning models for EEG decoding^{28,29}. These networks mimic the processing behavior of biological neural systems and demonstrate strong performance in EEG signal analysis. For example, Lawhern et al.¹⁶ proposed the compact CNN architecture EEGNet, performing temporal convolutions along the time dimension and depthwise convolutions along the spatial dimension. Subsequently, Wang et al.³⁰

introduced IFNet, which incorporates an interactive frequency convolution mechanism to effectively capture cross-frequency interactions, thereby improving MI-EEG decoding performance. FBCSPNet³¹ integrates the classical Filter Bank Common Spatial Pattern (FBCSP) concept into CNNs and demonstrates strong classification capabilities across multiple benchmark datasets. Additionally, researchers have explored hybrid models, such as Xu et al.³², who combined CNN and LSTM to leverage both spatial feature extraction and temporal sequence modeling.

For feature aggregation, existing EEG deep learning models commonly adopt average pooling operations^{33–36} to consolidate temporal information and reduce computational cost. However, single average pooling only captures first-order statistics and may overlook important distributional information. Recent studies have explored statistical pooling methods such as variance pooling to extract richer feature representations^{37,38}. Inspired by these works, this paper proposes a tri-view statistical pooling strategy, extracting fused spatiotemporal features from mean, variance, and maximum perspectives to obtain more discriminative and robust representations.

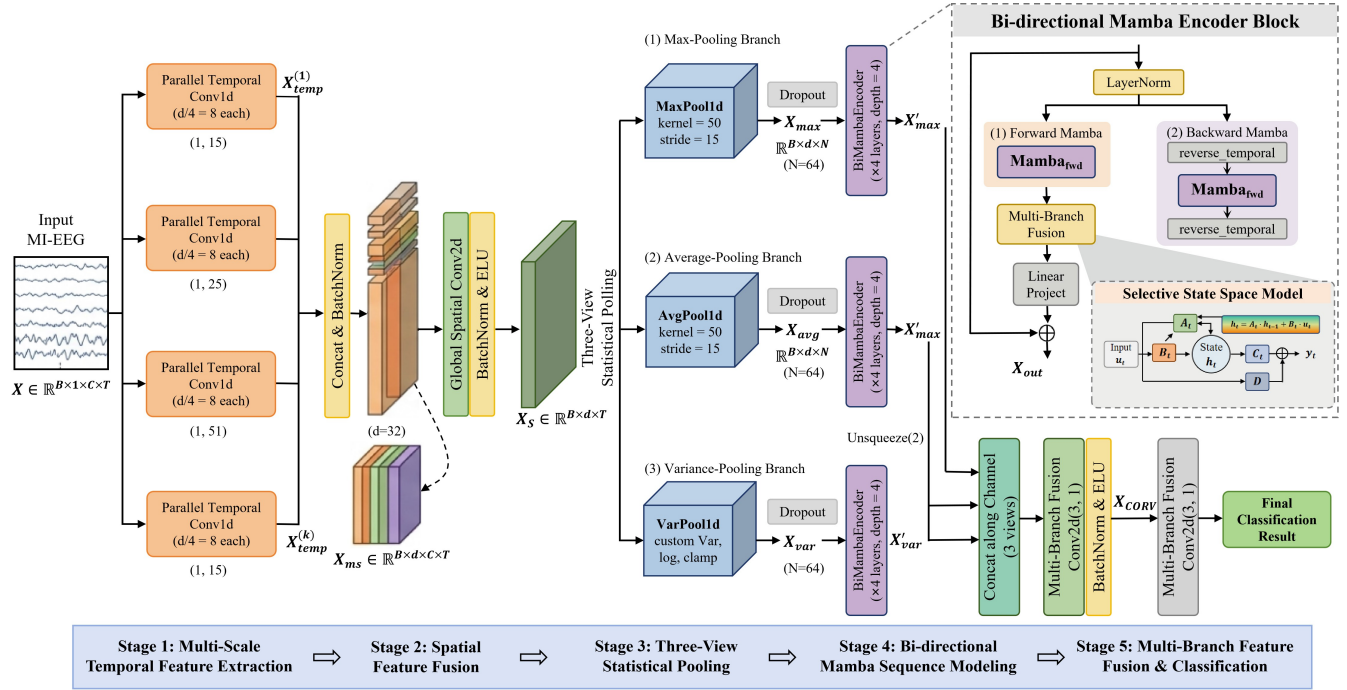


Figure 1. Overall architecture of the proposed STE-BiMamba model. The framework consists of five main stages: Stage 1: Multi-Scale Temporal Feature Extraction; Stage 2: Spatial Feature Fusion; Stage 3: Three-View Statistical Pooling; Stage 4: Bi-directional Mamba Sequence Modeling; Stage 5: Multi-Branch Feature Fusion & Classification.

The state-space model (SSM)-based Mamba architecture has attracted attention due to its efficient long-sequence modeling capability. Gu et al.³⁹ first proposed Mamba, introducing a selective mechanism that dynamically filters irrelevant information and enables global dependency modeling with linear-time computational complexity. This architecture has demonstrated superior performance compared with Transformers in tasks such as natural language processing and time series prediction^{40,41}. In EEG decoding, the application of Mamba remains at an early stage, with existing studies mostly using a unidirectional structure⁴². To capture bidirectional dependencies in MI-EEG signals more comprehensively, this paper proposes a bi-directional Mamba module, which processes sequences in both forward and backward directions to fully exploit contextual information.

Transformer models, based on self-attention mechanisms, are widely applied in sequence modeling tasks, including EEG decoding^{43–45}. Kostas et al.⁴⁶ proposed a Transformer pretraining and fine-tuning approach for EEG classification tasks. Song et al.⁴⁷ combined CNN and Transformer, first extracting local low-level features through convolution, and then capturing global correlations via self-attention. Inspired by these approaches, this paper integrates multi-scale convolutional spatiotemporal features with a bi-directional Mamba structure to form an efficient global sequence modeling module, which captures long-range dependencies while maintaining linear-time computational complexity, thereby improving model efficiency.

Materials and Methods

This section first introduces the motor imagery EEG dataset used in the experiments. Subsequently, the proposed STE-BiMamba motor imagery classification model is described in detail. As shown in Figure 1, the key modules of this model include multi-

scale temporal feature extraction, spatial feature fusion, three-view statistical pooling, and bidirectional Mamba sequence modeling. Finally, the evaluation metrics of the model are presented. The details are described as follows.

Dataset

VR-MI

We collected the VR-MI dataset⁴⁸ from 20 right-handed healthy participants in a virtual reality environment during EEG recording (as shown in Figure 2). The dataset is publicly available at <https://doi.org/10.21227/89ez-a093>. The participants' handedness was verified with the Edinburgh Handedness Questionnaire. The study was approved by the Academic Committee of Fujian Agriculture and Forestry University (approval number: PZCASFAFU25090). The experiment was carried out in accordance with the relevant guidelines and regulations. Informed consent was obtained from all subjects and/or their legal guardian(s). EEG signals were acquired simultaneously with the PICO 4 Ultra VR headset and the Emotiv Flex 2.0 system at a sampling rate of 256 Hz, using 32 scalp electrodes positioned in accordance with the international 10–20 system. The electrode positions were Cz, Fz, Fp1, F7, F3, FC1, C3, FC5, FT9, T7, CP5, CP1, P3, P7, PO9, O1, Pz, Oz, O2, PO10, P8, P4, CP2, CP6, T8, FT10, FC6, C4, FC2, F4, F8, and Fp2.

The experimental task involved binary classification of left- versus right-hand motor imagery (MI). In the self-developed VR environment [Figure 3(a)], participants were prompted to enter the MI state by watching virtual avatar movements that signaled left- or right-hand actions. Each participant performed a total of 200 trials (100 trials per hand) across two independent sessions, with 160 trials designated for training and 40 for testing. The structure of a single trial is presented in Figure 3(b). Each trial lasted 8 s and included a 3 s preparation phase (0–3 s), a 3 s MI phase (3–6 s), and a 2 s rest phase (6–8 s). Participants received a 1-minute break after every ten trials.

BCIC-IV-2a

The BCIC-IV-2a dataset⁴⁹ consists of nine healthy subjects. Each subject completed one training session and one evaluation session at different times. Each session contains 288 trials across four classes: left hand, right hand, both feet, and tongue motor imagery. The data were recorded from 22 EEG channels at a sampling rate of 250 Hz, with each trial lasting approximately 7.5 seconds.

The experimental paradigm of this dataset is illustrated in Figure 4. At $t = 0$ s, a fixed cross appears at the center of the screen. At $t = 2$ s, an arrow pointing left, right, down, or up is displayed for approximately 1.25 s, serving as a cue corresponding to left hand, right hand, both feet, or tongue motor imagery, respectively. Subjects then perform 3 s of motor imagery, followed by a 1.5 s rest period.

BCIC-IV-2b

The BCIC-IV-2b dataset⁵⁰ consists of nine healthy subjects. Each subject completed multiple sessions, including training sessions and evaluation sessions with feedback. Each session contains 120 trials across two classes: left hand and right hand motor imagery. The data were recorded from three bipolar EEG channels (C3, Cz, C4) at a sampling rate of 250 Hz, with each trial lasting approximately 7.5 seconds.

The experimental paradigm is illustrated in Figure 5. At $t = 0$ s, a fixed cross appears at the center of the screen. At $t = 2$ s, an arrow pointing left or right is displayed as a cue (corresponding to left hand or right hand motor imagery, respectively) and lasts for approximately 1.25 s. Subjects then perform 3 s of motor imagery, followed by a 1.5 s rest period.

HGD

The HGD dataset⁵¹ consists of 14 healthy subjects. Each subject performed 880 trials, covering four classes: left hand, right hand, both feet, and rest (all corresponding to actual motor execution tasks). Data were recorded from 128 EEG channels at a sampling rate of 500 Hz, with each trial lasting 4 seconds. Typically, 44 channels covering the motor cortex are selected for analysis, and the data are downsampled to 250 Hz.

The experimental paradigm of this dataset is illustrated in Fig. 6. In a single trial, a cue arrow is presented at the center of the screen for 4 seconds as a visual prompt. The arrow points left, right, down, or up, corresponding to sequential finger tapping of the left hand, sequential finger tapping of the right hand, repetitive toe gripping with both feet, or relaxation/rest, respectively. Subjects are required to perform the indicated movement at their own pace during the arrow presentation, followed by a random inter-trial rest period of 3–4 seconds before the next trial begins.



Figure 2. Virtual character action details designed based on mirror neuron theory.

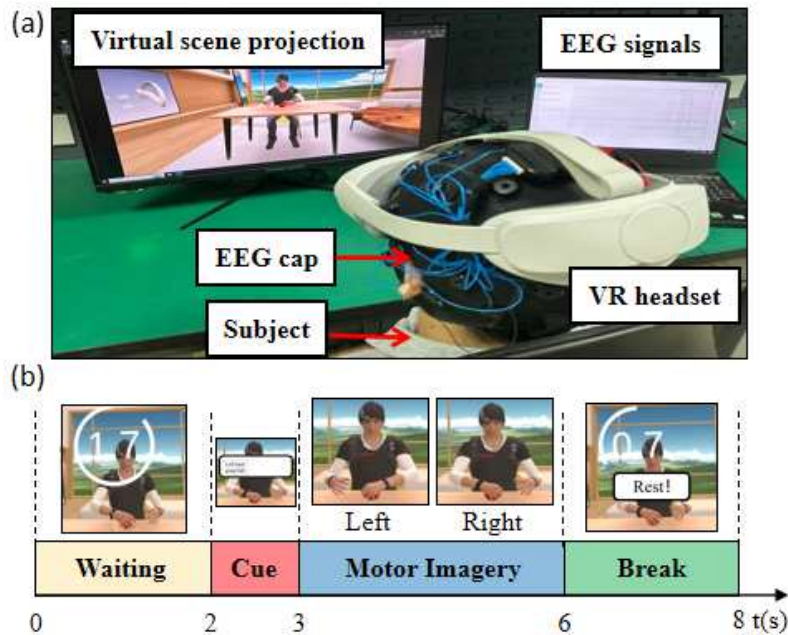


Figure 3. MI experimental tasks in the self-constructed VR environment. (a) Subject 2 performs motor imagery according to the guidance in a virtual scenario, and EEG data is collected in real-time during the motor imagery. (b) Each motor imagery stimulus in the experiment lasts for 8 seconds. The 0–2 seconds is the Waiting phase, 2–3 seconds is the Cue phase, 3–6 seconds is the motor imagery based on the guidance of a virtual character, and 6–8 seconds is the resting phase after the MI ends.

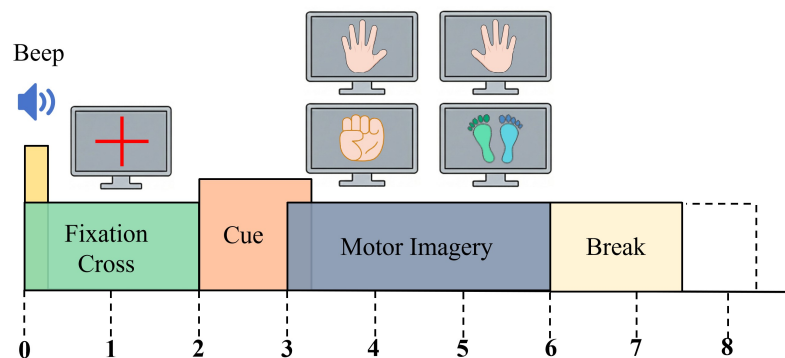


Figure 4. The BCIC-IV-2a dataset paradigm

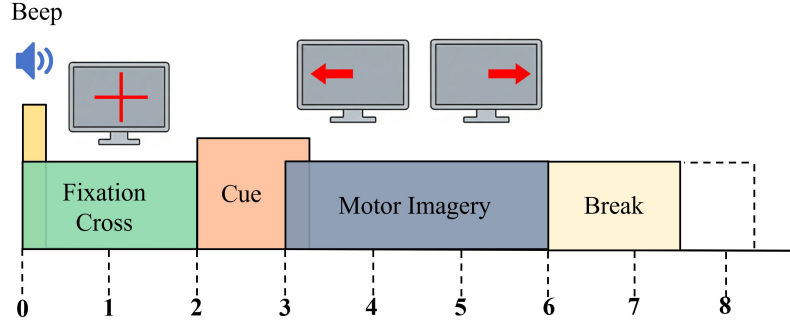


Figure 5. The BCIC-IV-2b dataset paradigm

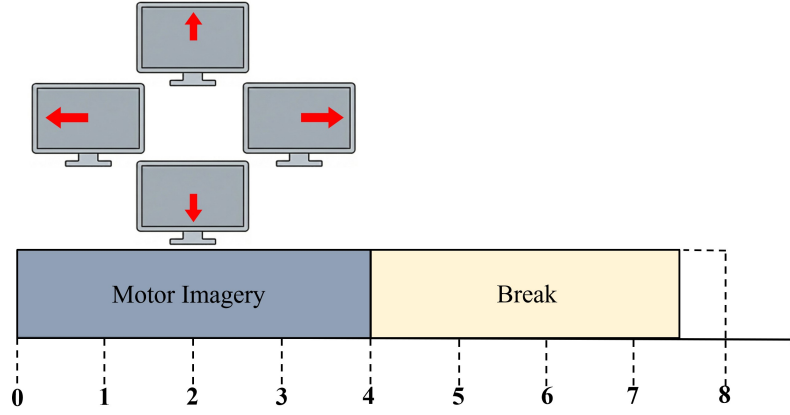


Figure 6. The HGD dataset paradigm

EEG Signal Preprocessing

For the self-collected VR-MI dataset, EEG signals were preprocessed using the EEGLAB toolbox in MATLAB following a standard pipeline. The pipeline consisted of electrode localization and channel selection, downsampling, bandpass filtering, epoch segmentation with baseline correction, bad-channel interpolation, bad-epoch rejection, re-referencing, independent component analysis (ICA), and artifact component removal.

Electrode locations were first imported and aligned to the standard 10–20 system. Non-EEG channels were removed. The data were then downsampled to 250 Hz to reduce computational load while preserving the frequency content relevant to motor imagery. A fifth-order Butterworth bandpass filter (0.5–40 Hz)⁵² was subsequently applied. Its transfer function is given by

$$H(s) = \frac{b_0 + b_1s + \dots + b_ns^n}{a_0 + a_1s + \dots + a_ns^n}, \quad (1)$$

where s is the complex frequency variable and the filter coefficients b_i and a_i are designed to attenuate low-frequency drifts and high-frequency noise while retaining the mu and beta rhythms critical for motor imagery decoding.

The continuous EEG data were then segmented into epochs corresponding to the motor imagery period of each trial (3–6 s). Baseline correction was performed by subtracting the mean amplitude of the pre-stimulus baseline window from each epoch:

$$x_{\text{corrected}}(t) = x(t) - \frac{1}{N_b} \sum_{t \in \text{baseline}} x(t), \quad (2)$$

where N_b denotes the number of samples in the baseline interval.

Bad channels were identified based on excessive variance and interpolated using spherical spline interpolation. Bad epochs containing large artifacts were rejected. The data were re-referenced to the average reference to reduce common-mode noise:

$$x'_i(t) = x_i(t) - \frac{1}{C} \sum_{j=1}^C x_j(t), \quad (3)$$

where C is the number of EEG channels and $x_i(t)$ is the signal at channel i .

Independent component analysis (ICA) was then performed to decompose the multi-channel EEG into statistically independent sources:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (4)$$

where $\mathbf{x}(t)$ denotes the observed EEG signals, $\mathbf{A}$ is the mixing matrix, and $\mathbf{s}(t)$ represents the underlying independent components. An unmixing matrix $\mathbf{W}$ was estimated such that the sources could be recovered as $\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t)$. Artifactual components corresponding to ocular activity, muscle activity, cardiac signals, and line noise were identified through a combination of automated classification and visual inspection of component topography, power spectra, and time courses, and subsequently removed.

For the publicly available datasets (BCIC-IV-2a, BCIC-IV-2b, and HGD), which had already undergone standard pre-processing in their original releases (including filtering, re-referencing, and epoching), only an additional bandpass filtering step (0.5–40 Hz) using the same fifth-order Butterworth filter was applied. This ensured that all datasets shared an identical frequency band for fair comparison while minimizing unnecessary re-processing of already cleaned public data.

All preprocessing parameters were kept consistent across datasets and subjects to maximize reproducibility.

Multi-Scale Temporal Feature Extraction

Motor imagery activities produce rich multi-scale temporal dynamics: short-scale local peaks reflect rapid event-related desynchronization (ERD) phenomena, while longer time scales contain stable oscillatory patterns such as mu and beta rhythms. To effectively capture this hierarchical temporal structure, the proposed STE-BiMamba employs a multi-scale CNN as the front-end feature extraction module.

Given the input MI-EEG sample $X \in \mathbb{R}^{B \times 1 \times C \times T}$ (where $C = 22$ denotes the number of EEG channels and $T = 1000$ represents the number of time points), the model applies four parallel temporal one-dimensional convolution layers with kernel sizes of 15, 25, 51, and 65, respectively:

$$X_{\text{temp}}^{(k)} = \text{Conv1d}(X; \text{kernel} = (1, k), \text{padding} = (0, p_k)), \quad k \in \{15, 25, 51, 65\} \quad (5)$$

where the padding values p_k are set to 7, 12, 25, and 32 respectively to preserve the original sequence length. Each branch outputs $d/4 = 8$ feature maps (with embedding dimension $d = 32$). The outputs of the four branches are then concatenated along the channel dimension and processed through a batch normalization layer:

$$X_{\text{ms}} = \text{BatchNorm}\left(\text{Concat}\left(X_{\text{temp}}^{(15)}, X_{\text{temp}}^{(25)}, X_{\text{temp}}^{(51)}, X_{\text{temp}}^{(65)}\right)\right) \quad (6)$$

This yields the multi-scale temporal feature map $X_{\text{ms}} \in \mathbb{R}^{B \times d \times C \times T}$. Through multi-scale convolution, the model can simultaneously perceive fine-grained transient changes and wide-span rhythmic patterns in MI-EEG signals, laying a solid foundation for subsequent spatial feature fusion and sequence modeling.

Spatial Feature Fusion

MI-EEG signals exhibit significant spatial correlations between different brain regions. To fully integrate this spatial information and obtain a compact spatio-temporal joint representation, the model introduces a spatial feature fusion module after the multi-scale temporal feature extraction stage.

To adapt to EEG datasets with different numbers of channels, the spatial convolution kernel size is set dynamically according to the input channel count C . The input multi-scale temporal feature map is $X_{\text{ms}} \in \mathbb{R}^{B \times d \times C \times T}$. The model applies a spatial convolution layer with kernel size $(C, 1)$, where both the input and output channel dimensions are d :

$$X_{\text{spatial}} = \text{Conv2d}(X_{\text{ms}}; \text{kernel} = (C, 1)) \quad (7)$$

This convolution performs a global mixing of all input EEG channels in a single operation, enabling the model to flexibly handle datasets with varying channel counts (e.g., 22 channels in BCIC-IV-2a/2b, 44 selected channels in HGD, or the channel configuration of the self-collected VR-MI dataset) without requiring manual channel selection or padding. It is then followed by a batch normalization layer and an ELU activation function:

$$X_s = \text{ELU}\left(\text{BatchNorm}(X_{\text{spatial}})\right) \quad (8)$$

This produces the final spatio-temporal joint feature $X_s \in \mathbb{R}^{B \times d \times T}$. This spatial fusion step effectively integrates the synergistic information across different brain regions, providing high-quality spatio-temporal representations for the subsequent three-view statistical pooling and bidirectional Mamba sequence modeling.

Three-View Statistical Pooling

The class differences in motor imagery not only manifest in temporal and spatial patterns but are more profoundly reflected in the statistical characteristics of the signals: different classes exhibit significant differences in peak intensity, overall energy level, and signal variability (variance). To comprehensively capture these statistical characteristics, this study proposes a three-view statistical pooling method.

This module takes the multi-scale spatio-temporal feature $X_s \in \mathbb{R}^{B \times d \times T}$ as input and designs three mutually independent pooling branches: (1) the max-pooling branch, which captures the most prominent peak activations in MI-EEG signals; (2) the average-pooling branch, which extracts the overall energy level of MI-EEG signals; and (3) the variance-pooling branch, which specifically models the fluctuation and variability of the signals.

Specifically, all three branches use the same pooling window size (kernel size = 50) and stride (stride = 15), producing an output sequence length of

$$N = \left\lfloor \frac{T - 50}{15} \right\rfloor + 1 = 64 \quad (9)$$

The max-pooling and average-pooling branches are implemented using standard pooling layers:

$$X_{\max} = \text{MaxPool1d}(X_s; \text{kernel} = 50, \text{stride} = 15) \quad (10)$$

$$X_{\text{avg}} = \text{AvgPool1d}(X_s; \text{kernel} = 50, \text{stride} = 15) \quad (11)$$

The variance-pooling branch employs a custom module that computes the variance of each sliding window, followed by a logarithmic transformation and numerical stabilization:

$$X_{\text{var}} = \log(\text{clamp}(\text{Var}(X_s[i : i + 50]), 10^{-6}, 10^6)) \quad (12)$$

All three branches are further regularized with Dropout to improve generalization. This yields three independent sequence features $X_{\max}, X_{\text{avg}}, X_{\text{var}} \in \mathbb{R}^{B \times d \times N}$. Through this three-view design, the model can simultaneously observe MI-EEG signals from three complementary statistical perspectives—peak, mean, and variance—thereby significantly enhancing the discriminability and richness of the extracted features.

Bi-directional Mamba Sequence Modeling

To further capture the global long-range temporal dependencies in MI-EEG signals, this study designs a bidirectional Mamba sequence modeling module (Bi-directional Mamba Sequence Modeling) for efficiently modeling the contextual information of long sequences.

The enriched spatio-temporal feature data $X_{\max}, X_{\text{avg}}, X_{\text{var}} \in \mathbb{R}^{B \times N \times d}$ (where $N = 64, d = 32$) are fed separately into four stacked BiMambaEncoder layers. The specific processing flow of each BiMambaEncoder is as follows (see Algorithm 1):

First, Layer Normalization is applied to the input features:

$$X_{\text{norm}} = \text{LayerNorm}(X_{\text{in}}) \quad (13)$$

Subsequently, the sequence is modeled through two branches: the forward Mamba and the backward Mamba. The forward Mamba processes the original sequence directly, while the backward Mamba first flips the temporal dimension:

$$X_{\text{norm}}^{\text{flip}} = X_{\text{norm}}[:, :: -1, :] \quad (14)$$

and flips it back after processing. The outputs of the two branches are concatenated, passed through a linear projection layer for fusion, and combined with a residual connection to produce the final output:

$$X_{\text{out}} = X_{\text{in}} + \text{Linear}\left(\text{Concat}\left(\text{Mamba}_{\text{fwd}}(X_{\text{norm}}), \text{Mamba}_{\text{bwd}}(X_{\text{norm}}^{\text{flip}})\right)\right) \quad (15)$$

The core of the Mamba module is the Selective State Space Model. Its discretized state update process is defined as:

$$h_t = \bar{A}_t h_{t-1} + \bar{B}_t u_t, \quad y_t = C_t h_t + D u_t \quad (16)$$

The state transition matrix and input matrix are computed from the input-dependent dynamic parameter Δ_t :

$$\bar{A}_t = \exp(\Delta_t A), \quad \bar{B}_t = \Delta_t B \quad (17)$$

where $A = -\exp(A_{\log})$ is a learnable diagonal state transition matrix ($A_{\log}$ is a model parameter), B and C are generated by the x -proj linear layer, Δ_t is dynamically produced by the dt -proj through a softplus projection, and D is a learnable skip connection parameter.

After processing through four BiMambaEncoder layers, the three branches respectively obtain the enhanced sequence features $X'_{\max}, X'_{\text{avg}}, X'_{\text{var}} \in \mathbb{R}^{B \times N \times d}$.

To achieve effective fusion of the three statistical-view features, the model further performs multi-branch feature fusion on the outputs of the above three branches. First, the three sequence features are dimensionally expanded along the channel dimension and concatenated:

$$X_{\text{fuse}} = \text{Concat}(X'_{\max}.\text{unsqueeze}(2), X'_{\text{avg}}.\text{unsqueeze}(2), X'_{\text{var}}.\text{unsqueeze}(2)) \in \mathbb{R}^{B \times N \times 3 \times d} \quad (18)$$

Subsequently, a convolution layer with kernel size $(3, 1)$ is used for cross-view fusion:

$$X_{\text{conv}} = \text{Conv2d}(X_{\text{fuse}}; \text{kernel} = (3, 1)) \quad (19)$$

The fused features are then processed by Batch Normalization and ELU activation functions, flattened into a vector, and passed through a fully connected layer to produce the final classification result.

Algorithm 1 Bi-directional Mamba Encoder (Detailed)

Input: Input enriched spatio-temporal features $X \in \mathbb{R}^{B \times N \times d}$

Output: Enhanced output features $X_{\text{out}} \in \mathbb{R}^{B \times N \times d}$

Step 1: Layer Normalization

1: $X_{\text{norm}} \leftarrow \text{LayerNorm}(X)$

Step 2: Forward Mamba Branch

2: $X_{\text{fwd}} \leftarrow \text{Mamba}_{\text{fwd}}(X_{\text{norm}})$

Step 3: Backward Mamba Branch

3: $X_{\text{flip}} \leftarrow \text{reverse_temporal}(X_{\text{norm}})$

▷ $X_{\text{norm}}[:, :, -1, :]$

4: $X_{\text{bwd_flip}} \leftarrow \text{Mamba}_{\text{bwd}}(X_{\text{flip}})$

5: $X_{\text{bwd}} \leftarrow \text{reverse_temporal}(X_{\text{bwd_flip}})$

▷ Flip back

Step 4: Multi-Branch Fusion

6: $X_{\text{cat}} \leftarrow \text{Concat}(X_{\text{fwd}}, X_{\text{bwd}})$

7: $X_{\text{proj}} \leftarrow \text{Linear}(X_{\text{cat}})$

Step 5: Residual Connection

8: $X_{\text{out}} \leftarrow X + X_{\text{proj}}$

▷ Residual addition

Selective State Space Model (Core of Mamba)

9: $\Delta_t \leftarrow \text{softplus}(\text{dt_proj}(x))$

▷ Dynamic parameter

10: $\bar{A}_t \leftarrow \exp(\Delta_t \cdot A)$

11: $\bar{B}_t \leftarrow \Delta_t \cdot B$

12: $h_t \leftarrow \bar{A}_t h_{t-1} + \bar{B}_t u_t$

13: $y_t \leftarrow C_t h_t + D u_t$

return X_{out}

Evaluation Metrics

To comprehensively evaluate the classification performance of the model, this study adopts Accuracy, F1-score, and Kappa coefficient (Cohen's Kappa) as the main evaluation metrics, with the specific formulas as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (20)$$

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

Precision is first defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (21)$$

The F1-score, which is the harmonic mean of Precision and Recall, is then calculated as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (22)$$

The Kappa coefficient considers random consistency and can more objectively reflect the performance of the model in multi-class tasks. Its calculation formula is:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (23)$$

where p_o is the observed agreement proportion and p_e is the expected agreement proportion by chance. All evaluation metrics are computed on the test set and reported as mean $\pm$ standard deviation. To verify the statistical significance of performance differences between different models, appropriate statistical tests were conducted, and the corresponding p-values are reported.

Results

The proposed STE-BiMamba model was evaluated on four motor imagery (MI) EEG datasets: our self-collected VR-MI dataset, BCI Competition IV 2a (BCIC-IV-2a), BCI Competition IV 2b (BCIC-IV-2b), and the High Gamma Dataset (HGD). To ensure a fair and realistic evaluation, all experiments were conducted under both within-subject cross-session and cross-subject settings.

Dataset Division Details

In the within-subject cross-session setting, we followed the official cross-session divisions of each dataset. For BCIC-IV-2a, both the training and test sets contained 288 trials per subject with a 1:1 ratio. For BCIC-IV-2b, the number of trials varied across subjects, with approximately 400–440 trials in the training set and 280–320 trials in the test set per subject. For the High Gamma Dataset (HGD), the training set consisted of approximately 880 trials and the test set consisted of 160 trials per subject. For the self-collected VR-MI dataset, 160 trials were used for training and 40 trials for testing per subject with a 4:1 ratio.

In the cross-subject setting, a leave-one-subject-out (LOSO) cross-validation strategy was adopted. For each fold, all trials from one subject were used as the test set, while the data from all remaining subjects were combined as the training set. The final performance metrics were obtained by averaging the results across all subjects.

STE-BiMamba Within-Subject Cross-Session Performance

As shown in Fig. 7, the STE-BiMamba model achieved excellent performance on the VR-MI dataset under the cross-session setting. Several subjects attained 100% motor imagery (MI) classification accuracy, with the lowest accuracy of 87.50% observed in the 10th subject. The model attained an average accuracy of **97.50% $\pm$ 3.35%** and an average Kappa value of **0.9500 $\pm$ 0.0671**.

Table 1 summarizes the subject-wise classification performance of STE-BiMamba on the three public datasets under the cross-session setting. On the BCIC-IV-2a dataset, STE-BiMamba achieved an average accuracy of **82.95% $\pm$ 10.52%**, an average Kappa coefficient of **0.7726 $\pm$ 0.1403**, and an average F1-score of **0.8280 $\pm$ 0.1068**. On the BCIC-IV-2b dataset, the model achieved an average accuracy of **89.20% $\pm$ 8.42%**, a Kappa coefficient of **0.7840 $\pm$ 0.1675**, and an F1-score of **0.8923 $\pm$ 0.0844**.

Additionally, STE-BiMamba demonstrated excellent performance on the HGD dataset, achieving an average accuracy of **95.67% $\pm$ 2.87%**, a Kappa coefficient of **0.9422 $\pm$ 0.0383**, and an F1-score of **0.9564 $\pm$ 0.0291**.

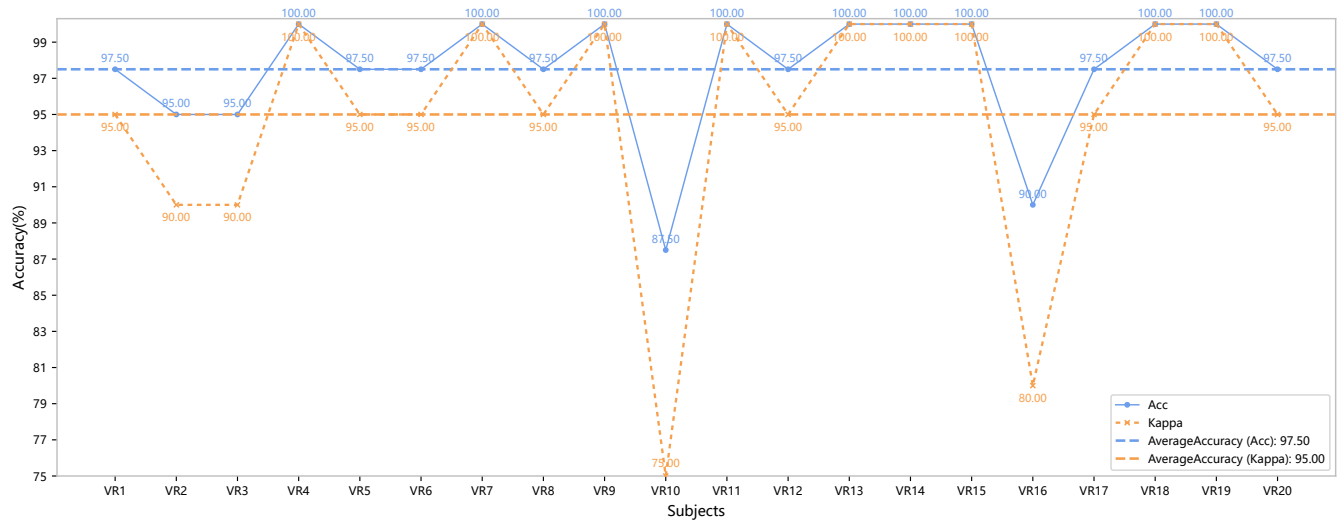


Figure 7. STE-BiMamba performance on VR-MI dataset (cross-session).

Table 1. Subject-wise classification performance of STE-BiMamba under the cross-session setting. Accuracy is reported in %, whereas Kappa and F1-score are unitless.

Dataset	Metric	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	Mean $\pm$ SD
BCIC-IV-2a	Acc.	87.15	63.54	97.22	81.25	78.47	69.10	94.79	88.54	86.46	–	–	–	–	–	82.95 $\pm$ 10.52
	Kappa	0.829	0.514	0.963	0.750	0.713	0.588	0.931	0.847	0.819	–	–	–	–	–	0.773 $\pm$ 0.140
	F1	0.870	0.626	0.972	0.813	0.783	0.692	0.948	0.885	0.863	–	–	–	–	–	0.828 $\pm$ 0.107
BCIC-IV-2b	Acc.	79.06	71.43	86.25	98.56	96.88	91.56	93.75	95.31	90.00	–	–	–	–	–	89.20 $\pm$ 8.42
	Kappa	0.581	0.431	0.725	0.969	0.938	0.831	0.875	0.906	0.800	–	–	–	–	–	0.784 $\pm$ 0.168
	F1	0.792	0.712	0.865	0.982	0.971	0.918	0.940	0.952	0.898	–	–	–	–	–	0.892 $\pm$ 0.084
HGD	Acc.	95.63	94.38	100.00	100.00	94.38	95.63	94.34	96.88	98.13	95.63	89.38	98.13	95.60	91.25	95.67 $\pm$ 2.87
	Kappa	0.942	0.925	1.000	1.000	0.925	0.942	0.925	0.958	0.975	0.942	0.858	0.975	0.941	0.883	0.942 $\pm$ 0.038
	F1	0.956	0.943	1.000	1.000	0.944	0.956	0.943	0.969	0.981	0.956	0.892	0.981	0.956	0.912	0.956 $\pm$ 0.029

STE-BiMamba Cross-Subject Performance

The STE-BiMamba model was further evaluated under the more challenging cross-subject setting. As shown in Fig. 8, on the self-collected VR-MI dataset the model achieved an average accuracy of **85.05% $\pm$ 7.15%** and an average Kappa value of **0.7010 $\pm$ 0.1431**.

Table 2 summarizes the subject-wise classification performance of STE-BiMamba on the three public datasets under the cross-subject setting.

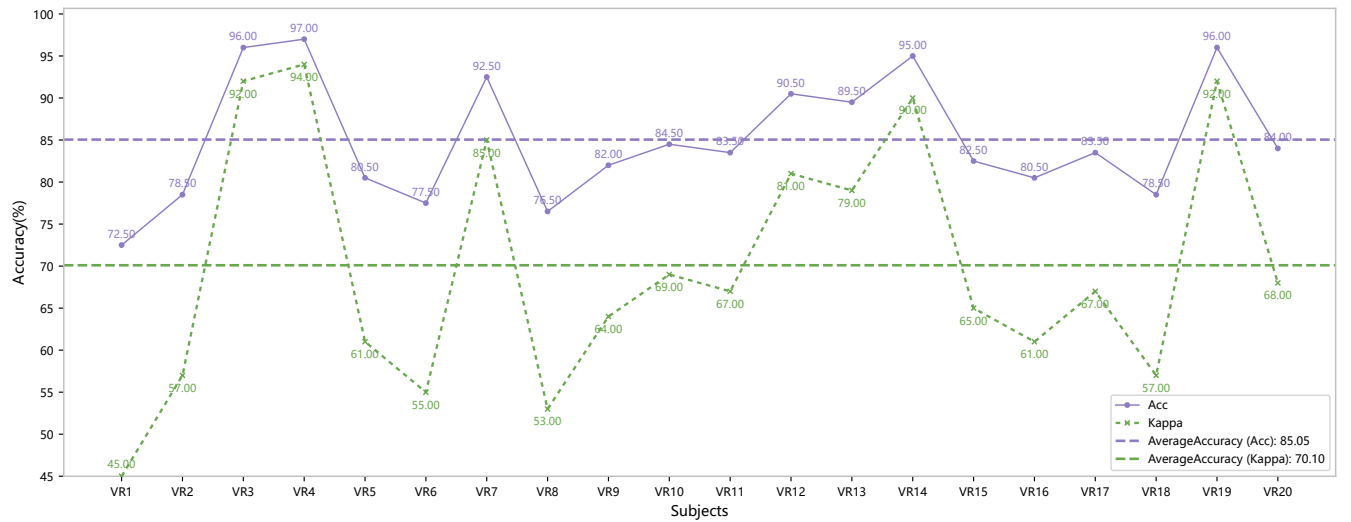


Figure 8. STE-BiMamba performance on VR-MI dataset (cross-subject).

Table 2. Subject-wise classification performance of STE-BiMamba under the cross-subject setting. Accuracy is reported in %, whereas Kappa and F1-score are unitless.

Dataset	Metric	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	Mean $\pm$ SD
BCIC-IV-2a	Acc.	75.17	47.57	84.72	63.02	57.29	51.74	73.09	76.74	74.13	–	–	–	–	–	67.05 $\pm$ 11.94
	Kappa	0.669	0.301	0.796	0.507	0.431	0.357	0.641	0.690	0.655	–	–	–	–	–	0.561 $\pm$ 0.159
	F1	0.756	0.478	0.845	0.626	0.571	0.519	0.727	0.768	0.738	–	–	–	–	–	0.670 $\pm$ 0.119
BCIC-IV-2b	Acc.	72.50	64.29	71.88	92.81	95.31	87.82	89.06	93.75	85.31	–	–	–	–	–	83.64 $\pm$ 10.59
	Kappa	0.450	0.288	0.438	0.856	0.975	0.788	0.781	0.875	0.706	–	–	–	–	–	0.673 $\pm$ 0.219
	F1	0.725	0.643	0.718	0.930	0.989	0.844	0.889	0.938	0.854	–	–	–	–	–	0.837 $\pm$ 0.110
HGD	Acc.	92.69	84.04	84.52	84.27	87.60	75.68	83.70	69.90	71.06	71.12	67.21	89.68	79.25	87.02	80.55 $\pm$ 7.88
	Kappa	0.903	0.787	0.794	0.790	0.835	0.676	0.783	0.599	0.614	0.615	0.563	0.862	0.723	0.827	0.741 $\pm$ 0.105
	F1	0.925	0.839	0.845	0.843	0.879	0.759	0.840	0.702	0.713	0.712	0.672	0.895	0.793	0.871	0.806 $\pm$ 0.078

On the BCIC-IV-2a dataset, STE-BiMamba achieved an average accuracy of $67.05\% \pm 11.94\%$, an average Kappa coefficient of 0.5607 ± 0.1593 , and an average F1-score of 0.6699 ± 0.1188 .

On the BCIC-IV-2b dataset, the model achieved an average accuracy of $83.64\% \pm 10.59\%$, a Kappa coefficient of 0.6729 ± 0.2192 , and an F1-score of 0.8366 ± 0.1102 .

Additionally, STE-BiMamba demonstrated competitive performance on the HGD dataset, achieving an average accuracy of $80.55\% \pm 7.88\%$, a Kappa coefficient of 0.7407 ± 0.1051 , and an F1-score of 0.8062 ± 0.0780 .

Comparison with Previous Methods

To comprehensively evaluate the performance of the proposed STE-BiMamba model, we compared it with the following representative baseline methods on the three public datasets (BCIC-IV-2a, BCIC-IV-2b, and HGD) under both the within-subject cross-session and cross-subject settings

- **EEGNet¹⁶**: A lightweight convolutional neural network designed for EEG-based BCI that employs depthwise separable convolutions and is characterized by a small number of parameters and strong generalization ability.
- **Conformer⁴⁷**: A hybrid architecture that combines convolutional neural networks and Transformers, enabling the model to capture both local spatiotemporal features and global long-range dependencies in EEG signals.
- **FBCSPNet³²**: A deep learning network based on filter bank common spatial patterns (FBCSP) that integrates traditional feature extraction concepts into an end-to-end framework.
- **IFNet³⁰**: A feature fusion network that adopts a multi-branch Inception structure for multi-scale spatiotemporal feature extraction from EEG signals.
- **FBCSP¹⁴**: The classical filter bank common spatial pattern algorithm, which is the most representative traditional machine learning method in the field of motor imagery EEG classification.
- **Deep ConvNet¹⁸**: A deep convolutional neural network that performs end-to-end feature learning and classification directly from raw EEG signals through multiple convolutional layers.
- **CNN-Mamba⁵³**: A hybrid architecture that combines CNN and the Mamba state space model, enhancing long-sequence modeling capability while maintaining high computational efficiency.
- **CNN-LSTM³¹**: A classic hybrid model that combines convolutional neural networks and long short-term memory networks to simultaneously extract spatial features and temporal dependencies.

Table 3 presents the quantitative comparison results between the proposed STE-BiMamba and various baseline models across the three datasets. All p-values reported in the table are Bonferroni-corrected based on statistical difference tests of accuracy compared to the proposed model.

The experimental results demonstrated that STE-BiMamba achieved the best performance on all three datasets, with statistically significant improvements over both traditional machine learning methods and state-of-the-art deep learning baselines. On the BCIC-IV-2a dataset, STE-BiMamba achieved an average accuracy of $82.95\% \pm 10.52\%$, which was 2.36 percentage points higher than the second-best CNN-Mamba (80.59%) and 4.06 percentage points higher than Conformer (78.89%). Most comparisons reached statistical significance at $p < 0.001$ or $p < 0.01$ levels. Similar significant advantages were also observed in Kappa and F1-score.

On the BCIC-IV-2b dataset, STE-BiMamba attained an average accuracy of $89.20\% \pm 8.42\%$, outperforming CNN-Mamba (87.43%) and Conformer (87.64%) with statistical significance in most cases. The Kappa coefficient (0.7840) and F1-score (0.8923) further confirmed its superiority.

On the HGD dataset, the performance advantage of STE-BiMamba was particularly pronounced, achieving an average accuracy of $95.67\% \pm 2.87\%$ ($p < 0.001$ or better against most baselines). The Kappa coefficient and F1-score reached 0.9422 and 0.9564, respectively, demonstrating substantial and statistically significant leads over all competing models.

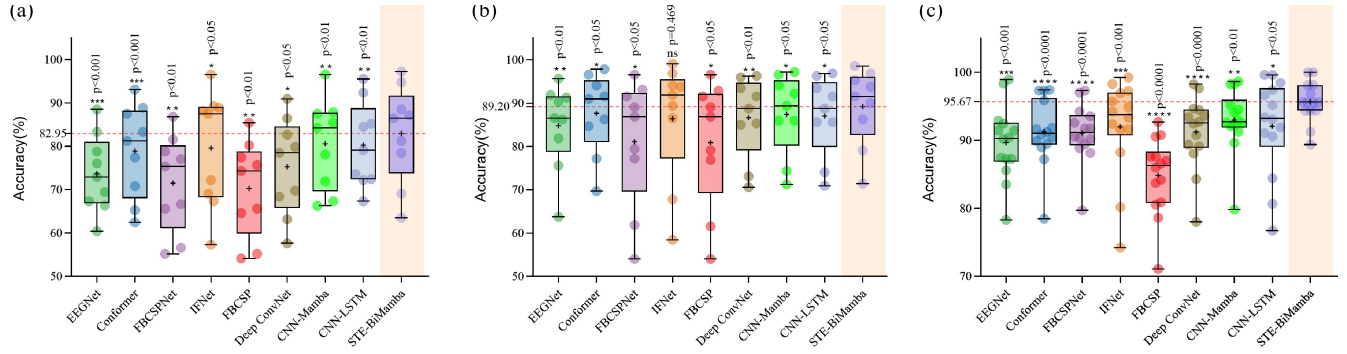
The distribution of decoding accuracy across different subjects for all models on the three datasets is shown in Fig. 9. It can be observed that the proposed STE-BiMamba model achieved higher average decoding accuracy than other baseline models across all datasets. We further conducted statistical significance analysis. For example, on the BCIC-IV-2a dataset, STE-BiMamba showed statistically significant improvements compared to EEGNet and Conformer ($p < 0.001$); on the HGD dataset, STE-BiMamba exhibited even stronger significant differences compared to FBCSPNet and DeepConvNet ($p < 0.0001$).

In the cross-subject setting, as shown in Table 4, the proposed STE-BiMamba model demonstrated strong generalization ability and achieved statistically significant improvements over all baseline models across the three public datasets. Specifically,

Table 3. Performance comparison of different models on BCIC-IV-2a, BCIC-IV-2b, and HGD datasets (cross-session).

Model	BCIC-IV-2a			BCIC-IV-2b			HGD		
	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std
EEGNet	73.68% $\pm$ 8.40%***	0.6359 $\pm$ 0.1244	0.7334 $\pm$ 0.1011	84.79% $\pm$ 9.31%**	0.6958 $\pm$ 0.1862	0.8477 $\pm$ 0.0928	89.68% $\pm$ 5.26%**	0.8624 $\pm$ 0.0701	0.8976 $\pm$ 0.0526
Conformer	78.89% $\pm$ 10.12%***	0.7186 $\pm$ 0.1341	0.7891 $\pm$ 0.1012	87.64% $\pm$ 8.77%*	0.7528 $\pm$ 0.1754	0.8760 $\pm$ 0.0883	91.29% $\pm$ 4.81%***	0.8839 $\pm$ 0.0641	0.9132 $\pm$ 0.0481
FBCSPNet	71.53% $\pm$ 10.44%**	0.6204 $\pm$ 0.1464	0.7171 $\pm$ 0.1106	81.78% $\pm$ 13.78%*	0.6215 $\pm$ 0.2756	0.8111 $\pm$ 0.1371	91.11% $\pm$ 4.14%***	0.8814 $\pm$ 0.0552	0.9107 $\pm$ 0.0421
IFNet	79.59% $\pm$ 12.54%*	0.7279 $\pm$ 0.1720	0.7957 $\pm$ 0.1288	86.42% $\pm$ 13.12%	0.7285 $\pm$ 0.2624	0.8642 $\pm$ 0.1316	92.02% $\pm$ 6.78%***	0.8936 $\pm$ 0.0904	0.9204 $\pm$ 0.0673
FBCSP	70.29% $\pm$ 10.36%**	0.6039 $\pm$ 0.1381	0.7033 $\pm$ 0.1031	80.94% $\pm$ 13.18%*	0.6187 $\pm$ 0.2762	0.8082 $\pm$ 0.1383	84.85% $\pm$ 5.46%***	0.7981 $\pm$ 0.0728	0.8479 $\pm$ 0.0539
Deep ConvNet	75.27% $\pm$ 10.49%*	0.6703 $\pm$ 0.1476	0.7525 $\pm$ 0.1106	86.67% $\pm$ 8.80%**	0.7333 $\pm$ 0.1761	0.8671 $\pm$ 0.0875	91.21% $\pm$ 5.13%***	0.8828 $\pm$ 0.0684	0.9120 $\pm$ 0.0507
CNN-Mamba	80.59% $\pm$ 9.74%**	0.7413 $\pm$ 0.1420	0.8231 $\pm$ 0.1103	87.43% $\pm$ 8.78%*	0.7486 $\pm$ 0.1746	0.8740 $\pm$ 0.0877	93.28% $\pm$ 4.52%**	0.9066 $\pm$ 0.0603	0.9298 $\pm$ 0.0454
CNN-LSTM	80.26% $\pm$ 9.19%**	0.7349 $\pm$ 0.1478	0.8209 $\pm$ 0.1092	87.01% $\pm$ 8.68%*	0.7406 $\pm$ 0.1735	0.8705 $\pm$ 0.0854	92.05% $\pm$ 6.78%*	0.8941 $\pm$ 0.0903	0.9200 $\pm$ 0.0683
STE-BiMamba	82.95% $\pm$ 10.52%	0.7726 $\pm$ 0.1403	0.8280 $\pm$ 0.1068	89.20% $\pm$ 8.42%	0.7840 $\pm$ 0.1675	0.8923 $\pm$ 0.0844	95.67% $\pm$ 2.87%	0.9422 $\pm$ 0.0383	0.9564 $\pm$ 0.0291

Note: Cyan cells denote the proposed STE-BiMamba model. All p-values are Bonferroni-corrected for multiple comparisons (5 comparisons) based on statistical difference tests of accuracy compared to the proposed model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

**Figure 9.** Per-subject decoding accuracy distribution of STE-BiMamba and baseline models under the within-subject cross-session setting. (a) BCIC-IV-2a dataset; (b) BCIC-IV-2b dataset; (c) HGD dataset.

STE-BiMamba attained an average accuracy of **67.05% $\pm$ 11.94%** ($p < 0.01$ or better against most baselines), with a Kappa coefficient of **0.5610 $\pm$ 0.1588** and an F1-score of **0.6699 $\pm$ 0.1188** on BCIC-IV-2a; an accuracy of **83.64% $\pm$ 10.59%**, Kappa of **0.6729 $\pm$ 0.2192**, and F1-score of **0.8366 $\pm$ 0.1102** on BCIC-IV-2b; and an accuracy of **80.55% $\pm$ 7.88%** ($p < 0.01$ or better), Kappa of **0.7410 $\pm$ 0.1051**, and F1-score of **0.8062 $\pm$ 0.0780** on the HGD dataset, reaching state-of-the-art performance among all compared methods.

Figures 10, 11, and 12 respectively illustrate the per-subject classification accuracy distribution of different models on the BCIC-IV-2a, BCIC-IV-2b, and HGD datasets. It can be observed that although STE-BiMamba exhibits some fluctuations in a few subjects, it demonstrates stable and superior performance in most subjects, with its overall average accuracy significantly outperforming other baseline models.

These results fully demonstrate the effectiveness and robustness of the proposed STE-BiMamba model in motor imagery EEG classification under both within-subject cross-session and cross-subject settings.

Ablation Study

To gain a deeper understanding of the core mechanisms of the STE-BiMamba model in cross-session motor imagery (MI) EEG decoding tasks, we conducted systematic ablation experiments on four datasets: our self-collected VR-MI dataset, BCIC-IV-2a, BCIC-IV-2b, and HGD. By successively removing key components (the Multi-scale temporal feature extraction module, the 3-Perspective Pooling module, and the Bi-Mamba module), we were able to clearly observe the contribution of each module to the final decoding performance.

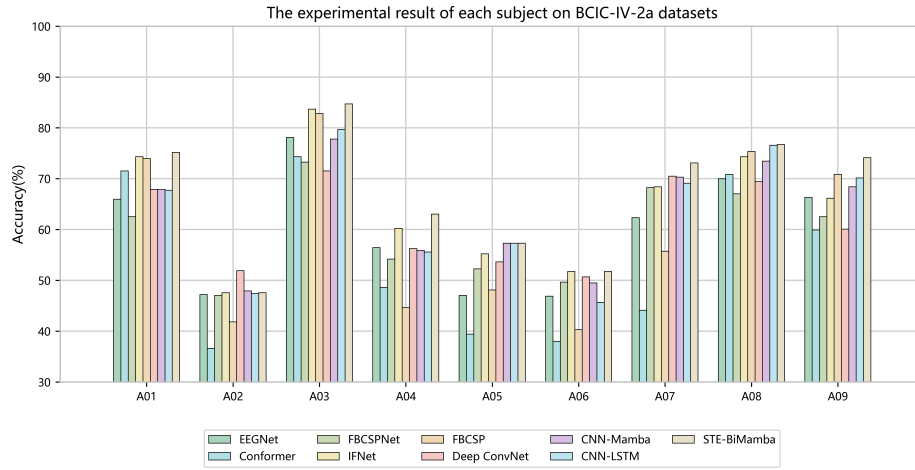


Figure 10. Per-subject classification accuracy comparison of STE-BiMamba and baseline models on the BCIC-IV-2a dataset (cross-subject setting).

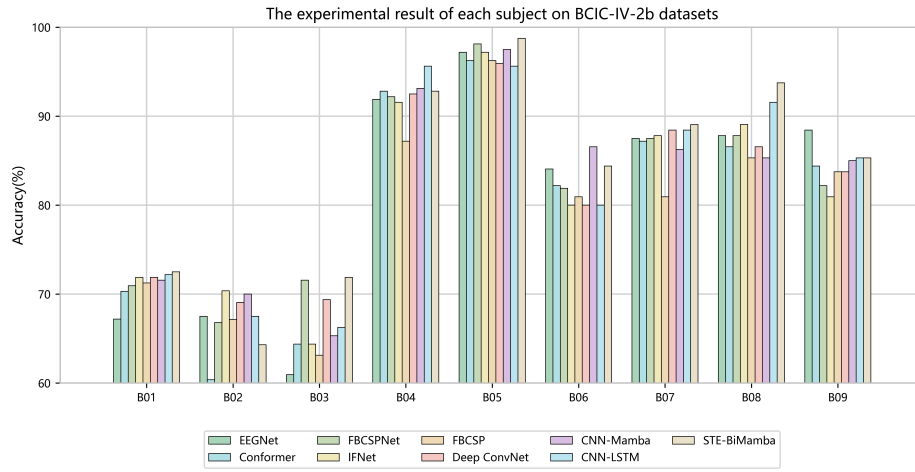


Figure 11. Per-subject classification accuracy comparison of STE-BiMamba and baseline models on the BCIC-IV-2b dataset (cross-subject setting).

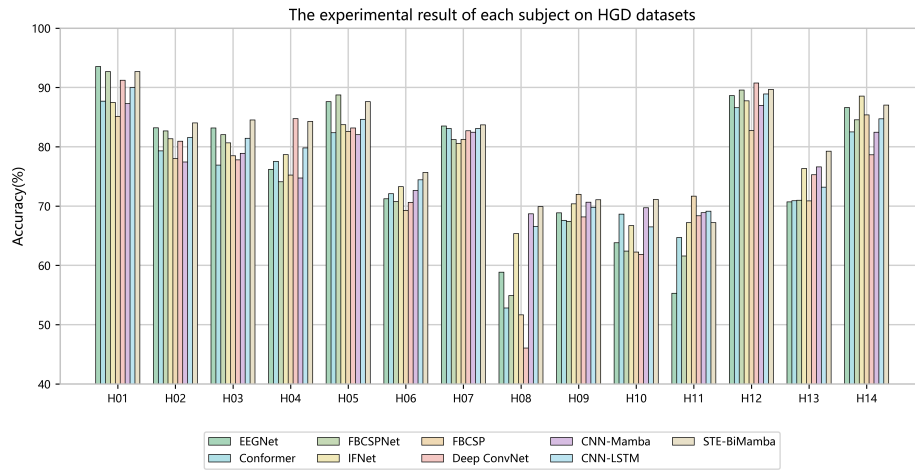


Figure 12. Per-subject classification accuracy comparison of STE-BiMamba and baseline models on the HGD dataset (cross-subject setting).

Table 4. Performance comparison of different models on BCIC-IV-2a, BCIC-IV-2b, and HGD datasets (cross-subject).

Model	BCIC-IV-2a			BCIC-IV-2b			HGD		
	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1 $\pm$ std
EEGNet	60.03% $\pm$ 10.67%**	0.4671 $\pm$ 0.1509	0.6032 $\pm$ 0.1115	81.39% $\pm$ 12.05%*	0.6278 $\pm$ 0.2411	0.8132 $\pm$ 0.1202	76.51% $\pm$ 11.48%**	0.6869 $\pm$ 0.1588	0.7660 $\pm$ 0.1149
Conformer	53.70% $\pm$ 14.66%***	0.3827 $\pm$ 0.2073	0.5335 $\pm$ 0.1535	80.49% $\pm$ 11.87%**	0.6097 $\pm$ 0.2376	0.8054 $\pm$ 0.1196	75.20% $\pm$ 9.32%**	0.6693 $\pm$ 0.1290	0.7514 $\pm$ 0.0932
FBCSPNet	59.63% $\pm$ 8.65%**	0.4617 $\pm$ 0.1223	0.5951 $\pm$ 0.0983	82.11% $\pm$ 9.95%*	0.6424 $\pm$ 0.1986	0.8213 $\pm$ 0.0983	75.98% $\pm$ 12.25%**	0.6797 $\pm$ 0.1557	0.7598 $\pm$ 0.1124
IFNet	64.62% $\pm$ 11.22%**	0.5283 $\pm$ 0.1586	0.6440 $\pm$ 0.1171	81.46% $\pm$ 10.31%*	0.6292 $\pm$ 0.2063	0.8145 $\pm$ 0.1034	77.72% $\pm$ 7.75%*	0.7029 $\pm$ 0.1072	0.7781 $\pm$ 0.0777
FBCSP	59.28% $\pm$ 15.56%***	0.4570 $\pm$ 0.2200	0.5895 $\pm$ 0.1651	79.54% $\pm$ 9.92%***	0.5910 $\pm$ 0.1982	0.7952 $\pm$ 0.0985	74.75% $\pm$ 9.12%***	0.6568 $\pm$ 0.1066	0.7478 $\pm$ 0.0910
Deep ConvNet	61.32% $\pm$ 8.06%**	0.4843 $\pm$ 0.1140	0.6100 $\pm$ 0.0923	81.95% $\pm$ 9.45%*	0.6389 $\pm$ 0.1889	0.8202 $\pm$ 0.0935	75.74% $\pm$ 11.65%**	0.6766 $\pm$ 0.1611	0.7577 $\pm$ 0.1163
CNN-Mamba	63.16% $\pm$ 10.15%*	0.5088 $\pm$ 0.1435	0.6341 $\pm$ 0.1072	82.29% $\pm$ 10.29%*	0.6458 $\pm$ 0.2059	0.8223 $\pm$ 0.1022	77.11% $\pm$ 6.25%*	0.6948 $\pm$ 0.0864	0.7711 $\pm$ 0.0662
CNN-LSTM	63.23% $\pm$ 11.57%**	0.5098 $\pm$ 0.1636	0.6321 $\pm$ 0.1218	82.50% $\pm$ 10.91%	0.6500 $\pm$ 0.2182	0.8241 $\pm$ 0.1091	78.13% $\pm$ 7.79%*	0.7083 $\pm$ 0.1078	0.7821 $\pm$ 0.0783
STE-BiMamba	67.05% $\pm$ 11.94%	0.5610 $\pm$ 0.1588	0.6699 $\pm$ 0.1188	83.64% $\pm$ 10.59%	0.6729 $\pm$ 0.2192	0.8366 $\pm$ 0.1102	80.55% $\pm$ 7.88%	0.7410 $\pm$ 0.1051	0.8062 $\pm$ 0.0780

Note: Cyan cells denote the proposed STE-BiMamba model. All p-values are Bonferroni-corrected for multiple comparisons (5 comparisons) based on statistical difference tests of accuracy compared to the proposed model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Table 5. Ablation study results of STE-BiMamba on VR-MI, BCIC-IV-2a, BCIC-IV-2b, and HGD datasets (cross-session).

Dataset	Method	Avg. Acc. $\pm$ std	Avg. Kappa $\pm$ std	Avg. F1-score $\pm$ std	p-value
VR-MI	Ours w/o Multi-scale	96.38% $\pm$ 3.31%	0.9275 $\pm$ 0.0661	0.9636 $\pm$ 0.0324	ns
	Ours w/o 3-Perspective Pooling	95.75% $\pm$ 3.80%	0.9150 $\pm$ 0.0760	0.9573 $\pm$ 0.0378	<0.05*
	Ours w/o Bi-Mamba	93.88% $\pm$ 4.35%	0.8775 $\pm$ 0.0887	0.9386 $\pm$ 0.0439	<0.01**
	Ours	97.50% $\pm$ 3.35%	0.9500 $\pm$ 0.0671	0.9749 $\pm$ 0.0338	-
BCIC-IV-2a	Ours w/o Multi-scale	81.82% $\pm$ 10.65%	0.7452 $\pm$ 0.1462	0.8065 $\pm$ 0.1095	<0.05*
	Ours w/o 3-Perspective Pooling	79.94% $\pm$ 10.58%	0.7168 $\pm$ 0.1495	0.7882 $\pm$ 0.1118	<0.01**
	Ours w/o Bi-Mamba	78.53% $\pm$ 10.82%	0.6841 $\pm$ 0.1538	0.7737 $\pm$ 0.1146	<0.01**
	Ours	82.95% $\pm$ 10.52%	0.7726 $\pm$ 0.1403	0.8280 $\pm$ 0.1068	-
BCIC-IV-2b	Ours w/o Multi-scale	87.58% $\pm$ 9.02%	0.7575 $\pm$ 0.1748	0.8732 $\pm$ 0.0935	<0.01**
	Ours w/o 3-Perspective Pooling	85.65% $\pm$ 9.08%	0.7342 $\pm$ 0.1765	0.8548 $\pm$ 0.0948	<0.001***
	Ours w/o Bi-Mamba	83.87% $\pm$ 9.31%	0.7073 $\pm$ 0.1802	0.8385 $\pm$ 0.0972	<0.001***
	Ours	89.20% $\pm$ 8.42%	0.7840 $\pm$ 0.1675	0.8923 $\pm$ 0.0844	-
HGD	Ours w/o Multi-scale	94.41% $\pm$ 3.02%	0.9265 $\pm$ 0.0412	0.9438 $\pm$ 0.0315	ns
	Ours w/o 3-Perspective Pooling	93.18% $\pm$ 3.18%	0.9094 $\pm$ 0.0435	0.9305 $\pm$ 0.0338	<0.01**
	Ours w/o Bi-Mamba	92.52% $\pm$ 3.41%	0.8867 $\pm$ 0.0468	0.9234 $\pm$ 0.0364	<0.01**
	Ours	95.67% $\pm$ 2.87%	0.9422 $\pm$ 0.0383	0.9564 $\pm$ 0.0291	-

Note: Cyan cells denote the complete model (Ours) in each dataset group. All p-values are Bonferroni-corrected based on statistical difference tests of accuracy compared to the proposed model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

As shown in Table 5, the complete STE-BiMamba model (Ours) achieved the best performance across all four datasets, with statistically significant improvements compared to its ablated variants (most comparisons $p < 0.05$ or better after Bonferroni correction). Removing any single component led to noticeable declines in accuracy, Kappa coefficient, and F1-score. In particular, the removal of the Bi-Mamba module caused the most significant performance degradation on most datasets ($p < 0.01$ or $p < 0.001$), indicating that bidirectional state-space modeling played a critical role in capturing long-term temporal dependencies in EEG signals. The absence of the Multi-scale feature extraction module weakened the model’s ability to perceive brainwave patterns at different time scales, while the removal of the 3-Perspective Pooling module also substantially reduced the expressive power of spatio-temporal features. These ablation results, supported by statistical difference tests, collectively verified that the synergistic design of multi-scale temporal extraction, three-view statistical pooling, and bidirectional Mamba modeling significantly enhanced the robustness and generalization capability of STE-BiMamba in real-world MI-EEG decoding scenarios.

We further selected the 3rd subject (A03) from the BCIC-IV-2a dataset, the 5th subject (B05) from the BCIC-IV-2b dataset, and the 12th subject (H12) from the HGD dataset to perform t-SNE two-dimensional visualization of the features learned by the complete model (Ours) and each ablation variant (w/o Multi-scale, w/o 3-Perspective Pooling, w/o Bi-Mamba).

As shown in Fig. 13, compared with the ablation models, the feature points of the complete STE-BiMamba model exhibited clearer inter-class separation boundaries and tighter intra-class clustering. Specifically, the Ours w/o Multi-scale model showed obvious boundary blurring between pairwise classes; the Ours w/o 3-Perspective Pooling model exhibited difficulty in distinguishing multiple classes; and the Ours w/o Bi-Mamba model had the most chaotic feature distribution, with severe overall class mixing that made it almost impossible to clearly separate different motor imagery classes. This phenomenon was consistent across all three datasets, fully demonstrating that the synergistic effect of Multi-scale feature extraction, 3-Perspective Pooling, and Bi-Mamba bidirectional state-space modeling significantly enhanced the model’s ability to distinguish subtle spatiotemporal differences in motor imagery EEG signals, thereby providing more discriminative feature representations for robust decoding in cross-session scenarios.

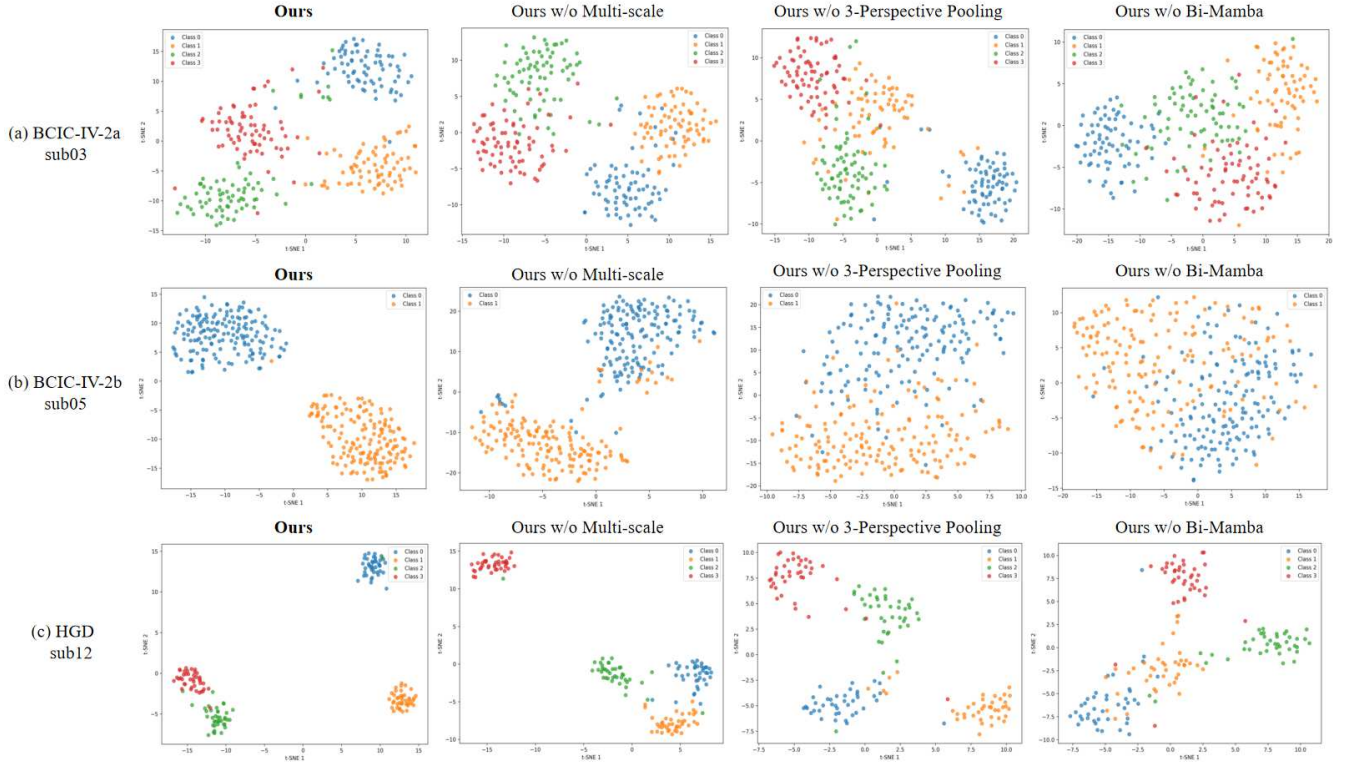


Figure 13. t-SNE visualization of learned features. (a) Subject A03 on BCIC-IV-2a; (b) Subject B05 on BCIC-IV-2b; (c) Subject H12 on HGD.

To further visually analyze the impact of each key module on the classification decision boundary, we plotted the confusion matrices of the complete model and each ablation variant on the same subjects. As shown in Fig. 14, the confusion matrices of the complete STE-BiMamba model (Ours) were highly concentrated on the main diagonal across all three datasets, with almost all samples correctly classified into their respective motor imagery classes, indicating extremely high classification accuracy and extremely low misclassification rates.

Specifically, a small number of off-diagonal elements began to appear in the Ours w/o Multi-scale model, indicating blurred boundaries between some classes and highlighting the importance of the multi-scale feature extraction module in capturing brainwave patterns at different time scales.

The confusion in the Ours w/o 3-Perspective Pooling model was more pronounced, with cross-misclassifications among multiple classes, which reflected the key role of three-perspective pooling in enhancing the discriminability of multi-dimensional spatiotemporal features. The Ours w/o Bi-Mamba model exhibited the most severe confusion, with a large number of samples misclassified into other classes and significant overall inter-class mixing. This further confirmed the core contribution of Bi-Mamba bidirectional state-space modeling in capturing long-term sequential dependencies.

These confusion matrix results were highly consistent with the preceding t-SNE visualizations, once again verifying that the synergistic design of multi-scale feature extraction, multi-perspective pooling, and bidirectional sequence modeling in

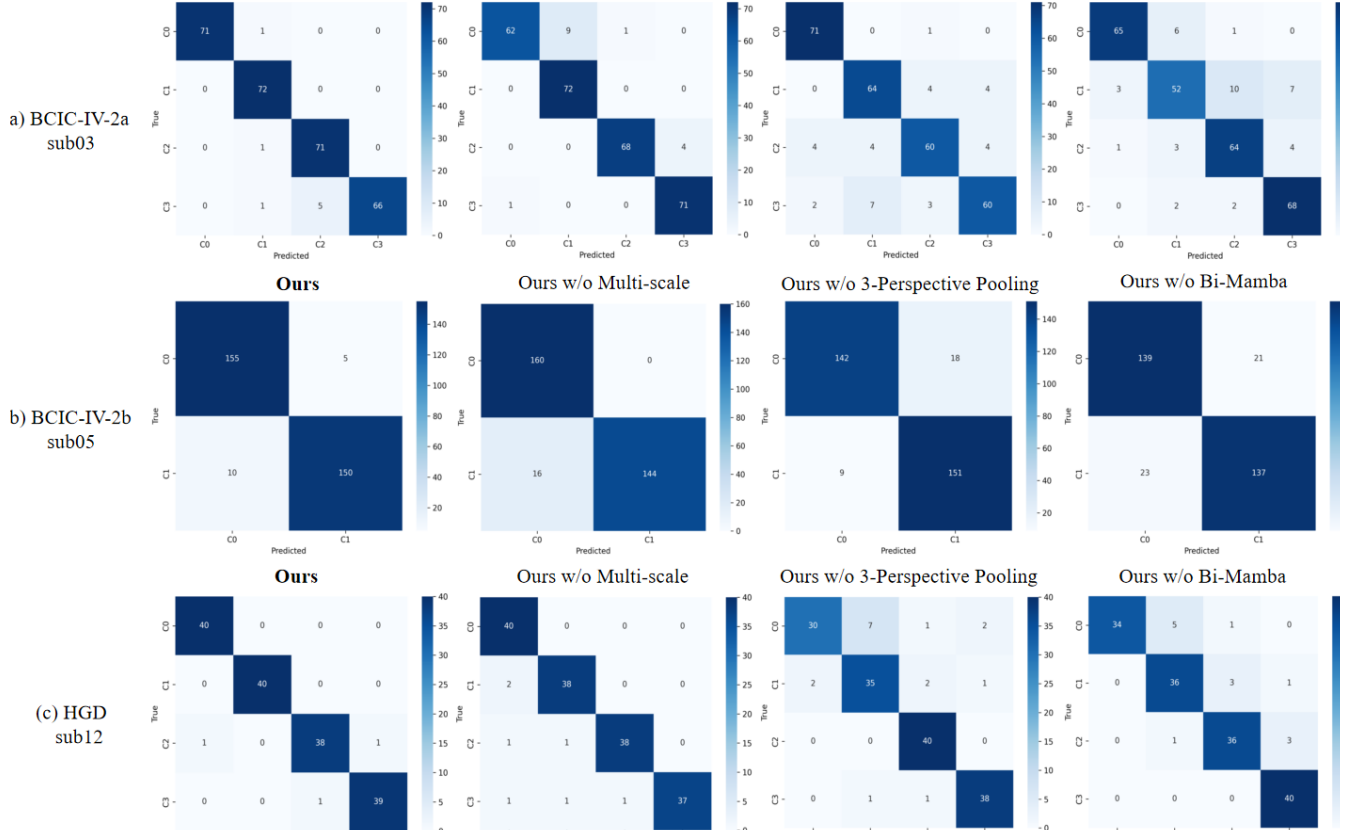


Figure 14. Confusion matrices of STE-BiMamba and its ablation variants. (a) Subject A03 on BCIC-IV-2a; (b) Subject B05 on BCIC-IV-2b; (c) Subject H12 on HGD.

STE-BiMamba significantly improved the discriminative power and robustness of the model in cross-session MI-EEG decoding tasks.

In the further group-wise ablation experiments, we analyzed the impact of removing different combinations of modules on model performance. As shown in Fig. 15, when multiple key modules are removed simultaneously, the model performance exhibits a more pronounced decline. For example, on the VR-MI dataset, after removing the Multi-scale and 3-Perspective Pooling modules (w/o MS+3PP), the accuracy dropped to 94.88%; after removing the Multi-scale and Bi-Mamba modules (w/o MS+BM), it dropped to 93.38%; after removing the 3-Perspective Pooling and Bi-Mamba modules (w/o 3PP+BM), it was 92.75%; and when all three modules were removed simultaneously (w/o All), it dropped to 92.25%. The complete model (Ours) achieved the highest average accuracy of 97.50% on the VR-MI dataset.

Similar phenomena were also observed on the BCIC-IV-2a, BCIC-IV-2b, and HGD datasets. Overall, removing any module combination led to performance degradation, with combinations involving the simultaneous removal of Bi-Mamba and other modules typically causing greater performance losses. This further confirms the critical role of the Bi-Mamba module in capturing long-term temporal dependencies in EEG signals. The synergistic interaction among the modules plays a vital role in enhancing the overall performance of the STE-BiMamba model.

Hyperparameter Tuning and Model Selection

As shown in Table 6, we performed a comprehensive hyperparameter tuning for the proposed STE-BiMamba model on the BCIC-IV-2a dataset under the intra-subject setting. A total of 21 representative configurations were evaluated by systematically varying the key hyperparameters, including embedding dimension d , multi-scale temporal kernel sizes k_{temp} , statistical pooling parameters (s_{pool} , str_{pool}), number of Mamba layers L , state dimension d_{state} , expansion factor e , local convolution kernel size k_{conv} , dropout rate p_{drop} , and spatial convolution output channels c_{spatial} .

The optimal configuration (Model 20) was selected based on the highest average classification accuracy while maintaining reasonable model complexity. The final adopted hyperparameters are: $d = 32$, $k_{\text{temp}} = [15, 25, 51, 65]$, $s_{\text{pool}} = 50$, $\text{str}_{\text{pool}} = 15$, $L = 4$, $d_{\text{state}} = 16$, $e = 2$, $k_{\text{conv}} = 4$, $p_{\text{drop}} = 0.5$, and $c_{\text{spatial}} = d$. This setting achieves the best performance with an average accuracy of $82.95\% \pm 10.52$ and Kappa of 0.7726 ± 0.1403 .

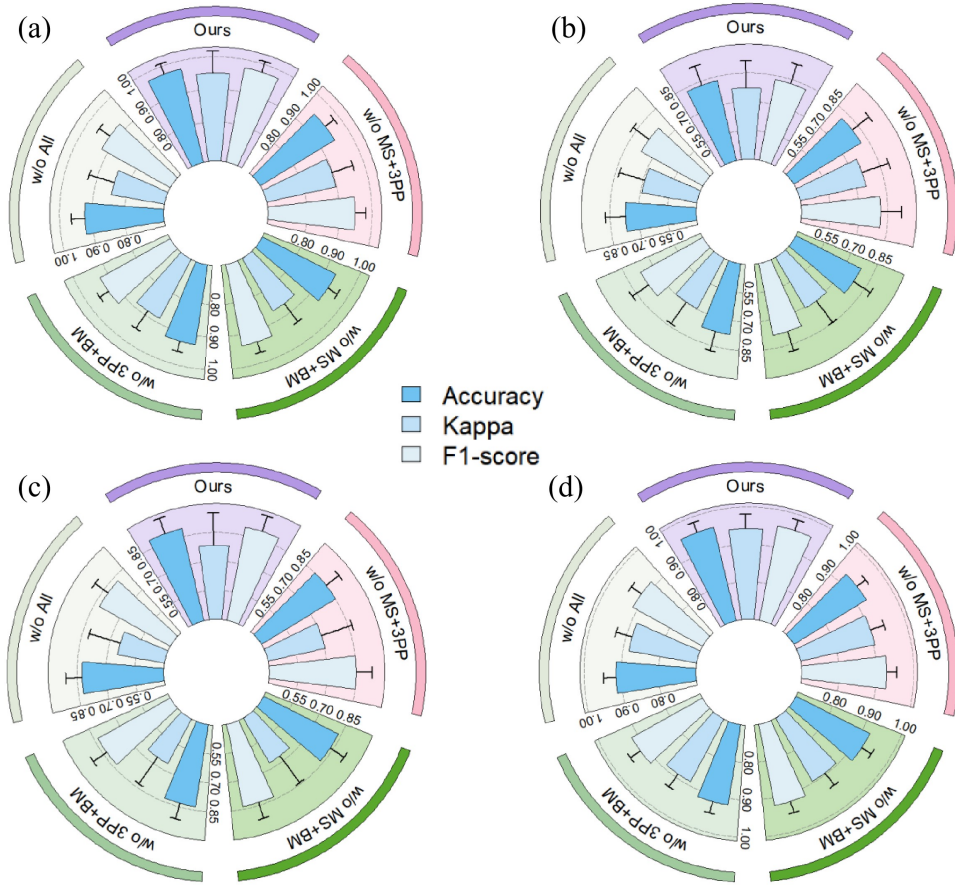


Figure 15. Group-wise ablation results of STE-BiMamba when removing different module combinations (w/o MS+3PP, w/o MS+BM, w/o 3PP+BM, w/o All) compared with the complete model. (a) VR-MI dataset; (b) BCIC-IV-2a dataset; (c) BCIC-IV-2b dataset; (d) HGD dataset.

All models were trained and evaluated under identical experimental conditions to ensure fair and reproducible comparison.

Efficiency Analysis

To comprehensively evaluate the deployment potential of the STE-BiMamba model in practical brain-computer interface (BCI) systems, we paid special attention to its computational efficiency and model complexity. Since the core of the model incorporates the Mamba state space model, we performed a fair comparison with the classical Transformer-based Conformer model in terms of inference time and parameter count. The experiments were conducted under the same hardware environment, with all models using a batch size of 1 for online inference to simulate real cross-session MI-EEG decoding scenarios.

Table 7 presents the quantitative comparison results between the two models.

From the table, it can be clearly seen that the proposed STE-BiMamba model exhibited significant advantages in efficiency: its single inference time was only approximately 52% of that of Conformer (48 ms vs. 93 ms), while the parameter count was substantially reduced to approximately 28% of Conformer (223.31K vs. 789.82K). This remarkable efficiency improvement was mainly attributed to the linear complexity characteristic of the Mamba architecture. Unlike the traditional Transformer architecture, the self-attention mechanism in Conformer requires computation of a full attention matrix, resulting in both time and space complexity of $O(n^2)$ (where n is the sequence length), which incurs extremely high computational cost when processing long-sequence EEG signals. In contrast, Mamba reduces the complexity of sequence modeling to linear $O(n)$ through its selective state space mechanism (Selective SSM) while still maintaining powerful long-range dependency modeling capability. Therefore, under the same hardware conditions, STE-BiMamba achieved much lower inference latency and memory usage than Conformer.

These efficiency gains are of great significance for the real-time deployment of practical MI-EEG brain-computer interface systems, enabling STE-BiMamba to achieve low-latency, high-reliability motor imagery decoding on resource-constrained edge devices or portable BCI systems.

Table 6. Hyperparameter Tuning of STE-BiMamba Model on BCIC-IV-2a Dataset (Intra-Subject Setting)

Model	d	k_{temp}	s_{pool}	str_{pool}	L	d_{state}	e	k_{conv}	p_{drop}	c_{spatial}	Avg. ACC. $\pm$ std	Avg. Kappa $\pm$ std
1	16	9,17,33,49	25	10	4	8	1	2	0.1	$d/2$	81.60% $\pm$ 10.54%	0.7546 $\pm$ 0.1405
2	32	9,17,33,49	25	10	4	8	1	2	0.1	$d/2$	81.71% $\pm$ 10.55%	0.7562 $\pm$ 0.1407
3	64	9,17,33,49	25	10	4	8	1	2	0.1	$d/2$	81.67% $\pm$ 10.60%	0.7556 $\pm$ 0.1414
4	32	15,25,51,65	25	10	4	8	1	2	0.1	$d/2$	81.87% $\pm$ 10.51%	0.7582 $\pm$ 0.1401
5	32	24,41,61,81	25	10	4	8	1	2	0.1	$d/2$	81.79% $\pm$ 10.53%	0.7572 $\pm$ 0.1404
6	32	15,25,51,65	50	10	4	8	1	2	0.1	$d/2$	82.18% $\pm$ 10.58%	0.7623 $\pm$ 0.1411
7	32	15,25,51,65	75	10	4	8	1	2	0.1	$d/2$	82.10% $\pm$ 10.53%	0.7613 $\pm$ 0.1405
8	32	15,25,51,65	50	15	4	8	1	2	0.1	$d/2$	82.22% $\pm$ 10.51%	0.7629 $\pm$ 0.1401
9	32	15,25,51,65	50	25	4	8	1	2	0.1	$d/2$	82.02% $\pm$ 10.60%	0.7603 $\pm$ 0.1413
10	32	15,25,51,65	50	15	6	8	1	2	0.1	$d/2$	81.98% $\pm$ 10.56%	0.7598 $\pm$ 0.1408
11	32	15,25,51,65	50	15	8	8	1	2	0.1	$d/2$	81.91% $\pm$ 10.52%	0.7588 $\pm$ 0.1403
12	32	15,25,51,65	50	15	4	16	1	2	0.1	$d/2$	82.37% $\pm$ 10.60%	0.7649 $\pm$ 0.1413
13	32	15,25,51,65	50	15	4	32	1	2	0.1	$d/2$	82.29% $\pm$ 10.54%	0.7639 $\pm$ 0.1405
14	32	15,25,51,65	50	15	4	16	2	2	0.1	$d/2$	82.48% $\pm$ 10.53%	0.7664 $\pm$ 0.1404
15	32	15,25,51,65	50	15	4	16	3	2	0.1	$d/2$	82.41% $\pm$ 10.52%	0.7654 $\pm$ 0.1407
16	32	15,25,51,65	50	15	4	16	2	4	0.1	$d/2$	82.60% $\pm$ 10.55%	0.7680 $\pm$ 0.1403
17	32	15,25,51,65	50	15	4	16	2	8	0.1	$d/2$	82.56% $\pm$ 10.51%	0.7675 $\pm$ 0.1401
18	32	15,25,51,65	50	15	4	16	2	4	0.3	$d/2$	82.68% $\pm$ 10.56%	0.7690 $\pm$ 0.1408
19	32	15,25,51,65	50	15	4	16	2	4	0.5	$d/2$	82.75% $\pm$ 10.55%	0.7701 $\pm$ 0.1407
20	32	15,25,51,65	50	15	4	16	2	4	0.5	d	82.95%$\pm$10.52%	0.7726$\pm$0.1403
21	32	15,25,51,65	50	15	4	16	2	4	0.5	$2d$	82.79% $\pm$ 10.54%	0.7706 $\pm$ 0.1405

Note: Cyan cells indicate the model with optimal performance. Gray cells indicate the best parameter choice within the current hyperparameter group.

Table 7. Efficiency comparison between STE-BiMamba (Ours) and Conformer.

Model	Inference Time (ms)	Parameters (K)
Conformer	93	789.82
Ours	48	223.31

Note: Cyan cells denote the results of STE-BiMamba.

Visualization of Learned Features

We further employed Hybrid CAM (Hybrid Class Activation Mapping) technology⁵⁴ to perform topographic map visualization of the raw EEG data and the learned features after model activation. The same three representative subjects were selected: subject A03 from the BCIC-IV-2a dataset, subject B05 from the BCIC-IV-2b dataset, and subject H12 from the HGD dataset.

As shown in Fig. 16, the topographic maps of the raw EEG data exhibited relatively dispersed and diffuse activation patterns, with signal energy distributed fairly uniformly across scalp regions and noticeable noise interference, making it difficult to clearly distinguish spatial differences among different motor imagery classes. In contrast, the learned features after Hybrid CAM activation showed significant changes: the energy distribution of the feature maps became highly focused and structured, primarily concentrated in the sensorimotor cortex regions closely related to motor imagery. At the same time, the spatial activation patterns among different classes became much more distinct, with significantly amplified inter-class differences and markedly improved intra-class consistency.

This before-and-after comparison vividly demonstrated that, through the synergistic effect of Multi-scale feature extraction, 3-Perspective Pooling, and Bi-Mamba bidirectional sequence modeling, STE-BiMamba successfully transformed the subtle spatiotemporal patterns hidden in the raw EEG signals into highly discriminative neural representations.

Discussion and Conclusions

To enhance the decoding performance of multi-channel MI-EEG signals in challenging cross-session and cross-subject scenarios, this paper proposes STE-BiMamba, a novel end-to-end model that integrates multi-scale temporal feature extraction, three-view statistical pooling, and bidirectional Mamba sequence modeling.

By effectively capturing both fine-grained transient patterns and long-range spatiotemporal dependencies while maintaining linear computational complexity, the model overcomes key limitations of existing CNN- and Transformer-based approaches.

Experimental results on three public datasets (BCIC-IV-2a, BCIC-IV-2b, and HGD) under the cross-session setting demon-

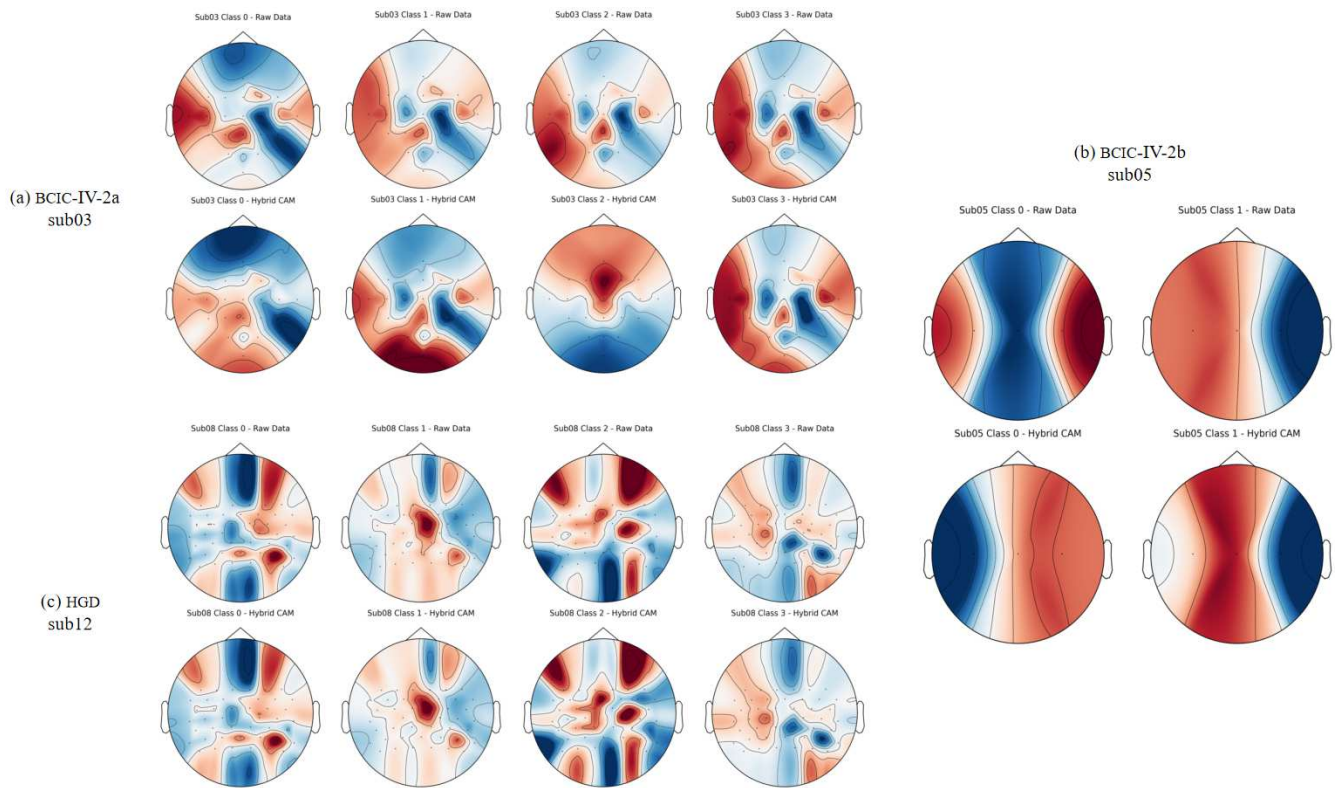


Figure 16. Brain topographic maps before and after feature learning. (a) Subject A03 on BCIC-IV-2a; (b) Subject B05 on BCIC-IV-2b; (c) Subject H12 on HGD.

strate that STE-BiMamba consistently achieves state-of-the-art performance, significantly outperforming strong baselines such as CNN-Mamba and Conformer in accuracy, Kappa coefficient, and F1-score. Ablation studies, t-SNE visualizations, confusion matrices, and Hybrid CAM topographic maps further confirm the critical contribution and physiological interpretability of each module. In addition, the model reduces inference time to 48 ms (52% of Conformer) and parameter count to 223.31K (28% of Conformer), highlighting its strong potential for real-time BCI deployment.

In conclusion, STE-BiMamba provides an accurate, efficient, and interpretable solution for MI-EEG decoding. Future work will focus on cross-subject MI-EEG decoding to further improve the model's generalization capability across different individuals.

References

1. Chaudhary, U., Birbaumer, N. & Ramos-Murguialday, A. Brain–Computer Interfaces for Communication and Rehabilitation. *Nat. Rev. Neurol.* **12**, 513–525, DOI: [10.1038/nrneurol.2016.113](https://doi.org/10.1038/nrneurol.2016.113) (2016).
2. Birbaumer, N. Breaking the Silence: Brain–Computer Interfaces (BCI) for Communication and Motor Control. *Psychophysiology* **43**, 517–532, DOI: [10.1111/j.1469-8986.2006.00456.x](https://doi.org/10.1111/j.1469-8986.2006.00456.x) (2006).
3. Vaid, S., Singh, P. & Kaur, C. EEG Signal Analysis for BCI Interface: A Review. In *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 143–147, DOI: [10.1109/acct.2015.72](https://doi.org/10.1109/acct.2015.72) (IEEE, Haryana, India, 2015).
4. Rakshit, A., Konar, A. & Nagar, A. K. A Hybrid Brain-Computer Interface for Closed-Loop Position Control of a Robot Arm. *IEEE/CAA J. Autom. Sinica* **7**, 1344–1360, DOI: [10.1109/jas.2020.1003336](https://doi.org/10.1109/jas.2020.1003336) (2020).
5. Tang, Z. *et al.* An Upper-Limb Rehabilitation Exoskeleton System Controlled by MI Recognition Model With Deep Emphasized Informative Features in a VR Scene. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 4390–4401, DOI: [10.1109/tnsre.2023.3329059](https://doi.org/10.1109/tnsre.2023.3329059) (2023).
6. Saravanakumar, D. & Ramasubba Reddy, M. A Virtual Speller System Using SSVEP and Electrooculogram. *Adv. Eng. Informatics* **44**, 101059, DOI: [10.1016/j.aei.2020.101059](https://doi.org/10.1016/j.aei.2020.101059) (2020).

7. Wierzgała, P., Zapała, D., Wojcik, G. M. & Masiak, J. Most Popular Signal Processing Methods in Motor-Imagery BCI: A Review and Meta-Analysis. *Front. Neuroinformatics* **12**, 78, DOI: [10.3389/fninf.2018.00078](https://doi.org/10.3389/fninf.2018.00078) (2018).
8. Khan, M. A., Das, R., Iversen, H. K. & Puthusserypady, S. Review on Motor Imagery Based BCI Systems for Upper Limb Post-Stroke Neurorehabilitation: From Designing to Application. *Comput. Biol. Medicine* **123**, 103843, DOI: [10.1016/j.combiomed.2020.103843](https://doi.org/10.1016/j.combiomed.2020.103843) (2020).
9. Mauro, M. C. *et al.* Restoring of Interhemispheric Symmetry in Patients With Stroke Following Bilateral or Unilateral Robot-Assisted Upper-Limb Rehabilitation: A Pilot Randomized Controlled Trial. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **32**, 3590–3602, DOI: [10.1109/tnsre.2024.3460485](https://doi.org/10.1109/tnsre.2024.3460485) (2024).
10. Jackson, A. F. & Bolger, D. J. The Neurophysiological Bases of EEG and EEG Measurement: A Review for the Rest of Us. *Psychophysiology* **51**, 1061–1071, DOI: [10.1111/psyp.12283](https://doi.org/10.1111/psyp.12283) (2014).
11. Subasi, A. & Ismail Gursoy, M. EEG Signal Classification Using PCA, ICA, LDA and Support Vector Machines. *Expert. Syst. with Appl.* **37**, 8659–8666, DOI: [10.1016/j.eswa.2010.06.065](https://doi.org/10.1016/j.eswa.2010.06.065) (2010).
12. Richhariya, B. & Tanveer, M. EEG Signal Classification Using Universon Support Vector Machine. *Expert. Syst. with Appl.* **106**, 169–182, DOI: [10.1016/j.eswa.2018.03.053](https://doi.org/10.1016/j.eswa.2018.03.053) (2018).
13. Jiang, X., Meng, L., Chen, X., Xu, Y. & Wu, D. CSP-Net: Common Spatial Pattern Empowered Neural Networks for EEG-Based Motor Imagery Classification. *Knowledge-Based Syst.* **305**, 112668, DOI: [10.1016/j.knosys.2024.112668](https://doi.org/10.1016/j.knosys.2024.112668) (2024).
14. An, Y., Lam, H. K. & Ling, S. H. Multi-Classification for EEG Motor Imagery Signals Using Data Evaluation-Based Auto-Selected Regularized FBCSP and Convolutional Neural Network. *Neural Comput. Appl.* **35**, 12001–12027, DOI: [10.1007/s00521-023-08336-z](https://doi.org/10.1007/s00521-023-08336-z) (2023).
15. Rajwal, S. & Aggarwal, S. Convolutional Neural Network-Based EEG Signal Analysis: A Systematic Review. *Arch. Comput. Methods Eng.* **30**, 3585–3615, DOI: [10.1007/s11831-023-09920-1](https://doi.org/10.1007/s11831-023-09920-1) (2023).
16. Lawhern, V. J. *et al.* EEGNet: A Compact Convolutional Network for EEG-Based Brain-Computer Interfaces. *J. Neural Eng.* **15**, 056013, DOI: [10.1088/1741-2552/aace8c](https://doi.org/10.1088/1741-2552/aace8c) (2018).
17. Wang, T., Dong, E., Du, S. & Jia, C. A Shallow Convolutional Neural Network for Classifying MI-EEG. In *2019 Chinese Automation Congress (CAC)*, 5837–5841, DOI: [10.1109/cac48633.2019.8996981](https://doi.org/10.1109/cac48633.2019.8996981) (IEEE, Hangzhou, China, 2019).
18. Schirrmeister, R. T. *et al.* Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Hum. Brain Mapp.* **38**, 5391–5420, DOI: [10.1002/hbm.23730](https://doi.org/10.1002/hbm.23730) (2017).
19. Emsawas, T., Morita, T., Kimura, T., Fukui, K. & Numao, M. Multi-Kernel Temporal and Spatial Convolution for EEG-Based Emotion Classification. *Sensors* **22**, 8250, DOI: [10.3390/s22218250](https://doi.org/10.3390/s22218250) (2022).
20. Altaheri, H., Karray, F. & Karimi, A.-H. Temporal Convolutional Transformer for EEG Based Motor Imagery Decoding. *Sci. Reports* **15**, 32959, DOI: [10.1038/s41598-025-16219-7](https://doi.org/10.1038/s41598-025-16219-7) (2025).
21. Rivas, F., Sierra-Garcia, J. E. & Camara, J. M. Comparison of LSTM- and GRU-Type RNN Networks for Attention and Meditation Prediction on Raw EEG Data from Low-Cost Headsets. *Electronics* **14**, 707, DOI: [10.3390/electronics14040707](https://doi.org/10.3390/electronics14040707) (2025).
22. Du, X. *et al.* An Efficient LSTM Network for Emotion Recognition From Multichannel EEG Signals. *IEEE Transactions on Affect. Comput.* **13**, 1528–1540, DOI: [10.1109/taffc.2020.3013711](https://doi.org/10.1109/taffc.2020.3013711) (2022).
23. Ouyang, J., Wu, M., Li, X., Deng, H. & Wu, D. BRIEDGE: EEG-Adaptive Edge AI for Multi-Brain to Multi-Robot Interaction. Preprint at <https://arxiv.org/abs/2403.15432> (2024).
24. Wang, G. *et al.* EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals. In *Advances in Neural Information Processing Systems* **37**, 39249–39280, DOI: [10.52202/079017-1239](https://doi.org/10.52202/079017-1239) (2024).
25. Pfeffer, M. A., Wong, J. K. W. & Ling, S. H. Trends and Limitations in Transformer-Based BCI Research. *Appl. Sci.* **15**, 11150, DOI: [10.3390/app152011150](https://doi.org/10.3390/app152011150) (2025).
26. Zhou, X., Huang, D., Peng, X. & Yin, L. miMamba: EEG-Based Emotion Recognition with Multi-Scale Inverted Mamba Models. *IEEE Transactions on Affect. Comput.* **16**, 3266–3278, DOI: [10.1109/taffc.2025.3587443](https://doi.org/10.1109/taffc.2025.3587443) (2025).
27. Yu, W. & Wang, X. MambaOut: Do We Really Need Mamba for Vision? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4484–4496, DOI: [10.1109/cvpr52734.2025.00423](https://doi.org/10.1109/cvpr52734.2025.00423) (2025).

28. Hossain, K. M., Islam, M. A., Hossain, S., Nijholt, A. & Ahad, M. A. R. Status of Deep Learning for EEG-Based Brain–Computer Interface Applications. *Front. Comput. Neurosci.* **16**, 1006763, DOI: [10.3389/fncom.2022.1006763](https://doi.org/10.3389/fncom.2022.1006763) (2023).
29. Kim, M. & Im, C.-H. HiRENet: Novel Convolutional Neural Network Architecture Using Hilbert-Transformed and Raw Electroencephalogram (EEG) for Subject-Independent Emotion Classification. *Comput. Biol. Medicine* **178**, 108788, DOI: [10.1016/j.compbiomed.2024.108788](https://doi.org/10.1016/j.compbiomed.2024.108788) (2024).
30. Wang, J., Yao, L. & Wang, Y. IFNet: An Interactive Frequency Convolutional Neural Network for Enhancing Motor Imagery Decoding From EEG. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 1900–1911, DOI: [10.1109/tnsre.2023.3257319](https://doi.org/10.1109/tnsre.2023.3257319) (2023).
31. Melinda, M. *et al.* Feature-Based EEG Classification: A Comparative Study of EMD and EWT With EEGNet and Shallow FBCSP ConvNet. *IEEE Access* **13**, 185095–185110, DOI: [10.1109/access.2025.3625234](https://doi.org/10.1109/access.2025.3625234) (2025).
32. Xu, Y. *et al.* Autism Spectrum Disorder Diagnosis with EEG Signals Using Time Series Maps of Brain Functional Connectivity and a Combined CNN–LSTM Model. *Comput. Methods Programs Biomed.* **250**, 108196, DOI: [10.1016/j.cmpb.2024.108196](https://doi.org/10.1016/j.cmpb.2024.108196) (2024).
33. Schirrmeister, R. T. *et al.* Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Hum. Brain Mapp.* **38**, 5391–5420, DOI: [10.1002/hbm.23730](https://doi.org/10.1002/hbm.23730) (2017).
34. Tveitstøl, T. *et al.* Introducing Region Based Pooling for Handling a Varied Number of EEG Channels for Deep Learning Models. *Front. Neuroinformatics* **17**, 1272791, DOI: [10.3389/fninf.2023.1272791](https://doi.org/10.3389/fninf.2023.1272791) (2024).
35. Wang, Z., Li, S., Chen, X. & Wu, D. MVCNet: Multi-View Contrastive Network for Motor Imagery Classification. *Knowledge-Based Syst.* 114205, DOI: [10.1016/j.knosys.2025.114205](https://doi.org/10.1016/j.knosys.2025.114205) (2025).
36. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep Learning for Electroencephalogram (EEG) Classification Tasks: A Review. *J. Neural Eng.* **16**, 031001, DOI: [10.1088/1741-2552/ab0ab5](https://doi.org/10.1088/1741-2552/ab0ab5) (2019).
37. Han, C. *et al.* A Spatial-Spectral and Temporal Dual Prototype Network for Motor Imagery Brain-Computer Interface. *Knowledge-Based Syst.* **315**, 113315, DOI: [10.1016/j.knosys.2025.113315](https://doi.org/10.1016/j.knosys.2025.113315) (2025).
38. Zhu, T. *et al.* A Study of Motor Imagery EEG Classification Based on Feature Fusion and Attentional Mechanisms. *Front. Hum. Neurosci.* **19**, 1611229, DOI: [10.3389/fnhum.2025.1611229](https://doi.org/10.3389/fnhum.2025.1611229) (2025).
39. Gu, A. & Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. Preprint at <https://arxiv.org/abs/2312.00752> (2023).
40. Wang, Z. *et al.* Is Mamba Effective for Time Series Forecasting? *Neurocomputing* **619**, 129178, DOI: [10.1016/j.neucom.2024.129178](https://doi.org/10.1016/j.neucom.2024.129178) (2025).
41. Waleffe, R. *et al.* An Empirical Study of Mamba-Based Language Models. Preprint at <https://arxiv.org/abs/2406.07887> (2024).
42. Yang, X. & Jia, Z. Spatial-Temporal Mamba Network for EEG-Based Motor Imagery Classification. In *International Conference on Advanced Data Mining and Applications*, 418–432, DOI: [10.1007/978-981-96-0821-8_28](https://doi.org/10.1007/978-981-96-0821-8_28) (Springer Nature Singapore, Singapore, 2024).
43. Vaswani, A. *et al.* Attention Is All You Need. In *Advances in Neural Information Processing Systems*, vol. 30 (2017).
44. Zhang, H. & Li, H. Transformer-Based EEG Decoding: A Survey. Preprint at <https://arxiv.org/abs/2507.02320> (2025).
45. Vafaei, E. & Hosseini, M. Transformers in EEG Analysis: A Review of Architectures and Applications in Motor Imagery, Seizure, and Emotion Classification. *Sensors* **25**, 1293, DOI: [10.3390/s25051293](https://doi.org/10.3390/s25051293) (2025).
46. Kostas, D., Aroca-Ouellette, S. & Rudzicz, F. BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. *Front. Hum. Neurosci.* **15**, 653659, DOI: [10.3389/fnhum.2021.653659](https://doi.org/10.3389/fnhum.2021.653659) (2021).
47. Song, Y., Zheng, Q., Liu, B. & Gao, X. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 710–719, DOI: [10.1109/tnsre.2022.3230250](https://doi.org/10.1109/tnsre.2022.3230250) (2023).
48. Li, R. *et al.* MBDSTGAT: A Multibranch Dynamic Spatiotemporal Graph Attention Network With Topology Perturbation for Motor Imagery EEG Decoding. *IEEE Internet Things J.* **13**, 21811–21824, DOI: [10.1109/jiot.2026.3667051](https://doi.org/10.1109/jiot.2026.3667051) (2026).
49. Brunner, C., Leeb, R., Müller-Putz, G. R., Schlögl, A. & Pfurtscheller, G. BCI Competition 2008 – Graz Data Set A. Institute for Knowledge Discovery, Graz University of Technology http://www.bbc.de/competition/iv/desc_2a.pdf (2008).

50. Leeb, R., Brunner, C., Müller-Putz, G. R., Schlögl, A. & Pfurtscheller, G. BCI Competition 2008 – Graz Data Set B. Graz University of Technology, Austria http://www.bbc.de/competition/iv/desc_2b.pdf (2008).
51. Tangermann, M. *et al.* Review of the BCI Competition IV. *Front. Neurosci.* **6**, DOI: [10.3389/fnins.2012.00055](https://doi.org/10.3389/fnins.2012.00055) (2012).
52. Widmann, A., Schröger, E. & Maess, B. Digital filter design for electrophysiological data – a practical approach. *J. Neurosci. Methods* **250**, 34–46, DOI: [10.1016/j.jneumeth.2014.08.002](https://doi.org/10.1016/j.jneumeth.2014.08.002) (2015).
53. Guo, M. *et al.* MI-Mamba: A Hybrid Motor Imagery Electroencephalograph Classification Model with Mamba’s Global Scanning. *Annals New York Acad. Sci.* **1544**, 242–253, DOI: [10.1111/nyas.15288](https://doi.org/10.1111/nyas.15288) (2025).
54. Lotey, T., Keserwani, P., Dogra, D. P. & Roy, P. P. Feature Reweighting for EEG-Based Motor Imagery Classification. *Biomed. Signal Process. Control.* **110**, 108215, DOI: [10.1016/j.bspc.2025.108215](https://doi.org/10.1016/j.bspc.2025.108215) (2025).

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and constructive suggestions, which have significantly improved the quality of this manuscript.

Author contributions statement

Conceptualization, R.L. and T.Z.; Methodology, R.L.; Formal analysis, R.L.; Software development, R.L., T.Z. and D.S.; Validation, R.L., T.Z. and Z.L.; Investigation, Z.L., H.L. and J.H.; Resources, T.Z. and X.L.; Data curation, R.L. and Q.D.; Writing – original draft preparation, R.L. and T.Z.; Writing – review and editing, T.Z. and X.K.; Visualization, Q.D. and J.L.; Supervision, F.W., G.H. and X.K.; Project administration, F.W., G.H. and X.K.; Funding acquisition, F.W., G.H. and X.K.; All authors have read and agreed to the published version of the manuscript.

Additional information

Funding: This work was supported by the National Key Research and Development Program of China under Grant 2023YFF1204203, the Natural Science Foundation of Fujian Province of China under Grant 2026J008100, and the Special Project of the Fujian Provincial Department of Finance under Grant KLY24107XA.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.