

Multiscale Pooling Spatial–Temporal Attention Network: Elevating Cross Session and Small Sample Decoding in Motor Imagery Brain–Computer Interfaces

Weijie Chen^{ID}, Ian Daly^{ID}, Yixin Chen^{ID}, Xiao Wu^{ID}, Xinjie He^{ID}, Xingyu Wang^{ID},
Andrzej Cichocki^{ID}, *Life Fellow, IEEE*, and Jing Jin^{ID}, *Senior Member, IEEE*

Abstract—Motor imagery (MI) is one of the most widely used paradigms in brain–computer interfaces (BCIs), known for its ability to trigger changes in brain activity without the need for an external “cue” stimulus. This unique characteristic has attracted significant attention from neuroscientists and researchers in fundamental science. However, compared to P300 and steady-state visual evoked potential (SSVEP), neural activity related to MI tends to be less stable and exhibits substantial variability between individuals. Consequently, accurately decoding MI, using both traditional machine learning and deep learning, has proven to be a considerable challenge. Moreover, given the difficulty of acquiring electroencephalography (EEG) data and the high data demands of deep learning, enhancing the accuracy of MI decoding with limited sample sizes remains a pressing issue that urgently needs to be addressed. This article addresses the challenges mentioned above by introducing a novel deep neural network designed for accurate MI decoding, which is designed to be effective with both small-sample sizes and larger datasets. This

network, named the multiscale pooling spatial–temporal attention network (MPSTANet), integrates mix pooling techniques with spatial–temporal attention mechanisms. MPSTANet first employs local and global spatial attention, along with multiscale temporal attention, to thoroughly extract spatial–temporal information from EEG signals. Next, MPSTANet utilizes feature fusion and the proposed mix pooling technique to preserve as much of the extracted spatial–temporal information as possible. Finally, channel interaction attention (CIA) and 3-D weight attention (3-DWA) are employed to recalibrate the weights of the fused channels and spatial–temporal features, respectively. To validate the performance of our proposed MPSTANet model, we conducted experiments on four public datasets, including both small-sample sizes and subject-independent scenarios. MPSTANet achieved cross-session decoding accuracies of 84.82%, 72.92%, 88.20%, and 46.54% on the BCI Competition IV 2a dataset, the Open BMI dataset, the BCI Competition IV 2b dataset, and the PhysioNet dataset, respectively. Furthermore, MPSTANet demonstrated a significant lead compared to other deep learning models in both small-sample and subject-independent experiments. These results demonstrate the robustness of MPSTANet in MI decoding and its promising potential for BCI applications.

Index Terms—Brain–computer interfaces (BCIs), cross session, motor imagery (MI), small sample.

Received 25 February 2025; revised 27 June 2025 and 9 October 2025; accepted 26 December 2025. This work was supported in part by the Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project under Grant 2022ZD0208900; in part by the National Natural Science Foundation of China under Grant 62176090; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX; and in part by the Project of Jiangsu Province Science and Technology Plan Special Fund in 2022 (Key Research and Development Plan Industry Foresight, Fundamental Research Fund for the Central Universities JKH01241605 and Key Core Technologies) under Grant BE2022064-1. This article was recommended by Associate Editor X. You. (*Corresponding author: Jing Jin.*)

Weijie Chen, Yixin Chen, Xiao Wu, Xinjie He, and Xingyu Wang are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: wjchen827@foxmail.com; yixinchen1022@foxmail.com; wuxiao121409@163.com; xinjieHe@mail.ecust.edu.cn; xywang@ecust.edu.cn).

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

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

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

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSMC.2025.3650196>.

Digital Object Identifier 10.1109/TSMC.2025.3650196

I. INTRODUCTION

BRAIN–COMPUTER interfaces (BCIs) represent a cutting-edge technology that attempts to bridge the gap between thoughts and physical actions by providing an alternative efferent pathway from the brain to the external world. Currently, three principal electroencephalography (EEG)-based paradigms have been introduced in the field of BCI: steady-state visual evoked potentials (SSVEPs) [1], [2], [3], [4], event-related potentials (ERPs) [5], and motor imagery (MI) [6], [7], [8], [9], [10], [11]. Notably, SSVEP and ERP, due to the relative stability of the signals they elicit, have led to well-developed applications in areas such as spellers [12], [13], [14], [15], [16], smart homes [17], and driving wheelchairs [18]. The distinctive feature of MI is its ability to induce specific neural activity related to the imagination of subjects, without the need for external stimuli to cue subjects. Consequently, MI is particularly well-suited for medical rehabilitation applications, such as motor impairment rehabilitation [19], stroke recovery [20],

and the control of wheelchairs and prosthetics [21], [22]. However, the absence of requisite external stimuli and the inherently subjective nature of MI make the quality of the resulting EEG signals heavily dependent on the subject's imaginative capacity and focus. Consequently, factors such as high subjectivity, poor signal quality, and significant individual variability considerably undermine the decoding accuracy.

To attenuate the impact of these aforementioned limitations, a considerable number of advanced machine learning and decoding techniques have been developed to enhance the accuracy of decoding the motor intentions of subjects [23], [24], [25], [26], [27]. These machine learning algorithms typically integrate feature extraction methodologies with sophisticated feature classification techniques. The objective of feature extraction is to identify informative attributes from the raw EEG signals that encapsulate characteristics of brain activity related to cognitive events of interest. Feature classification, on the other hand, aims to select subsets of the extracted EEG features to differentiate distinct brain states and intentions. Prevailing feature classification methods include support vector machines, logistic regression, and random forests.

Prominent feature extraction methods encompass common spatial patterns (CSPs), power spectral density estimation, and fast Fourier transforms. For example, the CSP algorithm operates by calculating and decomposing the covariance matrices of EEG signals to identify a set of optimal spatial filters. These filters are designed to maximize the separability of different classes of EEG signals within the filtered feature space [25]. Consequently, a multitude of CSP-based algorithms have been proposed to further enhance this process. For example, filter bank CSPs (FBCSPs) were proposed to ameliorate the sensitivity of the CSP algorithm to varied frequency ranges [26]. The fundamental concept of FBCSP involves decomposing the EEG signal into multiple frequency bands, applying the CSP algorithm to each band for feature extraction, and subsequently amalgamating the features from all frequency bands for classification. Furthermore, Luo [27] significantly advanced the decoding accuracy of EEG signals by synergizing CSP with a genetic algorithm, yielding commendable results.

Nevertheless, machine learning algorithms typically necessitate meticulous manual feature engineering, which entails extracting salient features from raw data. This process is not only labor-intensive but also heavily contingent upon domain-specific expertise. In addition, the generalization capacity of conventional machine learning algorithms is often inadequate, especially when confronted with extensive datasets encompassing multiple subjects.

Conversely, deep learning facilitates end-to-end learning paradigms, enabling the autonomous extraction of intricate feature representations directly from raw data. Its exceptional flexibility and expressive capacity allow it to capture complex nonlinear relationships within the data, thereby allowing it to learn from large-scale datasets. Convolutional neural networks (CNNs), which are among the most favored deep neural network architectures, are particularly suited for processing data that has both spatial and temporal structural elements due to their inherent design and attributes.

Despite the inherently 1-D nature of EEG signals as time-series data, the spatial configuration of electrodes on the scalp imparts a distinctive spatial structure. Therefore, EEG signals comprise not only rich temporal features but also vital spatial characteristics. At present, CNNs have ascended to prominence and widespread adoption in the domain of EEG signal analysis, chiefly because they offer a highly effective mechanism for the automatic learning and classification of complex time-series data [28], [29], [30], [31].

For instance, EEGNet, which employs depthwise separable convolutions, was introduced as a CNN-based solution to EEG decoding [32]. This model has garnered widespread acclaim for achieving great classification accuracies with a relatively streamlined network architecture. Following this, to fully exploit the frequency band characteristics inherent in EEG signals, a novel neural network architecture named FBCNet was proposed [33]. FBCNet effectively integrates frequency band analysis to efficiently capture and process the time–frequency features embedded within EEG signals, thereby demonstrating superior classification precision. Similarly, filter-bank multiscale convolutional neural network (FBMSNet) not only incorporates multiband information but also introduces mix convolutions to further extract intricate details contained within EEG signals [34]. This approach has exhibited significantly enhanced classification accuracy and robust generalization capabilities when dealing with complex EEG data. Furthermore, Liu et al. [35] introduced an advanced spatial–temporal transformer grounded in domain generalization, employing the transformer architecture to adeptly capture intrinsic spatiotemporal characteristics of the brain. They also incorporated domain generalization strategies to elucidate cross-subject variability, thereby substantially enhancing the model's generalization capabilities. Yang and Jia [36] architected STMambaNet, a spatial–temporal Mamba network that capitalizes on the Mamba state-space framework to proficiently process extensible sequences with linear scalability, effectively modeling intricate spatial and temporal dynamics. To facilitate the seamless integration of multiscale information, attention-based dual-scale fusion convolutional neural network (ADFCNN) was proposed [47]. ADFCNN is an attention-empowered dual-scale fusion CNN that synergistically extracts and amalgamates EEG spectral and spatial features across diverse scales, employing self-attention mechanisms to optimally fuse multiresolution information.

Nevertheless, existing models predominantly concentrate on the extraction of temporal and spatial features from the EEG and frequently overlook the critical need for interaction among these extracted features, particularly in the context of sequential signals such as EEG. Furthermore, while each electrode channel functions independently, there is an intrinsic exchange of information between different channels, underscoring the necessity for decoding models to take into account the interactions amongst channel features.

Moreover, pooling operations, including max pooling and average pooling, stand as some of the most extensively employed strategies in deep learning. These techniques not only substantially diminish computational complexity but also partially alleviate the risk of overfitting. However, pooling

introduces certain limitations. It has a propensity to excessively simplify and abstract features, which inevitably leads to some degree of information loss. This loss can be particularly detrimental in applications where the preservation of detailed original data is imperative.

To mitigate the prevalent limitations of current models in capturing intricate feature interactions and the potential loss of critical features during pooling, while simultaneously enhancing model performance in small-sample scenarios, we present an innovative solution termed the multiscale pooling spatial-temporal attention network (MPSTANet) model. MPSTANet amalgamates mix pooling technology, multiscale temporal attention mechanisms, both local and global spatial attention techniques, channel feature fusion strategies, channel interaction methodologies, and 3-D weight attention (3-DWA) mechanisms. Contrary to conventional pooling methods, our proposed mix pooling technique utilizes three distinct pooling scales accompanied by a fixed pooling stride, and thus attempts to significantly reduce feature loss. The multiscale temporal attention mechanism integrates four different sizes of temporal convolutional kernels: smaller kernels excel at extracting high-frequency information from EEG signals, whereas larger kernels are proficient in capturing low-frequency information.

Existing deep learning methodologies predominantly emphasize global spatial features, often at the expense of local spatial characteristics. To rectify this issue, we incorporate both local and global spatial attention mechanisms. These mechanisms initially extract local features from EEG signals before subsequently capturing global spatial attributes. Upon comprehensively extracting spatial-temporal features, we engage in feature amalgamation. This process commences with the application of grouped temporal convolutions, immediately followed by a pointwise multiplication, thus ensuring the thorough integration of temporal and spatial features across each channel.

Furthermore, we introduce advanced channel interaction techniques designed to enhance the interchannel information exchange within our model, thereby averting the excessive isolation of information within individual channels. Additionally, we employ 3-DWA mechanisms to measure the significance of each position within the feature map by calculating the respective energy levels, allowing our model to assign appropriate weights to each neuron. Crucially, we integrate R-Drop and L_2 regularization alongside the cross-entropy loss function to effectively alleviate overfitting concerns.

The primary contributions of this article can be articulated as follows.

- 1) We introduce an end-to-end MPSTANet, which integrates pioneering mix pooling technology, sophisticated temporal and spatial attention mechanisms, advanced feature fusion techniques, channel interaction methodologies, and the innovative incorporation of 3-DWA mechanisms for the classification of MI tasks from EEG. This framework effectively addresses the deficiencies of existing approaches that excessively emphasize feature extraction while overlooking the critical aspect of feature interaction.
- 2) MPSTANet has been rigorously validated and evaluated on four public datasets. In comparison to other

state-of-the-art (SOTA) methodologies, MPSTANet attains superior classification accuracy and exhibits remarkable generalization capabilities.

- 3) We have undertaken comprehensive experimental investigations with MPSTANet, including cross-session, small-sample, and subject-independent scenarios. The findings reveal that MPSTANet not only excels in performance metrics universally but also demonstrates significant advantages over contemporary SOTA methods.

II. METHODS

In this section, we provide a comprehensive exposition of the structural design of MPSTANet and our implementation of regularization techniques within the network. We thoroughly explore the functionalities of each component and elucidate the impact of the regularization techniques in augmenting model generalization and mitigating overfitting.

A. Structure of MPSTANet

1) *Local and Global Spatial Attention Techniques:* In the analysis of the spatial characteristics of EEG signals, contemporary research methodologies predominantly utilize global feature extraction techniques. These approaches typically process signals recorded from all electrodes collectively, thus capturing the comprehensive patterns of brain activity. Nevertheless, a potential limitation of this technique is its propensity to overlook the local spatial features of signals between adjacent EEG electrodes. These local features may encompass significant information, such as the activity patterns of localized neural networks and subtle distinctions between different brain regions.

To extract a more detailed view of the spatial features of EEG signals, as illustrated in Fig. 1, we initially employ local spatial attention with a kernel size of $[L_S, 1]$ to extensively capture the interconnections between local EEG electrode signals. Subsequently, we apply global spatial attention with a kernel size of $[C, 1]$ to acquire the global spatial features across the entire recorded brain region.

2) *Multiscale Temporal Attention Technique:* The temporal information intrinsic to EEG signals is exceptionally rich, encompassing low-frequency, mid-frequency, and high-frequency components. However, conventional research methodologies predominantly rely on single-scale temporal feature extraction, which often fails to adequately capture both low-frequency and high-frequency information, thereby resulting in the omission of crucial data.

To mitigate this limitation, by utilizing four kernels of distinct sizes, namely, $[1, M_{T1}]$, $[1, M_{T2}]$, $[1, M_{T3}]$, and $[1, M_{T4}]$, MPSTANet effectively extracts the low-frequency, mid-frequency, and high-frequency information embedded in EEG signals, as illustrated in the multiscale temporal attention framework in Fig. 1. The architecture of the four branches is designed not only to focus on localized temporal features but also to capture a broader temporal context. This improves the model's capacity to adapt to complex temporal structures, thereby offering a more comprehensive understanding of EEG signal characteristics, thereby enabling MPSTANet to more accurately understand the dynamic variations in the EEG signals over time.

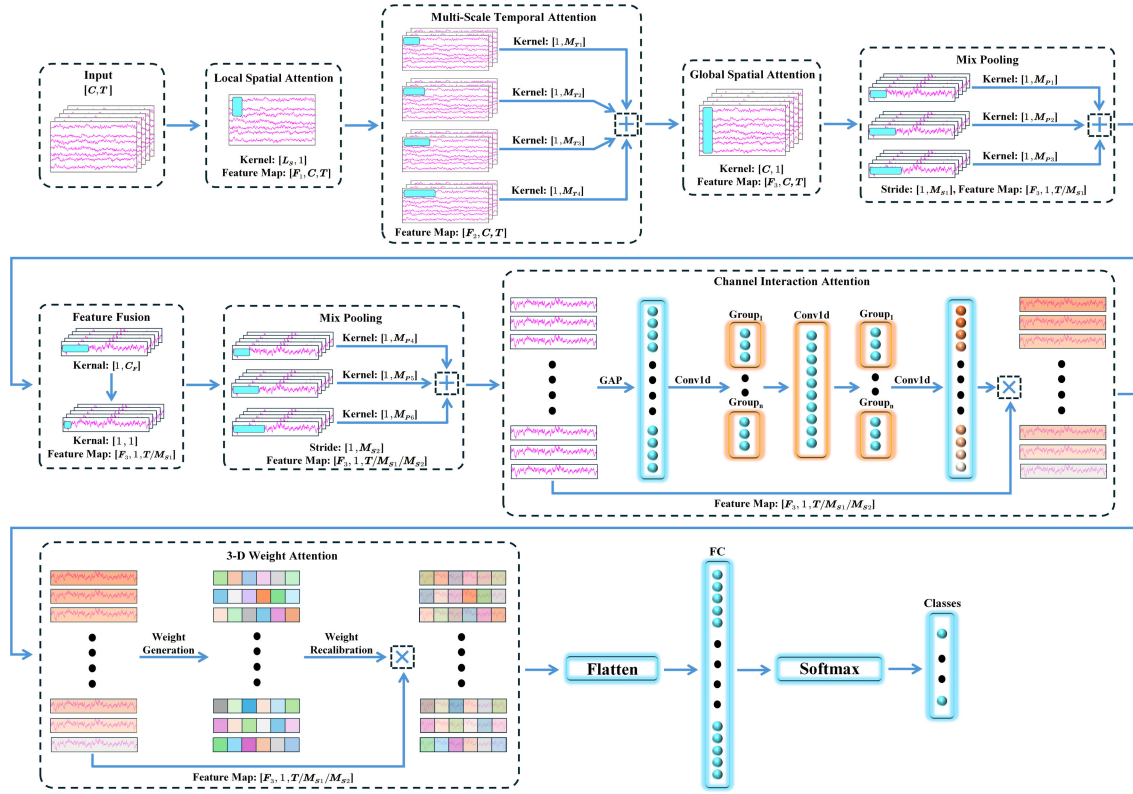


Fig. 1. Comprehensive architecture of MPSTANet; GAP stands for global average pooling and FC denotes a fully connected layer.

Moreover, the multiscale temporal attention technique augments the capacity of MPSTANet to learn more stable and robust features. These features exhibit enhanced performance even in the presence of noise interference and other sources of signal fluctuation, allowing the model to effectively adapt to variations across different temporal scales when confronted with session-to-session variability. This mitigates the influence of intersession discrepancies, thereby substantially augmenting the generalization capacity of MPSTANet.

3) *Mix Pooling Technique*: Max pooling and average pooling are two of the most ubiquitous pooling methods, facilitating the reduction of the spatial dimensions of features, diminishing the model's parametric complexity, and, to some extent, mitigating overfitting. However, these pooling mechanisms are not without limitations. Due to their fixed scales, they inherently lack flexibility, inevitably leading to the loss of salient features. Moreover, given the intersubject variability in the EEG, the optimal pooling scale may vary across different subjects.

To ameliorate the constraints associated with fixed pooling scales, we introduce a mix pooling technique (as shown in Fig. 1). Mix pooling amalgamates three pooling scales of disparate sizes, namely, $[1, M_{P1}]$, $[1, M_{P2}]$, and $[1, M_{P3}]$ (or $[1, M_{P4}]$, $[1, M_{P5}]$, and $[1, M_{P6}]$), in conjunction with a stride size of $[1, M_{S1}]$ (or $[1, M_{S2}]$). The mix pooling output is derived by aggregating the results of these three pooling operations. It is noteworthy that all parameters within mix pooling remain constant across each dataset.

Mix pooling captures features of EEG signals across a range of temporal scales. Smaller pooling scales are adept at capturing rapidly fluctuating features, whereas larger pooling

scales are more suited to capturing slowly varying features. This approach not only extracts the temporal dynamics of individual channels but also encapsulates the intricate temporal interactions across multiple channels. These interactions reflect distinctive patterns of functional activation or synchronized activities among different brain regions, thereby indirectly representing spatial dependencies within the EEG. By synthesizing features from these varied scales, a more holistic feature representation can be obtained, thereby reducing feature loss. Furthermore, mix pooling augments the robustness of MPSTANet, as pivotal features may manifest across different temporal scales, ensuring that these features are effectively captured and preserved.

4) *Feature Fusion Technique*: In the domain of EEG classification tasks, feature fusion is an indispensable step. The successful integration of features can profoundly enhance the capacity of a model to interpret and classify diverse EEG patterns. Following the comprehensive extraction of temporal and spatial features, we employ advanced feature fusion techniques to combine the extracted features, as depicted in Fig. 1. This technique encompasses grouped temporal convolutions with a kernel size of $[1, C_F]$ and a pointwise convolution. The merit of grouped temporal convolutions lies in their proficiency at capturing the intrinsic temporal characteristics within each channel, without intermingling information from other channels. Thus, the feature fusion technique attempts to achieve the meticulous extraction and integration of critical features, while significantly curtailing the model's parameter count and computational complexity, thereby preserving the model's lightweight architecture and operational efficiency.

5) *CIA Technique*: In the sophisticated analysis of EEG signals, numerous deep learning methodologies intended for EEG analysis predominantly emphasize feature extraction while inadequately addressing interchannel interactions. This limitation results in the overly isolated processing of information within individual channels, lacking the requisite integration and correlation. Consequently, this can adversely affect decoding accuracy and robustness, particularly in the context of complex tasks such as MI.

To effectively address this issue, we propose a novel channel interaction technique. This technique facilitates the allocation of weights according to the significance of each channel, enabling the extraction of interchannel relationships and the identification of task-relevant critical features, thereby adaptively refining the relative importance of each channel's attributes. As depicted in Fig. 1, initially, global average pooling (GAP) is employed to reduce the dimensionality of the EEG data, thereby computing the global spatial information for each channel. Subsequently, by setting a reduction ratio, grouped 1-D convolutions are utilized to downscale the channels, thus diminishing the parameter count and computational burden. Following this, a 1-D convolution is introduced to further refine the channel information, facilitating deeper exploration and adjustment of channel features. Next, grouped 1-D convolutions are employed to upscale the channels, restoring them to their original count, and a Sigmoid activation function is utilized to generate channel weights. Finally, each channel's weight is applied to the corresponding channel in the original feature map, thereby adaptively modulating the importance of each channel's features.

6) *3-DWA Technique*: In previous research, especially within EEG classification tasks, emphasis has largely been placed on channel and spatial attention mechanisms. These approaches, however, primarily concentrate on discerning and harnessing the significance of various channels and spatial positions within EEG signals. Although these methods have achieved certain levels of success, they inadvertently overlook a pivotal aspect of the EEG: the importance at the neuronal level within each spatial location.

To address this oversight, we have introduced the 3-DWA technique [38], as depicted in Fig. 1. EEG data inherently possess time-series characteristics, rendering temporal coherence critically significant. Diverging from conventional attention mechanisms, the 3-DWA technique effectively assesses the relevance of each temporal segment's latent state to the current classification task, mitigating the impact of irrelevant time intervals. This method emphasizes critical neural activity sequences along the temporal axis, recalibrates neuronal weights with precision, and augments the sensitivity to subtle intrinsic variations embedded within EEG signals.

First, we define the input feature as $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, and then formulate the ensuing energy function for each neuron

$$E_t(w_t, b_t, \mathbf{y}, x_i) = (y_t - \tilde{t})^2 + \frac{1}{B-1} \sum_{i=1}^{B-1} (y_o - \tilde{x}_i)^2 \quad (1)$$

where $\tilde{t} = w_t t + b_t$ and $\tilde{x}_i = w_t x_i + b_t$ represent the linear transformations of t and x_i , respectively, with t and

x_i signifying the target neuron and the remaining neurons within a singular channel of the input feature set. Concurrently, w_t and b_t are the variable weights and biases, respectively. Furthermore, $B = H \times W$ signifies the aggregate count of neurons in the specified channel. To ascertain the linear separability between the target neuron t and all other neurons within the same channel, it is necessary to minimize (1). For simplicity, we designate y_t and y_o as 1 and -1 , respectively. By integrating a regularization term ξw_t^2 into the equation, we derive the following equation:

$$\begin{aligned} E_t(w_t, b_t, \mathbf{y}, x_i) \\ = \frac{1}{B-1} \sum_{i=1}^{B-1} (-1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + \xi w_t^2. \end{aligned} \quad (2)$$

Equation (2) admits an analytical solution with respect to w_t and b_t , as illustrated by the following equations: (3) and (4), as shown at the bottom of the next page.

Assuming that all elements within a single channel adhere to the same distribution, the mean and variance can be computed over the entire set of neurons. Thus, the minimal energy can be ascertained using the following equation:

$$e_t = \frac{4 \left(\frac{1}{B} \sum_{i=1}^B \left(x_i - \frac{1}{B} \sum_{i=1}^B x_i \right)^2 + \xi \right)}{\left(t - \frac{1}{B} \sum_{i=1}^B x_i \right)^2 + 2 \frac{1}{B} \sum_{i=1}^B \left(x_i - \frac{1}{B} \sum_{i=1}^B x_i \right)^2 + 2\xi} \quad (5)$$

Equation (5) suggests that a lower value of energy e_t signifies a greater distinction for neuron t compared to others. Thus, the significance of each neuron can be denoted by $1/e_t$.

Finally, the weights of each neuron are reallocated using the sigmoid activation function, as shown in the following equation:

$$\mathbf{X}_R = \text{Sigmoid}(1/E) \odot \mathbf{X} \quad (6)$$

where $\mathbf{X}_R$ represents the final output, and $\mathbf{E}$ denotes the aggregation across all channels and spatial dimensions.

B. Implementation of Regularization Techniques

Dropout is among the most frequently utilized techniques to mitigate model overfitting. Dropout operates by randomly deactivating neurons in the model during the training phase to avert overfitting to the training data. In the testing phase, all neurons are fully activated for the purpose of prediction. This approach may lead to the model exhibiting robust performance on the training set yet falling short in terms of efficacy on the test set.

1) *Implementation of R-Drop*: To mitigate the aforementioned challenges, we introduced an advanced regularization technique termed R-Drop [39]. As depicted in Fig. 2, R-Drop mitigates the distributional discrepancies between submodel outputs derived from identical data samples by calculating and minimizing the bidirectional Kullback–Leibler (KL) divergence. In contrast to conventional dropout, which is confined to the hidden units of each layer, R-Drop extends its effect

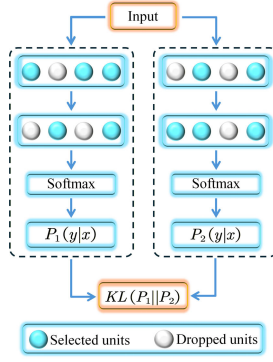


Fig. 2. Comprehensive structure of R-Drop.

to both the hidden units of each layer and the outputs of submodels generated through dropout sampling. This comprehensive approach reduces the performance differential between the training and test datasets and, in doing so, aims to bolster the model's robustness.

The computational approach for R-Drop comprises the following two steps: the initial step entails calculating the bidirectional KL divergence L_{KL}^i between the outputs of submodels generated from identical data samples

$$L_{\text{KL}}^i = \frac{1}{2} (\text{KL}(P_1(y_i|x_i) || P_2(y_i|x_i)) + \text{KL}(P_2(y_i|x_i) || P_1(y_i|x_i))). \quad (7)$$

The subsequent phase entails employing the cross-entropy loss function to quantify the divergence between the model's predicted probability distribution and the actual probability distribution, denoted as L_{CE}^i

$$L_{\text{CE}}^i = \frac{1}{2} (\text{CE}(y_i || P_1(y_i|x_i)) + \text{CE}(y_i || P_2(y_i|x_i))). \quad (8)$$

In this context, CE signifies the cross-entropy loss function, while the bidirectional KL divergence quantifies the discrepancy between the outputs of two submodels generated from identical data samples.

2) *Implementation of L_2 Regularization:* By integrating R-Drop as an ancillary loss component alongside the cross-entropy loss function, we aim to bolster the model's robustness. To further enhance the model's generalization, we also incorporated the L_2 regularization term. L_2 regularization, through the addition of a penalty term proportional to the sum of the squared weights within the loss function, effectively mitigates the risk of overfitting the model parameters

$$L_2\text{loss} = \varpi \sum_{\theta \in \Theta} \sum_i \theta_i^2 \quad (9)$$

where ϖ denotes the regularization strength parameter, Θ represents the ensemble of all model parameters, and θ_i signifies the i_h element within the parameter set.

By amalgamating the R-Drop technique, the cross-entropy loss function, and L_2 regularization, we derive the loss function L^i utilized for backpropagation in this study

$$\begin{aligned} L^i &= L_{\text{CE}}^i + \alpha L_{\text{KL}}^i + L_2\text{loss} \\ &= \frac{\alpha}{2} (\text{KL}(P_1(y_i|x_i) || P_2(y_i|x_i)) + \text{KL}(P_2(y_i|x_i) \\ &\quad \times || P_1(y_i|x_i))) \\ &\quad + \frac{1}{2} (\text{CE}(y_i || P_1(y_i|x_i)) + \text{CE}(y_i || P_2(y_i|x_i))) \\ &\quad + \varpi \sum_{\theta \in \Theta} \sum_i \theta_i^2. \end{aligned} \quad (10)$$

In this context, α signifies the coefficient that can modulate the intensity of the R-Drop regularization, while ϖ denotes the coefficient weight governing the intensity of L_2 regularization. And L^i represents the loss function that must be optimized throughout the ensuing training procedure.

In summary, regularization mitigates the excessive reliance on irrelevant signals by constraining the magnitude of the model weights, thereby addressing the issue of excessively large weights. It encourages MPSTANet to prioritize more prominent EEG channels and time intervals, rather than becoming overly dependent on specific channels or temporal points. This effectively prevents the model from learning overly intricate patterns, reducing the risk of overfitting, and consequently bolstering its robustness.

III. DATASETS AND EXPERIMENTAL CONFIGURATION

A. Datasets

1) *Dataset 1:* The BCI Competition IV 2a dataset is a four-class dataset encompassing left-hand, right-hand, foot, and tongue MI tasks. This dataset was recorded using the international 10–20 electrode placement system via 25 electrodes, which include 3 electrooculogram (EOG) channels and 22 EEG channels. It contains data from nine subjects, each of whom participated in two sessions conducted on different dates. Each session comprises 288 trials, with 72 trials per class, and was sampled at a rate of 250 Hz. The BCI Competition IV 2a dataset is available for download at <https://www.bbci.de/competition/iv/download>

2) *Dataset 2:* The OpenBMI dataset is a binary classification dataset encompassing left-hand and right-hand MI tasks. It was recorded using 62 EEG electrodes, positioned according to the International 10–20 system with a sampling

$$w_t = - \frac{2 \left(t - \frac{1}{B-1} \sum_{i=1}^{B-1} x_i \right)}{\left(t - \frac{1}{B-1} \sum_{i=1}^{B-1} x_i \right)^2 + 2 \left(\frac{1}{B} \sum_{i=1}^B \left(x_i - \frac{1}{B-1} \sum_{i=1}^{B-1} x_i \right)^2 \right) + 2\xi} \quad (3)$$

$$b_t = -\frac{1}{2} \left(t + \frac{1}{B-1} \sum_{i=1}^{B-1} x_i \right) w_t. \quad (4)$$

rate of 1000 Hz. The dataset features data from 54 subjects, including 29 males and 25 females, with 38 subjects being novice BCI users. Each subject engaged in one session per day on two separate occasions, with each session containing 200 trials. According to [34], we selected 20 channels that are judged to be the best placed to capture data from the motor areas of the brain: FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6. The Open BMI dataset is available for download at <https://gigadb.org/dataset/view/id/100542>

3) *Dataset 3*: The BCI Competition IV 2b dataset is a binary classification dataset designed for MI tasks involving the left and right hands. It comprises data from nine subjects, recorded using three EOG channels, and three EEG channels. Each subject’s data consists of five sessions, with the first two sessions conducted without visual feedback, and each session containing 120 trials (60 trials for each hand). The subsequent three sessions incorporate visual feedback, with each session comprising 160 trials (80 trials for each hand). All trials are recorded at a sampling rate of 250 Hz. The BCI Competition IV 2b dataset is available for download at <https://www.bbc.de/competition/iv/download>

4) *Dataset 4*: The PhysioNet dataset comprises data from 109 subjects, designed for a four-class MI classification task, encompassing left hand, right hand, both hands, and both feet [37]. Each subject engages in six sessions, comprising three sessions for the left hand and right hand, and three sessions for both hands and both feet. Recorded at a sampling rate of 160 Hz using 64 EEG channels positioned according to the international ten–ten system, the dataset provides a comprehensive representation of MI patterns. Data from four subjects (S088, S092, S100, and S104) were excluded due to inadequate sampling rates below the 160-Hz threshold. The PhysioNet dataset is available for download at <https://www.physionet.org/content/eegmmidb/1.0.0/>

B. Experimental Configuration

In the cross-session experiment, we used data from the first session for training and the subsequent session for testing for both Datasets 1 and 2. For Dataset 3, the first three sessions are utilized for training, while the subsequent two sessions for testing. For Dataset 4, the first four sessions are employed for training, with the final two sessions for testing. For the small-sample experiment, we initially select 20% of the data from the first session of each dataset for training purposes and subsequently test on the second session. Following this, we incrementally increase the proportion of training data, beginning with 20% of the first session’s data for training and assessing the model’s performance on the second session. This iterative process continues, progressively increasing the training data proportion until 90% of the first session’s data are utilized for training, with corresponding evaluations conducted on the second session. Prior to executing these experiments, we filter the data within the 0.5–40-Hz frequency range and extract a 4-s segment from each trial.

The batch size for both training and testing is configured to 32. The cross-entropy loss function is utilized in synergy with the R-Drop technique and L_2 regularization. Furthermore, the Adam optimizer is selected, with the learning rate fixed

TABLE I
DETAILS OF THE PARAMETER CONFIGURATION VALUES FOR MPSTANet

Module	Parameter setting
Local Spatial Attention	$L_S = 2, F_1 = 1$
Multi-Scale Temporal Attention	$M_{T1} = 5$
	$M_{T2} = 25$
	$M_{T3} = 45$
	$M_{T4} = 65$
Global Spatial Attention	$F_2 = 32$
	$C = \text{Channels}$
	$F_3 = 64$
Mix Pooling	$M_{P1} = M_{S1} = 4$
	$M_{P2} = 6$
	$M_{P3} = M_{P4} = M_{S2} = 8$
	$M_{P5} = 10$
Channel Interaction Attention	$M_{P6} = 12$
	Reduction ratio = 8
	Group = 8
R-Drop	Kernel = (1, 1)
	$\alpha = 0.5$
L_2 Regularization	$\varpi = 0.0005$

to 0.001, and training is conducted over 500 epochs, with hyperparameters α and λ set to 0.5 and 0.005, respectively, complemented by a cosine annealing learning rate schedule. The detailed parameter settings for MPSTANet are listed in Table I, where the number of channels is contingent upon the dataset. For the four datasets, the number of channels is configured as 22, 20, 3, and 64.

IV. EXPERIMENTAL RESULTS

This section describes the experimental outcomes of MPSTANet in conjunction with other SOTA methods across the four datasets. The results of all experiments are listed in Tables II–V, within these tables, the method that produces the highest mean accuracy for each subject is accentuated in bold. Furthermore, paired sample t-tests are utilized to evaluate the statistical significance of the results, with p -values indicated by asterisks (*: $P < 0.05$, **: $P < 0.01$, and ***: $P < 0.001$) in all tables.

A. Cross Session

The cross-session experimental results comparing the performance of MPSTANet to other SOTA methods on Dataset 1 are listed in Table II. It is evident that MPSTANet attained the highest classification accuracy, significantly outperforming recently proposed SOTA methods. In particular, compared to the classical machine learning algorithm FBCSP, MPSTANet exhibited a 17.07% improvement in accuracy. Relative to the widely used EEGNet, MPSTANet demonstrated a superior accuracy by 14.6%. Furthermore, in comparison to the contemporary SOTA methods ATCNet and M-FANet, MPSTANet achieved improvements in accuracy of 3.53% and 5.46%, respectively.

The cross-session experimental results of MPSTANet and other SOTA methods on Dataset 2 are listed in Table III. Over recent years, deep learning models have shown remarkable improvements in performance on Dataset 2, with several

TABLE II
COMPARISON OF CLASSIFICATION PERFORMANCES OF MPSTANet AND SOTA METHODS ON DATASET 1

Methods (year)	Subjects									Average Accuracy	STD
	A1	A2	A3	A4	A5	A6	A7	A8	A9		
FBCSP [26] (2017)	76.00	56.50	81.25	61.00	55.00	45.25	82.75	81.25	70.75	67.75***	13.73
EEGNet [32] (2018)	79.86	58.68	89.93	64.93	63.19	58.68	64.24	73.61	77.08	70.22***	10.72
EEG-TCNet [40] (2020)	85.77	65.02	94.51	64.91	75.36	61.40	87.36	83.76	78.03	77.35**	11.58
FBCNet [33] (2020)	85.42	60.42	90.63	76.39	74.31	53.82	84.38	79.51	80.90	76.20***	11.97
DRDA [41] (2021)	83.19	55.14	87.43	75.28	62.29	57.15	86.18	83.61	82.00	74.75***	12.96
Conformer [42] (2023)	87.59	61.71	92.94	77.63	58.27	64.33	87.36	84.92	87.92	78.07*	13.14
IDAN [43] (2023)	83.29	63.97	90.30	76.94	69.34	60.08	89.31	82.35	82.81	77.60***	10.85
ATCNet [44] (2023)	85.07	67.71	93.75	77.54	77.08	68.29	88.15	86.15	87.85	81.29***	9.14
M-FANet [45] (2024)	87.61	69.93	91.92	73.61	76.84	63.39	85.17	78.63	87.18	79.36***	9.39
MPSTANet	90.32	71.22	96.88	84.07	80.56	73.26	91.32	87.19	89.58	84.82	8.72

TABLE III
COMPARISON OF CLASSIFICATION PERFORMANCES OF MPSTANet AND SOTA METHODS ON DATASET 2

Methods (year)	Average Accuracy	STD
FBCSP [26] (2017)	60.36***	14.84
EEGNet [32] (2018)	60.41***	17.12
FBCNet [33] (2020)	67.19**	14.24
MIN2Net [46] (2022)	61.03***	14.47
FBMSNet [34] (2023)	70.05*	14.15
ADFCNN [47] (2024)	65.26**	-
TDLCNN [48] (2024)	70.43*	13.64
MPSTANet	72.92	13.21

models in the last two years achieving classification accuracies surpassing 70%. MPSTANet also yielded noteworthy classification results. Specifically, as shown in Table III, MPSTANet attained an accuracy of 72.92%, surpassing the machine learning algorithm FBCSP by approximately 12.56%. Furthermore, MPSTANet exhibited a 12.51% enhancement over EEGNet. Even when compared against recent SOTA methods such as FBMSNet, ADFCNN, and TDLCNN, MPSTANet still demonstrated superior advancements of 2.87%, 7.66%, and 2.49%, respectively.

The experimental outcomes of MPSTANet on Datasets 3 and 4 are presented in Tables IV and V, respectively. MPSTANet attained the highest classification accuracy on both datasets, exhibiting a notable superiority over the most recent SOTA methods. For instance, on Dataset 3, MPSTANet outperformed the contemporary SOTA method, M-FANet, by 5.4%. On Dataset 4, MPSTANet surpassed ADFCNN with an improvement of 5.19%.

B. Small-Sample Sizes

It is widely acknowledged that deep learning models generally require substantial datasets for training and that inadequately sized data can lead to overfitting on the training set and consequently poor performance on the test set. However, the acquisition of EEG signals for MI tasks is arduous, making it challenging to gather sufficient EEG data. Therefore, developing deep learning models for MI classification tasks that can be trained on small-sample training conditions is imperative. To address this issue, we conducted small-sample experiments

using MPSTANet and several other SOTA methods on the four datasets to evaluate each model's sensitivity to data volume.

The small-sample experimental results of MPSTANet and other SOTA methods on Dataset 1 are depicted in Fig. 3(a). It can be seen that, as the data volume increases, the classification accuracy of all the models enhances incrementally. Among these models, FBCNet and FBMSNet achieved high classification accuracies across all training data proportions. Notably, our proposed MPSTANet consistently outperformed other models across all dataset sizes.

The small-sample experimental results of MPSTANet and other SOTA methods that we applied to Dataset 2 are illustrated in Fig. 3(b). It is apparent that when the training data proportion is 20% or 30%, the classification accuracy across all models exhibited negligible variation. Additionally, when the training data proportion is 50% and 60%, the classification accuracy of the M-FANet and ShallowConvNet models remains relatively stable, while EEGNet and TCNet-Fusion exhibit a slight decline. Only EEG-TCNet, FBMSNet, and MPSTANet display an upward trend. As the data volume continues to increase, the classification accuracy of all the models progressively and consistently improves.

The small-sample experimental results of MPSTANet and other SOTA methods on Dataset 3 are depicted in Fig. 3(c). The accuracy of most models demonstrates a gradual improvement as the data volume increases, with the exception of EEGNet, ShallowConvNet, and TCNet-Fusion, which experience notable fluctuations when the data proportion ranges from 30% to 50%.

The small-sample experimental results of MPSTANet and other SOTA methods on Dataset 4 are illustrated in Fig. 3(d). It is apparent that the accuracy of the majority of models exhibits a consistent upward trend when the training data ratio falls between 20% and 60%, with the exception of EEGNet. However, once the training data ratio surpasses 60%, the accuracy of most models experiences a marked decline, whereas the accuracy of FBCNet continues to increase. It is noteworthy that MPSTANet maintains a leading position across all data ratios.

When the sample size is limited, fully extracting information from the available samples becomes particularly crucial. MPSTANet demonstrates significant advantages under small-sample conditions, primarily due to its integration of multiple

TABLE IV
COMPARISON OF CLASSIFICATION PERFORMANCES OF MPSTANet AND SOTA METHODS ON DATASET 3

Methods (year)	Subjects									Average Accuracy	STD
	A1	A2	A3	A4	A5	A6	A7	A8	A9		
FBCSP [26] (2017)	70.00	60.36	60.94	97.50	93.12	80.63	78.13	92.50	86.88	80.00**	13.85
EEGNet [32] (2018)	70.31	70.36	78.44	95.33	93.44	82.18	91.88	87.19	71.65	82.37*	10.15
EEG-TCNet [40] (2020)	69.69	69.07	77.22	96.14	91.13	74.57	84.96	91.25	74.38	81.21**	10.14
FBCNet [33] (2020)	67.38	61.21	67.19	90.06	93.75	72.31	80.56	88.44	71.56	76.94***	11.64
DRDA [41] (2021)	81.37	62.86	63.63	95.94	93.56	88.19	85.00	95.25	90.00	83.98	12.67
Conformer [42] (2023)	82.06	67.14	63.18	97.12	87.03	88.91	87.81	93.29	91.37	84.22	11.49
FBMSNet [34] (2023)	72.81	53.21	80.31	97.19	95.00	85.63	84.38	90.00	88.75	83.03*	13.39
M-FANet [45] (2024)	67.77	68.24	81.03	96.33	89.91	81.87	89.55	92.56	77.91	82.80*	10.25
MPSTANet	82.50	70.71	81.25	97.81	95.63	88.75	92.50	94.69	90.62	88.20	8.73

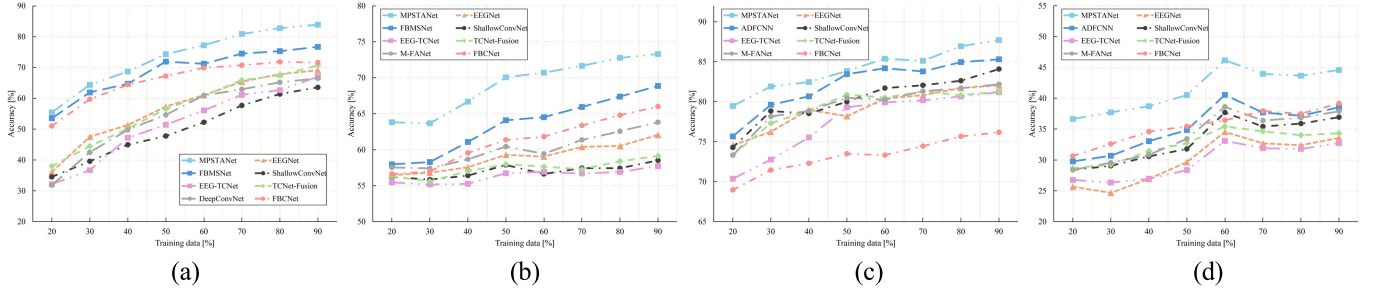


Fig. 3. Experimental results of MPSTANet and other SOTA methods on four datasets with small-sample sizes. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4.

TABLE V
COMPARISON OF CLASSIFICATION PERFORMANCES OF MPSTANet AND SOTA METHODS ON DATASET 4

Methods (year)	Average Accuracy	STD
EEGNet [32] (2018)	35.29**	13.15
FBCNet [33] (2020)	40.86*	14.79
EEG-TCNet [40] (2020)	34.63**	14.27
Conformer [42] (2023)	34.35**	13.13
ADFCNN [47] (2024)	41.35*	12.66
M-FANet [45] (2024)	39.87*	12.57
MPSTANet	46.54	11.67

TABLE VI
ABLATION EXPERIMENTS OF MPSTANet OVER THE FOUR DATASETS

Module			Average Accuracy			
LS	CIA	3-DWA	Dataset 1	Dataset 2	Dataset 3	Dataset 4
×	✓	✓	83.37	71.62	87.54	45.84
✓	×	✓	83.80	71.64	86.66	45.24
✓	✓	×	83.94	71.97	87.06	44.43
✓	✓	✓	84.82	72.92	88.20	46.54

multiple modules, MPSTANet achieves deep feature extraction and accurate classification of small-sample EEG data.

modules that not only efficiently extract effective information from EEG signals but also evaluate and rank the importance of the extracted features. Specifically, before extracting global channel information, MPSTANet employs local spatial attention to capture local spatial features, effectively complementing the representation of global spatial information. The combination of local and global spatial features enhances the richness and representativeness of spatial characteristics. Furthermore, the multiscale temporal convolution module comprehensively captures low-frequency, mid-frequency, and high-frequency components embedded in the signals, enabling multilevel temporal feature representation. To mitigate information loss during the dimensionality reduction process, the network introduces a mix pooling strategy that effectively preserves more valuable feature information. Finally, the channel interaction attention (CIA) and 3-DWA mechanisms jointly strengthen the expression of critical features, thereby improving the model's discriminative capability across different classes. In summary, through the collaborative optimization of

V. DISCUSSION

A. Ablation Study

The proposed MPSTANet model integrates various attention mechanisms. To ascertain the efficacy of these mechanisms, we performed an ablation study. The experimental outcomes of this study are listed in Table VI, where LS signifies local spatial attention, CIA stands for channel interaction attention, and 3-DWA denotes 3-D weight attention. The symbol “x” indicates the exclusion of the respective module.

Traditionally, the spatial feature extraction of EEG signals employs a global convolutional kernel, focusing on the overall characteristics of the EEG signals. However, this method may neglect the local spatial information within the EEG. To address this limitation, we introduced the local spatial attention (LS) technique. The results from the ablation study show the effect of this technique on the models' performance. Specifically, upon the removal of the LS module, the mean performance of MPSTANet deteriorated by 2.01%, 1.3%, 0.66%, and 0.7% on the four datasets, respectively.

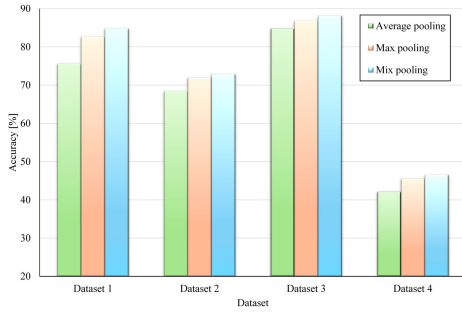


Fig. 4. Experimental results of MPSTANet employing average pooling, max pooling, and mix pooling over the four datasets.

In deep learning models used for EEG decoding, channels are typically considered independently, thereby overlooking interchannel information exchange. Consequently, we incorporated the CIA technique to facilitate the correlation and integration of information between channels, thereby enhancing the accuracy of MI classification tasks. When this CIA module was excised, the performance of MPSTANet declined by 1.02%, 1.28%, 1.54%, and 1.3% on the four datasets, respectively.

Channel attention techniques have garnered significant traction in MI classification tasks, yielding commendable results. However, they often disregard the significance of individual neurons. Therefore, we introduced the 3-DWA technique, which recalibrates the weights of each neuron within the channels. The ablation study results validate the effectiveness of this weight recalibration approach. Eliminating the 3-DWA module resulted in a performance decrease of 0.88%, 0.95%, 1.14%, and 2.11% on the four datasets, respectively.

B. Efficacy of Mix Pooling

This study introduces a mix pooling technique designed to mitigate the feature loss associated with fixed pooling strides. By employing pooling at varied scales, mix pooling endeavors to retain a maximal number of extracted EEG features, thereby ensuring that crucial features are not lost during the pooling process. To substantiate the efficacy of the mix pooling technique, we conducted experiments by substituting it with the conventional average pooling and max pooling methods on the four datasets.

The experimental results are depicted in Fig. 4. When MPSTANet utilized average pooling, its performance experienced a substantial decline across the four datasets. In contrast, substituting average pooling with max pooling resulted in a marked enhancement in the performance of MPSTANet on the four datasets. Nevertheless, when the proposed mix pooling was implemented, MPSTANet exhibited further improvement in performance. Consequently, we can conclude that the effectiveness of mix pooling is corroborated.

C. Efficacy of Multiscale Temporal Attention

The multiscale temporal attention technique adeptly captures both low-frequency and high-frequency information in EEG signals, thereby obtaining a more comprehensive

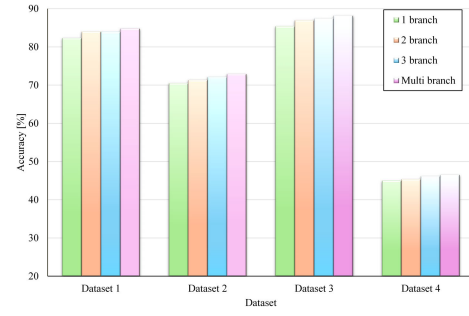


Fig. 5. Experimental results of MPSTANet with temporal attention using different numbers of branches on four datasets.

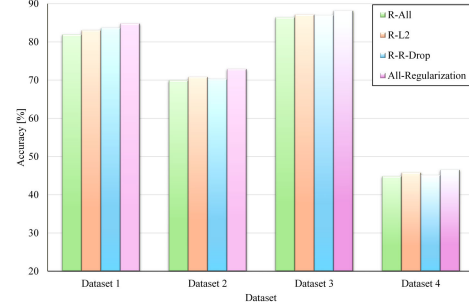


Fig. 6. Experimental results of MPSTANet utilizing diverse regularization techniques over the four datasets.

feature representation. Consequently, we conducted experiments on the number of multiscale temporal attention branches using four datasets, with the results presented in Fig. 5. The experimental outcomes reveal a trend where the classification accuracy of MPSTANet consistently improves with the increase in the number of branches, with the optimal performance achieved through the use of multiple branches. Specifically, with a single branch, the performance of MPSTANet on the four datasets shows a marked decline relative to multiple branches. The implementation of two branches results in an enhancement in performance on the four datasets, albeit still inferior to the multiple branches configuration. Additionally, when the number of branches is increased to three, the performance of MPSTANet on Dataset 1 experiences a slight decline compared to two branches, while the performance on the other three datasets with three branches surpasses that of the two-branch configuration.

D. Efficacy of Regularization

For the classification task of MI, the limited amount of EEG data and the significant variability in EEG signals from the same subject over different times poses a persistent challenge to enhancing the generalization capability of MI classification models. To substantiate the efficacy of the regularization techniques proposed in this article, we conducted evaluations using R-Drop and L_2 regularization, with the results illustrated in Fig. 6.

In Fig. 6, “R-All” denotes the removal of both R-Drop and L_2 regularization, relying solely on the cross-entropy loss function. “R-L2” signifies the exclusion of only L_2 regularization, while “R-R-Drop” pertains to the exclusion of only R-Drop. “All-Regularization” represents the application

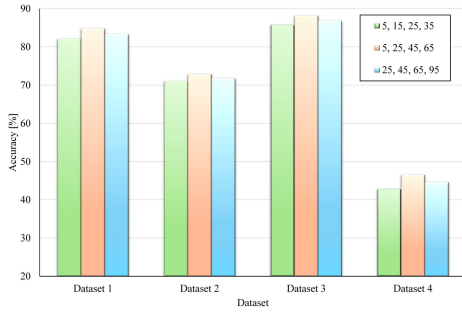


Fig. 7. Experimental results of MPSTANet using multiscale convolution kernels of different sizes on four datasets.

of both R-Drop and L_2 regularization in conjunction with the cross-entropy loss.

As depicted in Fig. 6, the omission of either R-Drop or L_2 regularization precipitates a marked decline in the performance of MPSTANet across the four datasets. This empirical evidence compellingly attests to the efficacy of regularization techniques in augmenting model performance. When both R-Drop and L_2 regularization are concurrently applied, the model's performance on both datasets surpasses that of the model when it is applied without any regularization techniques. This underscores that regularization techniques not only mitigate the risk of overfitting but also enhance the model's generalization ability, thereby rendering it more robust and dependable in addressing diverse datasets and tasks.

E. Influence of Convolution Kernels of Different Sizes

EEG signals encompass a wealth of information across low-, mid-, and high-frequency bands, making the selection of appropriately sized convolution kernels that effectively capture these spectral components crucial in multiscale convolutional frameworks. To elucidate the influence of varying convolution kernel sizes on the performance of MPSTANet, we conducted cross-session experiments across four datasets. The results, depicted in Fig. 7, reveal a pronounced degradation in MPSTANet's performance when employing kernel size combinations of either 5, 15, 25, and 35 or 25, 45, 65, and 95. This decline can be attributed to the fact that the former configuration, while accommodating low- and mid-frequency information, neglects high-frequency components, whereas the latter disregards the low-frequency constituents intrinsic to EEG signals. Conversely, the kernel set selected by MPSTANet, consisting of 5, 25, 45, and 65, facilitates a comprehensive extraction of low-, mid-, and high-frequency features, thereby achieving the highest decoding accuracy.

F. Subject-Independent Experiments

The generalizability and robustness of MI decoding models remain critical areas in need of enhancement. Consequently, subject-independent experiments serve as a pivotal metric for assessing the performance of MI decoding models. We conducted subject-independent evaluations of MPSTANet and SOTA approaches across four datasets, with the results presented in Table VII. MPSTANet consistently achieves the highest classification accuracy among the compared SOTA

TABLE VII
SUBJECT-INDEPENDENT EXPERIMENTS OF MPSTANet AND SOTA METHODS OVER THE FOUR DATASETS

Methods (year)	Average Accuracy			
	Dataset 1	Dataset 2	Dataset 3	Dataset 4
EEGNet [32] (2018)	55.32**	72.48**	71.07**	50.52*
FBCNet [33] (2020)	52.24***	70.21***	67.68***	46.31**
EEG-TCNet [40] (2020)	54.17**	73.51**	72.02**	45.50**
Conformer [42] (2023)	51.41***	71.94**	68.84**	46.86**
ADFCNN [47] (2024)	60.55*	75.19**	72.83*	48.67*
M-FANet [45] (2024)	58.81**	73.04**	70.49**	47.91**
MPSTANet	63.18	79.26	76.56	53.25

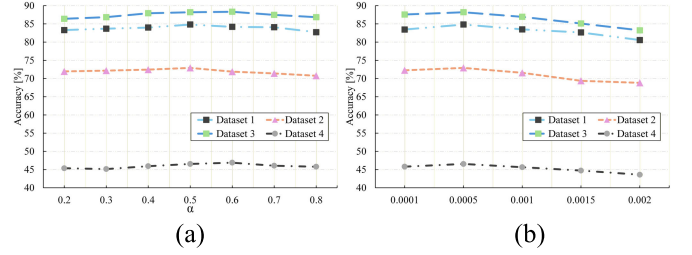


Fig. 8. Influence of different hyperparameter values in the loss function on the performance of MPSTANet across four datasets. (a) Influence of α . (b) Influence of τ .

methods across all four datasets. Specifically, when compared to the classical EEGNet model, MPSTANet demonstrates an advantage of 7.86%, 6.78%, 5.49%, and 2.73% over the four datasets. Even in comparison to the recently introduced ADFCNN, MPSTANet outperforms by margins of 4.37%, 4.07%, 3.73%, and 4.58%, respectively.

G. Influence of Hyperparameters in the Loss Function

To further investigate the influence of hyperparameters within the regularization framework, we conducted a comprehensive sensitivity analysis of MPSTANet across four datasets. The experimental results are depicted in Fig. 8. As illustrated in Fig. 8(a), the performance of MPSTANet exhibits a positive correlation with gradually increases in hyperparameter α . When α attains a value of 0.5, MPSTANet achieves optimal performance on Datasets 1 and 2. When α increases to 0.6, MPSTANet attains the best results on Datasets 3 and 4. However, surpassing the threshold of 0.7 precipitates a marked deterioration in performance across all four datasets. These observations suggest that MPSTANet maintains robust efficacy at α values of either 0.5 or 0.6, with our selection of $\alpha = 0.5$ justified by a notable performance decline on Dataset 2 at $\alpha = 0.6$.

Fig. 8(b) illustrates that increasing the hyperparameter τ from 0.0001 to 0.0005 exhibits an upward trajectory in MPSTANet's performance across all datasets. Nevertheless, exceeding the τ value of 0.0005 induces a progressive degradation in performance, attributable to excessive regularization constraining the model's representational capacity and impeding its ability to capture salient information within the data. Consequently, the hyperparameter configuration of $\alpha = 0.5$ and $\tau = 0.0005$ was adopted as optimal.

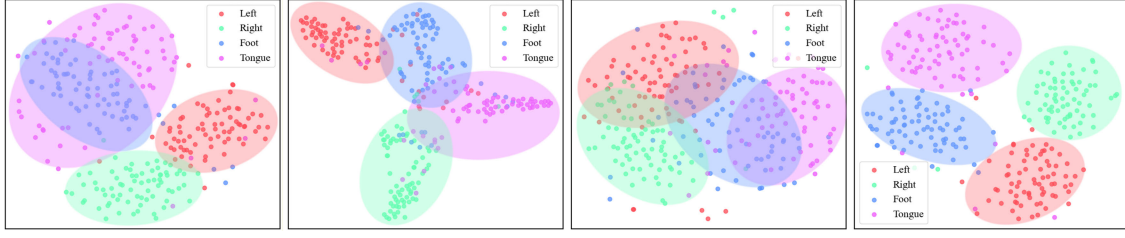


Fig. 9. t-SNE visualizations of EEGNet, EEGTCNet, M-FANet, and MPSTANet on Dataset 1 for subject 3.

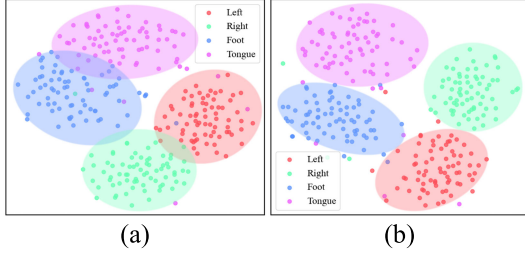


Fig. 10. Feature distribution of MPSTANet for subject 3 on Dataset 1 based on t-SNE visualization. (a) Without L_2 regularization and R-Drop and (b) with all regularization techniques.

H. Feature Visualization

To further evaluate the efficacy of MPSTANet in MI classification tasks, we employed the t-distributed stochastic neighbor embedding (t-SNE) method to visualize the output features preceding the final layer of EEGNet, EEG-TCNet, M-FANet, and MPSTANet. The t-SNE visualization for subject 3 in Dataset 1 is depicted in Fig. 9. It is evident that EEGNet, EEG-TCNet, and M-FANet lack distinct boundaries between different classes, with a significant overlap among data points from various categories. Conversely, MPSTANet exhibits superior classification performance relative to EEGNet, EEG-TCNet, and M-FANet, as evidenced by the increased interclass distances and reduced intraclass distances, thereby achieving more pronounced separation of the four classes in Dataset 1.

Inspired by [29], we visualized the feature distribution of MPSTANet both prior to and following the application of regularization techniques, as illustrated in Fig. 10. This visualization reveals that after incorporating L_2 regularization and R-Drop, the boundaries between distinct classes are markedly more pronounced, with greater interclass separation. This enhancement is conducive to optimizing the performance of MPSTANet.

VI. CONCLUSION

This article introduces MPSTANet, an innovative deep learning network that amalgamates multiple attention mechanisms for the classification of MI tasks in BCI. MPSTANet leverages both local and global attention techniques, thereby effectively integrating detailed local spatial information with comprehensive global spatial data derived from EEG signals. It also employs multiscale temporal attention mechanisms to exhaustively extract both low- and high-frequency temporal information from EEG signals, thus preventing the loss of critical temporal features. Moreover, we propose a mix

pooling technique that combines pooling with varying strides, which effectively reduces data dimensionality while retaining the extracted temporal and spatial features to the greatest extent possible. To further enhance interchannel information exchange within MPSTANet, we introduce a CIA mechanism to consolidate information across channels. Lastly, we implement a 3-DWA technique to recalibrate the weights of neurons within each channel. The efficacy of MPSTANet was assessed using two datasets, with the experimental results demonstrating that MPSTANet surpasses other SOTA methods. In summary, MPSTANet markedly improves the accuracy and robustness of MI classification through the innovative fusion of multiple attention mechanisms. Consequently, it holds significant potential for advancing the field of BCIs and fostering the development and application of related technologies. In future work, our objective is not only to focus on enhancing the performance of MI classification, but also to prioritize the optimization of the model's computational complexity. We aim to strike a balance between boosting performance and minimizing the computational load, with the ultimate goal of achieving a more efficient and lightweight model.

APPENDIX

A. Spatial Filters

The quantity of the spatial filters often exerts a substantial influence on the performance of neural networks. In this context, we investigated the performance of MPSTANet across four datasets with varying numbers of spatial filters, as detailed in Table VIII. The results reveal that MPSTANet performs optimally with 32 and 64 filters, surpassing other configurations, with the sole exception of Dataset 4. These findings underscore that both an excessively high and excessively low number of spatial filters can adversely affect the performance of MPSTANet, emphasizing the critical importance of selecting an optimal number of filters.

B. Analysis of Epoch Impact

The loss versus epoch curves serve as an essential tool for assessing whether a model is experiencing overfitting or underfitting. To evaluate the fitting capacity of MPSTANet, we plotted the training and testing loss curves across the four datasets, as illustrated in Fig. 11. The results reveal that MPSTANet's training loss consistently diminishes across all four datasets, while the testing loss progressively stabilizes as the number of epochs increases. This indicates that MPSTANet exhibits remarkable proficiency in mitigating overfitting.

TABLE VIII
EXPERIMENTS ON THE INFLUENCE OF SPATIAL FILTERS OF MPSTANet
OVER THE FOUR DATASETS

Datasets	Spatial filters	Average Accuracy	
		Cross Session	Subject Independent
Dataset 1	$F_2=8, F_3=16$	82.62	61.27
	$F_2=16, F_3=32$	83.39	62.06
	$F_2=32, F_3=64$	84.82	63.18
	$F_2=64, F_3=128$	83.02	62.87
Dataset 2	$F_2=8, F_3=16$	71.43	78.14
	$F_2=16, F_3=32$	72.44	78.60
	$F_2=32, F_3=64$	72.92	79.26
	$F_2=64, F_3=128$	72.13	78.44
Dataset 3	$F_2=8, F_3=16$	87.13	74.79
	$F_2=16, F_3=32$	87.71	75.77
	$F_2=32, F_3=64$	88.20	76.56
	$F_2=64, F_3=128$	87.24	75.51
Dataset 4	$F_2=8, F_3=16$	45.78	52.36
	$F_2=16, F_3=32$	46.02	52.74
	$F_2=32, F_3=64$	46.54	53.25
	$F_2=64, F_3=128$	46.95	52.57

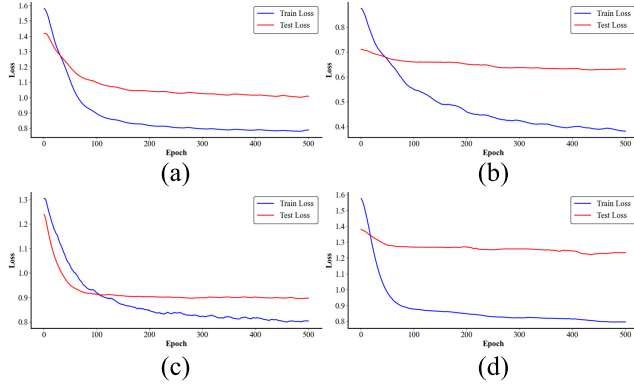


Fig. 11. Fluctuations in training and testing losses of MPSTANet across four datasets under different epochs. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4.

C. Dynamic Impact Analysis of Different Loss Functions

To explore the impact of various loss functions on the performance of MPSTANet, we systematically removed L_2 regularization, R-Drop, and both regularization techniques simultaneously across four datasets, and subsequently plotted the variations in both training and testing losses. The results are depicted in Fig. 12. As shown, even with the removal of L_2 regularization, R-Drop, or both techniques, the training losses across all four datasets exhibited a consistent decline and eventual stabilization. In Datasets 1 and 3, the pattern of testing loss closely mirrored that of the training loss. In Dataset 2, although the reduction in testing loss occurred at a slower pace, it ultimately stabilized at a relatively low level. In Dataset 4, despite the higher testing loss values, an initial downward trend was observed before the loss stabilized.

D. Computational Complexity

Although MPSTANet demonstrates impressive performance across the four datasets, its computational complexity is a significant consideration when evaluating the overall performance of the model. FLOPS, a widely adopted metric indicating the number of floating-point operations executed per

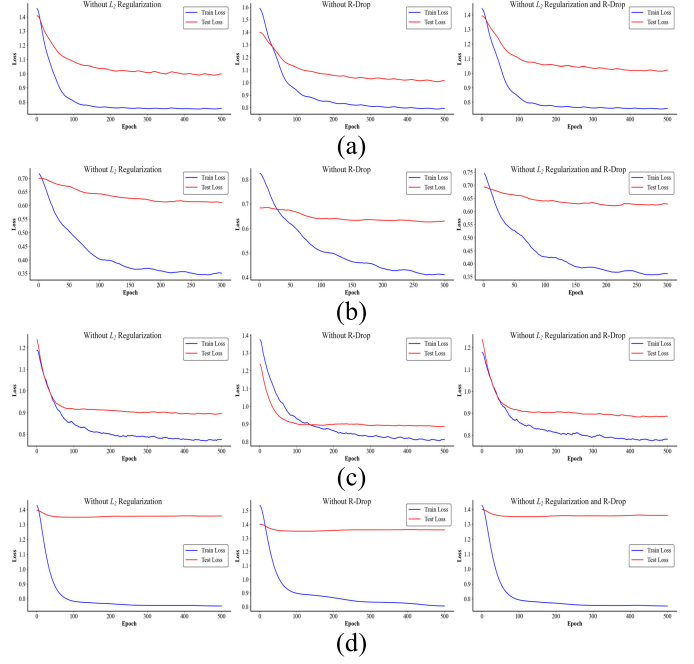


Fig. 12. Variations in training and testing losses of MPSTANet following the removal of L_2 regularization, R-Drop, and both regularization techniques across the four datasets. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4.

TABLE IX
DETAILS OF COMPUTATIONAL COMPLEXITY

Methods	Parameters (K)	FLOPS (G)	Time (s)	Memory (MB)
EEGNet	3.46	0.0122	0.2528	1564
EEG-TCNet	15.26	0.0067	0.3436	1593
Conformer	789.81	0.0639	0.4718	1686
ATCNet	113.71	0.0584	0.4616	1643
MPSTANet	20.89	0.1003	0.5884	2254

second, serves as a key standard for evaluating computational performance. Consequently, we performed a comprehensive comparative analysis of MPSTANet against several SOTA approaches in terms of parameter count, FLOPS, epochwise runtime, and memory consumption using Dataset 1, as detailed in Table IX. While MPSTANet surpasses EEGNet and EEG-TCNet in terms of parameters, it remains substantially more efficient than Conformer and ATCNet. Furthermore, MPSTANet exhibits relatively high FLOPS, exceeding those of the other four SOTA models, which corresponds to the longest epoch runtime and increased memory utilization, an area that necessitates optimization.

REFERENCES

- [1] X. He et al., "Leveraging transfer superposition theory for stable-state visual evoked potential cross-subject frequency recognition," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 11, pp. 3071–3084, Nov. 2024.
- [2] Z. Gao, W. Dang, M. Liu, W. Guo, K. Ma, and G. Chen, "Classification of EEG signals on VEP-based BCI systems with broad learning," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 11, pp. 7143–7151, Nov. 2021.
- [3] J. Ai, J. Meng, X. Mai, and X. Zhu, "BCI control of a robotic arm based on SSVEP with moving stimuli for reach and grasp tasks," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 3818–3829, Aug. 2023.

- [4] P. Autthasan et al., "A single-channel consumer-grade EEG device for brain-computer interface: Enhancing detection of SSVEP and its amplitude modulation," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3366–3378, Mar. 2020.
- [5] J. Jin, R. Xu, I. Daly, X. Zhao, X. Wang, and A. Cichocki, "MOCNN: A multiscale deep convolutional neural network for ERP-based brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 54, no. 9, pp. 5565–5576, Sep. 2024.
- [6] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.
- [7] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Dynamic convolution with multilevel attention for EEG-based motor imagery decoding," *IEEE Internet Things J.*, vol. 10, no. 21, pp. 18579–18588, Nov. 2023.
- [8] L. Zhou, X. Tao, F. He, P. Zhou, and H. Qi, "Reducing false triggering caused by irrelevant mental activities in brain-computer interface based on motor imagery," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3638–3648, Sep. 2021.
- [9] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.
- [10] S. U. Amin, H. Altaheri, G. Muhammad, W. Abdul, and M. Alsulaiman, "Attention-inception and long-short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5412–5421, Aug. 2022.
- [11] Z. Wang et al., "Motor imagery and action observation induced electroencephalographic activations to guide subject-specific training paradigm: A pilot study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2457–2467, 2023.
- [12] Y. Ke, J. Du, S. Liu, and D. Ming, "Enhancing detection of control state for high-speed asynchronous SSVEP-BCIs using frequency-specific framework," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1405–1417, 2023.
- [13] E. Yin, T. Zeyl, R. Saab, T. Chau, D. Hu, and Z. Zhou, "A hybrid brain-computer interface based on the fusion of P300 and SSVEP scores," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 693–701, Jul. 2015.
- [14] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A speedy hybrid BCI spelling approach combining P300 and SSVEP," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 473–483, Feb. 2014.
- [15] E. Yin, Z. Zhou, J. Jiang, Y. Yu, and D. Hu, "A dynamically optimized SSVEP brain-computer interface (BCI) speller," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1447–1456, Jun. 2015.
- [16] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm," *J. Neural Eng.*, vol. 10, no. 2, Apr. 2013, Art. no. 026012.
- [17] Y. Zhang, S. Q. Xie, H. Wang, and Z. Zhang, "Data analytics in steady-state visual evoked potential-based brain-computer interface: A review," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1124–1138, Jan. 2021.
- [18] S. Zhang, Y. Chen, L. Zhang, X. Gao, and X. Chen, "Study on robot grasping system of SSVEP-BCI based on augmented reality stimulus," *Tsinghua Sci. Technol.*, vol. 28, no. 2, pp. 322–329, Apr. 2023.
- [19] R. Foong et al., "Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 786–795, Mar. 2020.
- [20] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, Apr. 2017.
- [21] X. Mai, J. Ai, M. Ji, X. Zhu, and J. Meng, "A hybrid BCI combining SSVEP and EOG and its application for continuous wheelchair control," *Biomed. Signal Process. Control*, vol. 88, Feb. 2024, Art. no. 105530.
- [22] M. Jochumsen, T. A. M. Janjua, J. C. Arceo, J. Lauber, E. S. Buessinger, and R. L. Käseler, "Induction of neural plasticity using a low-cost open source brain-computer interface and a 3D-printed wrist exoskeleton," *Sensors*, vol. 21, no. 2, p. 572, Jan. 2021.
- [23] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
- [24] D.-H. Shin et al., "MARS: Multiagent reinforcement learning for spatial-spectral and temporal feature selection in EEG-based BCI," *IEEE Trans. Syst. Man Cybern., Syst.*, vol. 54, no. 5, pp. 3084–3096, Feb. 2024.
- [25] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
- [26] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.
- [27] T.-J. Luo, "Parallel genetic algorithm based common spatial patterns selection on time-frequency decomposed EEG signals for motor imagery brain-computer interface," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104397.
- [28] W. Dang et al., "Flashlight-net: A modular convolutional neural network for motor imagery EEG classification," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 54, no. 7, pp. 4507–4516, Jul. 2024.
- [29] M. Zhang et al., "Multimodal vigilance estimation with modality-pairwise contrastive loss," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 4, pp. 1139–1150, Apr. 2024.
- [30] P. Chen, Z. Gao, M. Yin, J. Wu, K. Ma, and C. Grebogi, "Multiattention adaptation network for motor imagery recognition," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 8, pp. 5127–5139, Aug. 2022.
- [31] P. Autthasan, R. Chaisaen, H. Phan, M. D. Vos, and T. Wilaiprasitporn, "MixNet: Joining force of classical and modern approaches toward the comprehensive pipeline in motor imagery EEG classification," *IEEE Internet Things J.*, vol. 11, no. 17, pp. 28539–28554, Sep. 2024.
- [32] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [33] R. Mane, N. Robinson, A. P. Vinod, S.-W. Lee, and C. Guan, "A multi-view CNN with novel variance layer for motor imagery brain computer interface," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 2950–2953.
- [34] K. Liu, M. Yang, Z. Yu, G. Wang, and W. Wu, "FBMSNet: A filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 2, pp. 436–445, Feb. 2023.
- [35] S. Liu, L. An, C. Zhang, and Z. Jia, "A spatial-temporal transformer based on domain generalization for motor imagery classification," in *Proc. IEEE Int. Conf. Syst. Man, Cybern. (SMC)*, vol. 30, Oahu, HI, USA, Oct. 2023, pp. 3789–3794.
- [36] X. Yang and Z. Jia, "Spatial-temporal mamba network for EEG-based motor imagery classification," in *Advanced Data Mining and Applications (Lecture Notes in Computer Science)*, vol. 15389, Cham, Switzerland: Springer, 2024.
- [37] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
- [38] L. Yang, R. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn. (PMLR)*, 2021, pp. 11863–11874.
- [39] X. Liang et al., "R-drop: Regularized dropout for neural networks," in *Proc. Neural Inf. Process. Syst.*, 2021, pp. 10890–10905.
- [40] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Toronto, ON, Canada, Oct. 2020, pp. 2958–2965.
- [41] H. Zhao, Q. Zheng, K. Ma, H. Li, and Y. Zheng, "Deep representation-based domain adaptation for nonstationary EEG classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 535–545, Feb. 2021.
- [42] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2023.
- [43] Q. She, T. Chen, F. Fang, J. Zhang, Y. Gao, and Y. Zhang, "Improved domain adaptation network based on Wasserstein distance for motor imagery EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1137–1148, 2023.
- [44] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 2249–2258, Feb. 2023.
- [45] Y. Qin, B. Yang, S. Ke, P. Liu, F. Rong, and X. Xia, "M-FANet: Multi-feature attention convolutional neural network for motor imagery decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 401–411, 2024.

- [46] P. Autthasan et al., “MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2105–2118, Jun. 2022.
- [47] T. Wei et al., “ADFCNN: Attention-based dual-scale fusion convolutional neural network for motor imagery brain–computer interface,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 154–165, 2023.
- [48] X. Ma, W. Chen, Z. Pei, J. Liu, B. Huang, and J. Chen, “A temporal dependency learning CNN with attention mechanism for MI-EEG decoding,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3188–3200, 2023.



Weijie Chen received the B.S. degree in electronic information engineering from Anhui Polytechnic University, Wuhu, China, in 2020, and the M.S. degree in electronic information from the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China, in 2023. He is currently pursuing the Ph.D. degree in control science and engineering with East China University of Science and Technology, Shanghai, China.

His main research interests include brain–computer interfaces, neural networks, and robotics.



Ian Daly received the M.Eng. degree in computer science and the Ph.D. degree in cybernetics from the University of Reading, Reading, U.K., in 2006 and 2011, respectively.

From May 2011 to 2013, he was a Post-Doctoral Researcher with the Laboratory of Brain-Computer Interfaces, Graz University of Technology, Graz, Austria. He is currently a Senior Lecturer with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. His research interests focus on brain–computer interfaces (BCIs), nonlinear dynamics, machine learning, signal processing, and connectivity analysis in electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). He is also interested in the neurophysiological correlates of motor control and stimuli perception and how they differ between healthy participants and individuals with neurological and physiological impairments.



Yixin Chen received the B.Eng. degree in robotics engineering from Qingdao University of Science and Technology, Qingdao, China, in 2019, where his work centered on embedded robotics technology. He is currently pursuing the M.Eng. degree in control science and engineering with East China University of Science and Technology, Shanghai, China, where his research focuses on electroencephalography (EEG) signal processing and brain–computer interfaces.

His research interests lie in the intersection of neural networks, robotics, and human–machine interaction, aiming to contribute to advancements in brain–computer interfaces and intelligent robotics.



Xiao Wu received the bachelor's degree in mathematics and applied mathematics from the Dongchang College, Liaocheng University, Liaocheng, China, in 2020, and the master's degree in mathematics from the School of Mathematical Sciences, Liaocheng University, in 2023. She is currently pursuing the Ph.D. degree in control science and engineering with East China University of Science and Technology, Shanghai, China.

Her main research interests include brain–computer interfaces, deep learning, and pattern recognition.



Xinjie He received the B.E. degree from the Electrical Information Institute, Hunan University of Engineering, Xiangtan, China, in 2018, and the M.D. degree in control science and engineering from Hunan University of Science and Technology, Xiangtan, in 2021. He is currently pursuing the Ph.D. degree with East China University of Science and Technology, Shanghai, China.

His main research interests spans several fields, focusing on signal processing during his undergraduate studies, and extending to brain–computer interfaces and pattern recognition for his doctoral research.



Xingyu Wang was born in Sichuan, China, in 1944. He received the B.S. degree in mathematics from Fudan University, Shanghai, China, in 1967, and the M.S. degree in control theory and the Ph.D. degree in industrial automation from East China University of Science and Technology (ECUST), Shanghai, in 1982 and 1984, respectively.

He is currently a Professor with the School of Information Science and Engineering, ECUST. His research interests include control theory, control techniques, applications to biomedical systems, and brain control.



Andrzej Cichocki (Life Fellow, IEEE) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (Habilitation) degrees in electrical engineering from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively.

He spent several years at the University of Erlangen, Erlangen, Germany, as an Alexander-von-Humboldt Research Fellow and a Guest Professor. From 1995 to 2017, he was a Senior Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako, Japan. He is currently with the Systems Research Institute, Polish Academy of Sciences, Warsaw, with RIKEN Advanced Intelligence Project, and also with Tokyo University of Agriculture and Technology, Fuchu, Japan. His research focuses on multiway uninformed source separation, tensor decomposition, tensor networks for big data mining, and brain–computer interface.



Jing Jin (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from East China University of Science and Technology (ECUST), Shanghai, China, in 2010.

His Ph.D. advisors were Prof. Xingyu Wang at ECUST, from 2006 to 2008, and Prof. Gert Pfurtscheller at Graz University of Technology, Graz, Austria, from 2008 to 2010. He is currently a Professor with ECUST and the Vice Dean of the School of Mathematics. His research interests include brain–computer interface, signal processing, and pattern recognition.

Prof. Jin currently serves as an Associate Editor for *Cognitive Neurodynamics*, the *Journal of Neuroscience Methods*, and *Frontiers in Neurobotics*, an Action Editor for *Neural Networks*, and an Editor for the *Journal of Neural Engineering*.