Enhancing the Capability and Accuracy of Motor Imagery Classification: A Deep Neural Network-Powered Multifaceted Strategy Model

Weijie Chen, Ian Daly, Yixin Chen, Xiao Wu, Wei Liang, Xinjie He, Xingyu Wang, Andrzej Cichocki, *Fellow, IEEE*, Jing Jin*, *Senior Member, IEEE*

***Abstract*—Motor imagery is a popular non-invasive brain computer interface paradigm, yet its decoding accuracy remains hindered by the inherent non-stationarity and low signal-to-noise ratio of EEG signals. Current decoding frameworks often fail to fully exploit the intricate spatial-temporal dependencies, leading to suboptimal feature representation and the omission of latent discriminative cues. To address these challenges, we introduce a deep neural network-powered multifaceted strategy model (DPMS-Net), a novel approach that employs dynamic convolution to unearth effective discriminative cues across multiple dimensions, including the temporal, spatial, and frequency domains. This model synergizes channel and temporal attention mechanisms to adeptly capture the salient features of electroencephalogram (EEG) signals across diverse spatial-temporal dimensions, thereby mitigating the risk of omitting critical information. Furthermore, we introduce a spectral-domain analysis component that unearths subtle oscillatory signatures hidden within the EEG spectrum, providing enriched evidence for classification. We evaluated the performance of DPMS-Net on two publicly available datasets and a self-collected dataset from stroke patients. On the BCI Competition IV 2a and BCI Competition IV 2b datasets, DPMS-Net achieved subject dependent classification accuracies of 83.93% and 88.38%, respectively, alongside subject independent classification accuracies of 65.88% and 76.01%. In the stroke patient dataset, DPMS-Net attained a subject dependent classification accuracy of 67.67% and a subject independent classification accuracy of 57.58%. Experimental results indicate that DPMS-Net possesses efficient decoding capabilities and robust stability, reflecting its potential for deployment in neurorehabilitation BCI systems.**

***Index Terms*—Brain computer interface, motor imagery, electroencephalogram, deep neural network.**

I. INTRODUCTION

NON-INVASIVE brain computer interfaces (BCIs) based on electroencephalography (EEG) possess the remarkable capability to sensitively capture signals emitted by neuronal activity through the placement of electrodes on the scalp [1-3]. At present, EEG-based brain-computer interfaces predominantly employ three principal paradigms: Event-Related Potentials [4], Steady-State Visual Evoked Potentials [5-8], and Motor Imagery (MI) [9-14]. These paradigms have found extensive application across various domains, including emotion recognition [15-16], spelling devices [17-19], mobile robots [20-21], and rehabilitation therapies [22].

The MI BCI paradigm requires subjects to mentally simulate limb movements. This mental imagery activates the corresponding motor cortical neurons and generates patterns of activity in the EEG signals which can be detected and decoded. The precision of the MI decoding outcomes often determines the feasibility of subsequent applications, influencing the alignment between the BCI system and user expectations, and impacting the efficacy of a range of rehabilitation training and intelligent assistive control projects derived from MI paradigm [23-25].

In order to mitigate the potential limitations of the model's decoding capabilities, numerous traditional machine learning classification algorithms have been introduced [26-31]. For instance, common spatial pattern (CSP) is employed to identify discriminative characteristics of EEG signals across different categories by utilizing covariance matrices [27]. To fully exploit the frequency information inherent in EEG signals and to diminish CSP's sensitivity to the frequency ranges of these signals, filter bank common spatial patterns was introduced [28]. In the literature [29], Luo et al. decomposed time segments and frequency bands based on the MapReduce framework to extract salient features from EEG signals. Furthermore, Gaur et al. proposed two innovative sliding window techniques to bolster binary classification of MI [30]. Huang et al. proposed a Multiple Regularized Knowledge Transfer learning method

This work was supported by the STI 2030-major projects 2022ZD0208900 and Grant National Natural Science Foundation of China under Grant 62176090; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX. This research is also supported by 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. (*Corresponding author: Jing Jin).

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

Weijie Chen, Yixin Chen, Xiao Wu, Wei Liang, 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; liangwei.gd@foxmail.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, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK (e-mail: i.daly@essex.ac.uk).

Andrzej Cichocki is with the Systems Research Institute of Polish Academy of Sciences, 01-447bWarsaw, and Nicolaus Copernicus University (UMK), 87-100 Torun, Poland (e-mail: cichockiand@gmail.com).

and incorporated four regularization strategies to enhance the robustness of motor imagery decoding [31].

While the evolution and implementation of machine learning has reached a level of maturity, numerous inherent limitations persist. Traditional machine learning is profoundly reliant on manual feature engineering, making it challenging to uncover deep and complex data patterns, as well as to capture nonlinear and dynamic relationships, thereby limiting the performance of the models. In contrast, deep learning facilitates the construction of richly hierarchical end-to-end neural networks capable of autonomously deriving effective feature representations from the data. Furthermore, deep learning excels at modeling complex data relationships, adeptly capturing abstract features that span from low-level representations to high-level abstractions.

Convolutional neural networks (CNNs) represent an advanced class of deep learning models inspired by the biological mechanisms underlying visual perception. In the realm of EEG signal classification tasks, by designing appropriate convolutional kernels and architectural frameworks, CNNs have been shown to autonomously and efficiently derive temporal and spatial features directly from EEG data [32-43]. For instance, Lu et al. proposed a dynamic hierarchical convolutional attention network that captures time-frequency information and extracts global and local spatial features from motor imagery neural patterns [34]. Wang et al. synergistically integrated multi-scale spatiotemporal convolutions with Riemannian geometry, facilitating classification through sophisticated feature extraction modules and manifold embeddings [36]. Ke et al. introduced frequency adapters to convert signals from the time domain into the frequency domain, applied adaptive attention-weighted mechanisms prior to inverse transformation, and amalgamated convolutional embeddings with temporal periodicity modules to proficiently extract spatiotemporal representations [38]. Altaheri et al. initially employed sliding window techniques for data augmentation, subsequently harnessing convolutional blocks alongside multi-head self-attention mechanisms to distill high-level temporal features, thereby accentuating salient information [39]. Furthermore, Chen et al. proposed an adaptive attention-modulated graph network that adaptively captures salient inter-channel relationships and alleviates the over smoothing issue in feature representations, thereby enhancing subject adaptability to variability in brain connectivity and representation [42]. Lin et al. segmented EEG signals into blocks and employed a hybrid attention module to capture both local and global temporal patterns while constructing spatial representations across EEG channels, effectively modeling the temporal semantic information in EEG data [43].

EEG signals, as temporal sequences, not only encompass rich temporal information but also convey substantial spatial information owing to the presence of multiple channels. Furthermore, frequency domain information is inherently rich and valuable. Niu et al. integrated FIR filters within end-to-end networks to effectively harness frequency domain features [44-45]. Nonetheless, prevailing models predominantly emphasize the extraction of temporal and spatial characteristics, while the thorough exploration and utilization of frequency domain information remain to be further investigated. Furthermore, existing models frequently fail to facilitate interactions among the extracted features, leading to an excessive degree of independence among them. Meanwhile, the efficacy of the MI paradigm is profoundly reliant on the subjective awareness of subjects, which is largely modulated by individual levels of concentration. This challenge is especially pronounced among stroke patients, as damage to their motor cortical areas results in EEG signals generated during MI that are not only diminished in strength, compared to healthy adults, but also exhibit a low signal-to-noise ratio. Such subpar signal quality significantly undermines the accuracy of MI decoding models, ultimately constraining the advancement and application of MI in clinical rehabilitation.

To address the aforementioned challenges and enhance both the subject dependent and subject independent performance of MI for both healthy individuals and stroke patients, we introduce a deep neural network-powered multifaceted strategy model (DPMS-Net). DPMS-Net is a MI decoding model that synergizes dynamic convolution [46] with a multifaceted strategy. A horizontal spatial correlation module and a global spatial perception module are employed to extract local spatial information between EEG channels and global spatial information, respectively. The multidimensional temporal perception module utilizes four convolution kernels of varying dimensions to comprehensively capture the temporal information embedded in the EEG signals. The temporal variance insight module is adept at delineating the complex temporal variations inherent in the signals. The adaptive fine-grained channel attention module integrates both local and global channel information, facilitating a holistic reallocation of channel weights. Lastly, the frequency-related information perception module extracts salient frequency domain features from the EEG signals. The neuron focus mechanism module calculates the significance of each neuron, enabling the allocation of optimal weights within the model. The abbreviations of the modules used are summarized in Table I

The principal contributions can be articulated as follows:

- This paper introduces DPMS-Net, which transcends the limitations of traditional single-domain feature extraction. By innovatively employing dynamic convolution, we construct a multifaceted module that simultaneously resolves features from temporal, spatial, and frequency perspectives, significantly enriching the signal representation compared to existing fixed-kernel models.
- To overcome limitations in feature interaction observed in existing models, DPMS-Net incorporates an adaptive fine-grained channel attention module that effectively integrates both local and global information to thoroughly assess channel significance. Furthermore, the introduction of a temporal variance insight and neuron focus mechanism specifically targets the dynamic and non-stationary patterns of EEG, thereby boosting the model's sensitivity to complex temporal variations.
- DPMS-Net has been evaluated on two public datasets and a self-collected dataset from stroke patients through both subject dependent and subject independent experiments.

TABLE I
ABBREVIATIONS OF THE MODULES USED.

Module	Abbreviation
Horizontal Spatial Correlation Module	HSCM
Adaptive Fine-Grained Channel Attention Module	AFCAM
Frequency-Related Information Perception Module	FIPM
Neuron Focus Mechanism Module	NFMM

Experimental results demonstrate that DPMS-Net achieves superior classification accuracy and robust generalization performance, highlighting its potential and practical value in neural signal decoding tasks.

II. METHODS

DPMS-Net is primarily composed of eight modules based on dynamic convolution. The comprehensive structure is depicted in Fig. 1. This section will offer an in-depth exposition of the design and functionality of dynamic convolution within each module.

A. Dynamic convolution

In traditional convolution, the convolutional kernels are static, which limits the ability to capture individualized features of various samples. In contrast, dynamic convolution introduces multiple sets of kernels along with an attention mechanism that dynamically generates weights for each input feature, facilitating a weighted amalgamation of these diverse kernel sets. This approach allows the convolutional kernels to be adaptively adjusted for each input sample, enabling a more precise alignment with the input features, thereby enhancing the model's capacity to capture complex data patterns and significantly augmenting its flexibility and adaptability. Moreover, dynamic convolution utilizes a lightweight weight generation module that not only elevates the model's expressive power but also maintains a low computational burden. When compared to straightforward methods that merely increase the parameter count, dynamic convolution effectively mitigates the risk of overfitting while substantially improving classification accuracy and generalization performance, constituting a significant advancement over traditional convolution.

Dynamic convolution can generate convolutional kernels $W_{dynamic}$ dynamically through a series of steps, as depicted in Fig. 2. Let the input be denoted as $x_d \in \square^{B\times F\times C\times T}$. First, a global average pooling operation is applied to the input to obtain $x_{da} \in \square^{B\times F\times 1\times 1}$ for capturing global information. Then, K sets of fixed convolutional kernels $\{W_1, W_2, \frac{1}{4}, W_K\}$ are designed for each layer. Subsequently, a lightweight attention mechanism, formulated as $f(\mathrm{g})$ and based on two-dimensional convolution, is utilized to dynamically derive the weighting coefficients $\{a_1, a_2, ..., a_K\}$ from the input data x using the formula $a = \mathrm{soft}\max(f(x))$. These coefficients are employed to effectuate a weighted amalgamation of the K sets of convolutions:

$$W_{dynamic} = \sum_{k=1}^{K} a_k W_k \tag{1}$$

Finally, the dynamic convolutional kernel $W_{dynamic}$ is employed to process the input features x through a standard two-dimensional convolution operation, thereby producing the output y of the dynamic convolution:

$$y = x_d' \; W_{dynamic} \hat{\mathrm{I}} \; \text{¡}^{B' F' C' T} \tag{2}$$

B. Horizontal spatial correlation module and global spatial perception module

Current mainstream deep learning models conventionally implement spatial convolutions with a kernel size corresponding to the number of EEG electrodes to allow them to capture global spatial information. However, this strategy may hinder the model's ability to effectively adapt to the localized correlation patterns inherent in EEG signals. Consequently, as depicted in Fig. 1, DPMS-Net integrates a horizontal spatial correlation module specifically designed to capture nuanced local spatial dependencies among adjacent electrodes along the horizontal axis. By emphasizing the correlations between neighboring electrodes, this module uses an $[S,1]$ convolutional kernel to mitigate the risk of information dilution commonly associated with indiscriminate global feature extraction across disparate regions, thereby enhancing the model's capacity to harness and exploit local relational features. Subsequently, we leverage the global spatial perception module, using a $[C,1]$ convolutional kernel to extract global spatial information, effectively integrating local and global features to significantly enhance the capacity for feature representation.

C. Multidimensional temporal perception module

Temporal convolution is recognized as the most prevalent and efficient method for extracting temporal information in EEG signal processing. However, traditional single-scale temporal convolution often confines its attention to fixed-length time windows, making it challenging to simultaneous consideration of low-frequency, mid-frequency, and high-frequency components inherent in EEG signals.

To adequately capture the temporal features embedded within EEG signals, DPMS-Net integrates the multidimensional temporal perception module. This module utilizes four temporal convolution kernels of varying sizes: $[1,T_1]$, $[1,T_2]$, $[1,T_3]$, and $[1,T_4]$, enabling it to address both short-term rapid dynamic fluctuations and long-term trending behavior. This approach facilitates the extraction of local temporal features alongside global dynamic patterns within the EEG. Such a design allows the model to concurrently capture both short-term and long-term temporal dependencies, enabling a comprehensive understanding of the diverse frequency and temporal scale characteristics of EEG. Moreover, this module not only effectively mitigates the risk of information loss commonly associated with single temporal scale convolutions but also significantly enhances feature diversity and the robustness of the model.

D. Temporal variance insight module

EEG signals encompass a wealth of temporal information; however, owing to their inherent vulnerability to noise and interference, as well as the non-uniform distribution of salient information across time points, not all temporal segments contribute constructively to classification tasks. Certain time points may be dominated by noise or irrelevant fluctuations,

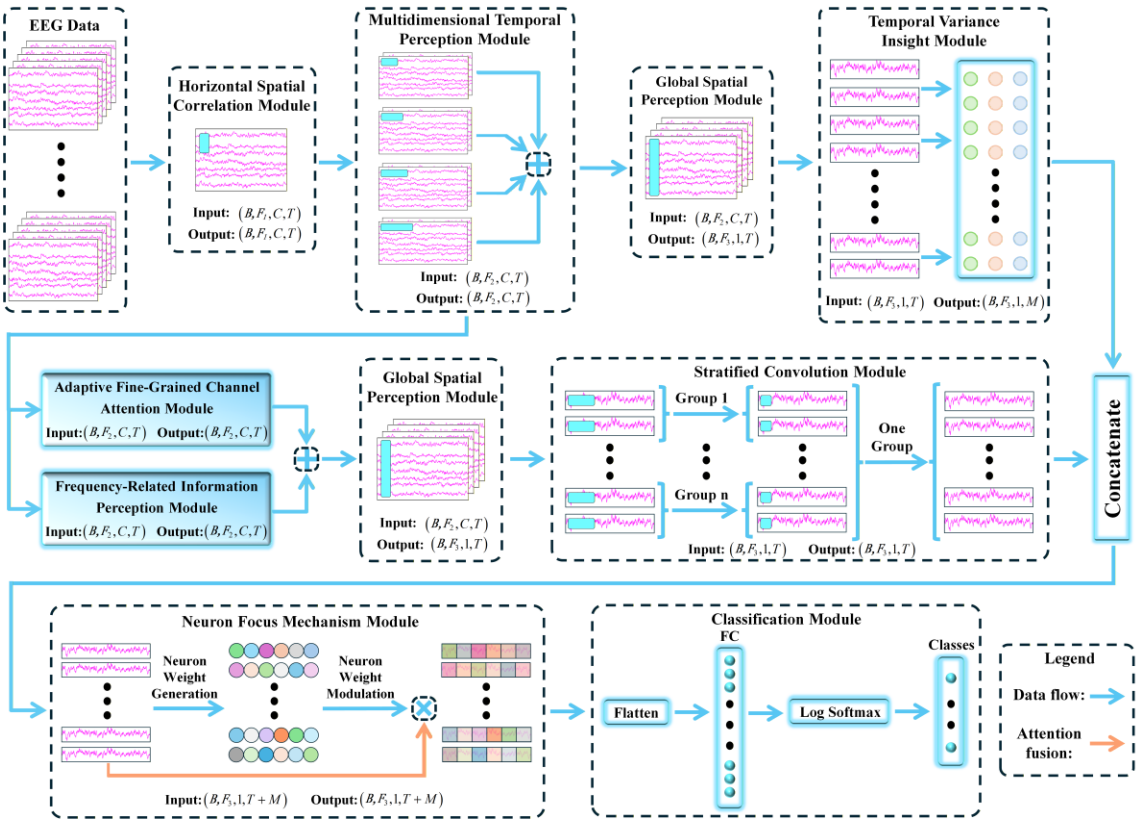


Fig. 1. The comprehensive architecture of DPMS-Net, where FC signifies the fully connected layer.

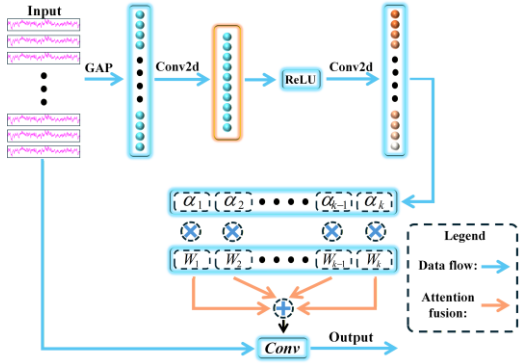


Fig. 2. The architecture of dynamic convolution.

and indiscriminately leveraging features from the entire temporal sequence can undermine the discriminative power of the model. The temporal variance insight module aims to address this challenge by concentrating on the overarching temporal dynamics, selectively amplifying temporal features that hold greater relevance for the classification objective. This mechanism mitigates temporal dependencies, filters out irrelevant and redundant information along the temporal dimension, and thereby augments the efficacy of feature extraction. Furthermore, EEG signals manifest pronounced inter-subject variability. The dynamic features derived from the temporal variance insight module adeptly capture the patterns of signal fluctuations, ultimately enhancing the model's robustness and generalization capacity.

Let the features fed into the temporal variance insight module be denoted as $x \in \mathbb{R}^{B\times F\times 1\times T}$, where B signifies the number of samples, F represents the number of channels, and T indicates the total number of time points. First, we perform a reshaping operation on x to derive $x_{re} \in \mathbb{R}^{B\times F\times M\times L}$, where M denotes the number of time windows and L is computed as T/M. Subsequently, for each sample $b \in [1,B]$ and each feature map $f \in [1,F]$, we compute the temporal variance for each time window $m \in [1,M]$:

$$\mathrm{Var}(x_{re}(b,f,m,:)) = \frac{1}{L}\sum_{l=1}^{L}\left(x_{re}(b,f,m,l) - m_{b,f,m}\right)^2 \quad (3)$$

where $m_{b,f,m}$ denotes the mean for each time window. Finally, we implement a logarithmic transformation on the variance as described by the subsequent formula:

$$x_{lv}(b,f,m) = \log(\mathrm{Var}(b,f,m) + z) \quad (4)$$

where x_{lv} is the output of the temporal variance insight module, $\mathrm{Var}(b,f,m)$ represents the variance of each time window and z is a small positive constant to prevent the logarithmic operation from resulting in negative infinity.

E. Adaptive fine-grained channel attention module

Channel attention is a widely used feature extraction technique for identifying channels with the most salient features. Nevertheless, contemporary channel attention methodologies predominantly emphasize global channel attention, thereby neglecting the interdependencies among local channels. To address this limitation, we introduce a novel approach that integrates both local and global channel attention: adaptive fine-grained channel attention [47], as illustrated in Fig. 3.

In this framework, we apply global average pooling to the input feature map, thereby extracting comprehensive information:

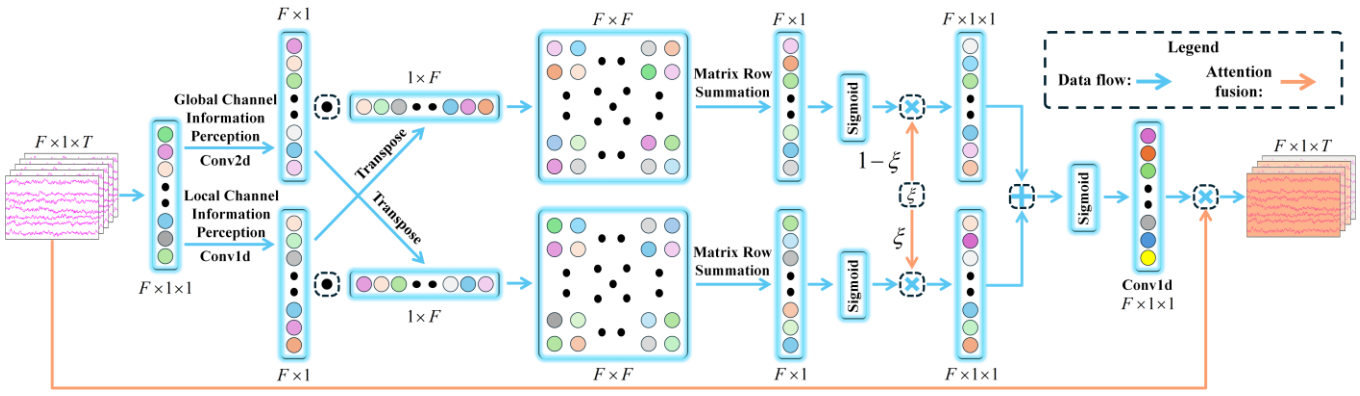
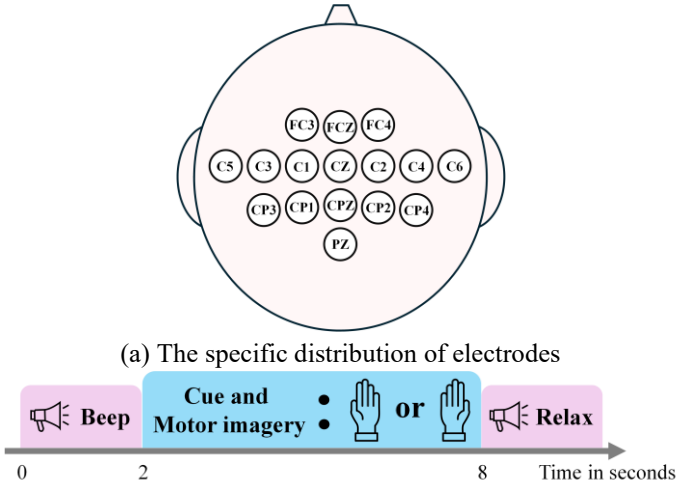


Fig. 3. The comprehensive architecture of adaptive fine-grained channel attention module.



(a) The specific distribution of electrodes

(b) The time configuration for each trial

Fig. 4. Electrode distribution and experimental design of Dataset 3.

$$x_a = \mathrm{AvgPool}(x_i) \in \mathbb{R}^{B\times F\times 1\times 1} \tag{5}$$

where x_i refers to the input of the adaptive fine-grained channel attention module, and x_a represents the output after global average pooling.

Then, Conv1d and Conv2d are employed to capture local detail information and global feature patterns from the respective local and global channels:

$$\begin{aligned} x_1 &= \mathrm{Conv1d}\left(\mathrm{squeeze}(x_a,-1)\right) \in \mathbb{R}^{B\times F\times 1} \\ x_2 &= \mathrm{squeeze}\left(\mathrm{Conv2d}(x_a),-1\right) \in \mathbb{R}^{B\times F\times 1} \end{aligned} \tag{6}$$

where x_1 and x_2 represent the outputs of Conv1d and Conv2d, respectively.

Subsequently, we compute a correlation matrix through a cross-correlation operation on the extracted local and global information, and then sum the matrix row-wise, aiming to capture the relationship between the two:

$$\begin{aligned} x_3 &= \sum\left(x_1 \times x_2^{Tr}\right) \in \mathbb{R}^{B\times F\times 1} \\ x_4 &= \sum\left(x_1^{Tr} \times x_2\right) \in \mathbb{R}^{B\times F\times 1} \end{aligned} \tag{7}$$

where Tr denotes the transpose operation, x_3 and x_4 represent the outputs of row-wise summation of the two matrices, respectively.

This correlation matrix is then adaptively fused with the learnable factor ξ using the sigmoid activation function to generate the final attention weights:

$$x_m = \sigma\left[\mathrm{Conv1d}\left(\mathrm{unsqueeze}\left(\xi\sigma(x_3)+(1-\xi)\sigma(x_4),-1\right)\right)\right] \in \mathbb{R}^{B\times F\times 1\times 1} \tag{8}$$

where, x_m denotes the generated attention weights, and $\sigma(\cdot)$ represents the sigmoid activation function.

Finally, these weights are element-wise multiplied with the input feature map, resulting in a weighted output x_{out} that enables a reasonable allocation of weights to the feature channels:

$$x_{out} = x_i \times x_m \in \mathbb{R}^{B\times F\times 1\times T} \tag{9}$$

Within the adaptive fine-grained channel attention module, local attention can focus on specific regions of the input feature map, capturing fine-grained features that are closely associated with motor imagery activities. Conversely, global attention encapsulates the long-range dependencies among channels across the entire spectrum, capturing overall trends and temporal variations. By amalgamating both local and global attention mechanisms, this design not only effectively extracts localized personalized features but also establishes global patterns across individuals, thereby augmenting the model's adaptability to complex tasks. This approach facilitates a thorough integration of local and global features, significantly enhancing the feature extraction capabilities and the overall performance of the model in motor imagery classification tasks.

F. Frequency-related information perception module

Existing models predominantly emphasize the extraction of temporal and spatial features, frequently neglecting the latent value of frequency domain characteristics. Therefore, we propose a frequency-related information perception module for the analysis of frequency domain information from the EEG. This module attempts to discern trends in critical frequency energy fluctuations, thereby capturing overall global patterns. Moreover, frequency domain operations inherently possess filtering attributes, enabling them to effectively suppress high-frequency noise and task-irrelevant interference, while simultaneously focusing on specific frequency band features relevant to the task at hand. This dual capability markedly enhances the model's robustness against noise. Additionally, through the synthesis of multi-domain features, this module is capable of delivering a more abundant representation of the signals, facilitating the mitigation of individual variances and

TABLE II
DETAILS OF THE PARAMETER CONFIGURATION VALUES FOR DPMS-NET.

Module	Parameter setting
Horizontal spatial correlation module	$S = 2$, $F_1 = 1$
Multidimensional temporal perception module	$T_1 = 5$, $T_2 = 25$, $T_3 = 45$, $T_4 = 65$
	$F_2 = 32$
Global spatial perception module	C = Channels
	$F_3 = 64$
Temporal variance insight module	$M = 10$
	ζ = 1e-6
Stratified convolution module	$T_5 = 25$
Dynamic convolution	$K = 4$
	Compression rate = 0.25

consequently bolstering the accuracy and generalization abilities of the classifier.

Let the features input into this module be represented as x_F Î ¡ $^{B' F' C' T}$, where C denotes the number of EEG electrodes. Subsequently, we conduct a fast Fourier transform along the C and T dimensions of x_F:

$$F(x_F) = FFT(x_F(B,F,:,:)) \tag{10}$$

Next, we decompose $F(x_F)$ into its real and imaginary components:

$$F(x_F) \to \left[Re(F(x_F)), Im(F(x_F)) \right] \tag{11}$$

Then, we execute a convolution on the amalgamated frequency domain features derived from both the real and imaginary parts.

$$F'(x_F) = Conv2D\left(\left[Re(F(x_F)), Im(F(x_F)) \right]\right) \tag{12}$$

Finaly, we reconstruct the convolved frequency domain features into their complex form and apply the inverse Fourier transform to revert them to the time domain, thereby producing $x' \in$ □ $^{B\times F\times C\times T}$ as the resultant output.

$$x' = IFFT\left(Re(F'(x_F)) + j \cdot Im(F'(x_F))\right) \tag{13}$$

G. Stratified convolution module

Integrating spatiotemporal information extracted from EEG signals is critically important. A prevalent approach employs depthwise separable convolution, which markedly reduces computational complexity. However, this technique treats each channel in isolation, resulting in insufficient inter-channel information exchange. To overcome this limitation, we introduce the stratified convolution module. Initially, channels are partitioned into groups, with the number of groups set to half the total number of feature maps, and the convolution kernel is configured as $[1, T_5]$. Subsequently, a pointwise convolution is applied across all feature maps, effectively treating all channels as a single group at this stage. The stratified convolution module adeptly captures spatial correlations among multiple channels within localized groups, mitigating the drawbacks of completely independent channel processing that may neglect synergistic inter-channel relationships. This suppressing feature redundancy, and enhancing DPMS-Net's proficiency in salient feature learning. The neuron focus design facilitates an integration of localized and global channel fusion while maintaining a minimal parameter footprint and computational overhead, ultimately augmenting feature fusion efficacy and representational capacity.

Subsequently, we concatenate the outputs of the stratified convolution module and the temporal variance insight module. Let the output of the stratified convolution module be $x_s \in$ □ $^{B\times F\times 1\times T}$, and the output of the temporal variance insight module be $x_t \in$ □ $^{B\times F\times M\times 1}$. First, x_t is reshaped to have the dimensions $[B, F, 1, M]$. Then, x_s and x_t are concatenated along the temporal dimension to obtain the final output $x_{out} =$ □ $^{B\times F\times 1\times (T+M)}$.

H. Neuron focus mechanism module

The majority of prevailing models predominantly emphasize channel attention and spatial attention, overlooking the temporal nature of EEG signals. To effectively capture the essential temporal features embedded within the EEG, we propose the incorporation of a temporal attention mechanism, the neuron focus mechanism module [48]. This module dynamically generates attention weights based on the statistical properties of the features, facilitating weighted adjustments without necessitating additional parameter learning, thereby significantly enhancing classification performance.

First, we compute the mean μ_N of the input features $x_N \in$ □ $^{B\times F\times 1\times T'}$ across the spatial-temporal dimensions for each channel, where T' denotes the time points fed into the module:

$$\mu_N = \frac{1}{1\times T'} \sum_{t=1}^{T'} x_N \tag{14}$$

Then, we determine the variance for each channel:

$$v_N = \frac{1}{1\times T' - 1} \sum_{t=1}^{T'} (x_N - \mu_N)^2 \tag{15}$$

Thereafter, we evaluate the significance of each feature and utilize the Sigmoid activation function for weighting, culminating in the resultant output $x_N^{'}$:

$$x_N^{'} = x_N \times \text{Sigmoid}\left(\frac{(x_N - \mu_N)^2}{4\times (v_N + 10^{-4})} + 0.5 \right) \tag{16}$$

I. The analysis of module mechanisms

DPMS-Net employs dynamic convolution to adaptively modulate the convolutional kernel weights across its constituent modules, enabling the extraction of salient EEG signal characteristics. By integrating the horizontal spatial correlation module alongside the multidimensional temporal perception module, it synthesizes spatiotemporal feature maps with superior discriminative capacity. Subsequently, the frequency-related information perception module leverages spectral analysis to construct a rich, joint time-frequency feature space, enabling the model to learn discriminative information from a more comprehensive signal distribution, thereby amplifying the representational richness and distinctiveness of the features. Concurrently, the adaptive fine-grained channel attention module dynamically calibrates channel-wise weights, facilitating the effective fusion of local and global information, mechanism module implements a weighted mechanism to selectively emphasize critical channels and temporal segments,

effectuating dynamic weighted projections within the feature manifold. The multidimensional synergistic modeling and effective fusion of these modules builds a high-order feature space characterized by reinforced discriminative boundaries and robustness, significantly enhancing the model's capability in recognizing intricate EEG signal patterns. The synergistic collaboration of these mechanisms attenuates the interference from irrelevant features during training, augments the specificity and stability of feature extraction, thereby boosting the generalization ability of the DPMS-Net.

III. Datasets Description and Experimental Settings

A. Datasets description

1) Dataset 1

The BCI Competition IV 2a dataset encompasses motor imagery data from nine subjects, recorded through three EOG electrodes and twenty-two EEG electrodes positioned in accordance with the international 10-20 system, at a sampling frequency of 250 Hz. This dataset comprises four distinct categories of motor imagery tasks (left hand, right hand, both feet, and tongue), with each subject undertaking two sessions, resulting in a total of 576 trials. Each session consists of 288 trials, with 72 trials allocated to each category. The BCI Competition IV 2a dataset is available for download at https://www.bbci.de/competition /iv/download.

2) Dataset 2

The BCI Competition IV 2b dataset encompasses motor imagery data from nine subjects, recorded utilizing three EOG electrodes and three EEG electrodes in accordance with the international 10-20 system, at a sampling frequency of 250Hz. This dataset comprises two distinct categories of motor imagery tasks (left hand and right hand), with each subject engaging in five sessions. The initial two sessions are conducted without visual feedback, with each session comprising 120 trials, allocated as 60 trials for the left hand and 60 trials for the right hand. The subsequent three sessions incorporate visual feedback, with each session consisting of 160 trials, divided into 80 trials for the left hand and 80 trials for the right hand. The BCI Competition IV 2b dataset is available for download at https://www.bbci.de/competition/iv/download.

3) Dataset 3

Dataset 3 constitutes a laboratory self-collected dataset encompassing motor imagery data from twenty four stroke patients, recorded using the g.USBamp system (g.tec medical engineering GmbH, Austria), featuring 16 EEG channels (FC-3/4/Z, C-1/2/3/4/5/6/Z, CP-1/2/3/4/Z and PZ) at a sampling frequency of 256 Hz, The distribution of these electrodes conforms to the international 10-20 electrode placement system, with each electrode's impedance maintained below 10 kΩ. The specific locations of the electrodes are shown in Figure 4(a). Continuous real-time monitoring of electrode impedance is conducted throughout the acquisition process to ensure signal quality. The experiments are executed by experienced researchers rigorously complying with the laboratory's standardized operating protocols, including comprehensive equipment calibration and data acquisition guidelines. The data are thoroughly inspected to ensure accuracy and consistency after data collection. Moreover, all data are comprehensively backed up and securely stored to ensure the guarantee integrity. During the experimental protocol, subjects are instructed to maintain a state of relaxation while striving to keep their bodies stable. This dataset comprises two distinct categories of motor imagery tasks (left hand and right hand), with each subject undertaking 60 trials. Each trial commences with an auditory cue, followed by a visual prompt appearing after 2 seconds, which directs subjects to engage in a 6-second motor imagery task with either the left or right hand, as depicted in Fig. 4(b).

B. Experimental settings

In the subject dependent experiments, the first session of Dataset 1 is designated as the training set, while the subsequent session is allocated as the test set. For Dataset 2, the first three sessions are employed as the training set, whereas the last two sessions are reserved for testing. In the case of Dataset 3, the first 40 trials are utilized as the training set, with the remaining 20 trials serving as the test set. Additionally, both Dataset 1 and Dataset 2 employ 4-second motor imagery segments, while Dataset 3 utilizes 6-second motor imagery segments. In the subject independent experiments, the complete dataset from a single subject is retained for testing, while the data from all other subjects is utilized for training.

The detailed parameter configuration of DPMS-Net is presented in Table II. All datasets are processed using a finite impulse response bandpass filter within the 0.5-40 Hz range, effectively eliminating low-frequency baseline drift and high-frequency electromyographic artifacts, while preserving auxiliary neural activity features beyond the μ and β frequency bands. To maintain consistency with Datasets 1 and 2, Dataset 3 is resampled to 250Hz using the resample function in MNE. Additionally, the loss function combines cross-entropy with R-Drop [49]. Dropout is set to 0.4. The batch size is fixed at 32, the Adam optimizer is employed, and a cosine annealing decay strategy is used to set the learning rate to 0.001.

IV. Experimental Results

This section describes the experimental results achieved with DPMS-Net across three datasets for both the subject dependent and subject independent evaluations, comparing these findings with other state-of-the-art (SoTA) methods. It is noteworthy that all DPMS-Net results in the table are reported as the mean over five independent runs. The best performing accuracy is highlighted in bold, and a two-tailed paired t-test is utilized to evaluate the statistical significance of the results. Statistical significance in all tables is indicated by an asterisk (*: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$), with the exact p-values explicitly reported. Additionally, the degrees of freedom are 8 for Dataset 1 and Dataset 2, and 23 for Dataset 3. Furthermore, we incorporated the F1-score as an additional metric to comprehensively assess the model's performance. The F1-score, defined as the harmonic mean of Precision and Recall, serves as an indicator of the model's balance between precision and sensitivity in identifying positive instances. The computation is

TABLE III
COMPARISON OF SUBJECT DEPENDENT CLASSIFICATION PERFORMANCE BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 1.

Methods	Subjects									Average accuracy	Standard deviation	F1-score	p value
	A1	A2	A3	A4	A5	A6	A7	A8	A9				
EEGNet [50]	79.86	58.68	89.93	64.93	63.19	58.68	64.24	73.61	77.08	70.22***	10.72	0.6891	0.0002
EEG-TCNet [51]	85.77	65.02	94.51	64.91	75.36	61.40	87.36	83.76	78.03	77.35**	11.58	0.7603	0.0097
FBCNet [35]	85.42	60.42	90.63	76.39	74.31	53.82	84.38	79.51	80.90	76.20**	11.97	0.7449	0.0010
DRDA [52]	83.19	55.14	87.43	75.28	62.29	57.15	86.18	83.61	82.00	74.75**	12.96	0.7258	0.0016
Conformer [53]	87.59	61.71	92.94	77.63	58.27	64.33	87.36	84.92	87.92	78.07*	13.14	0.7634	0.0309
IDAN [54]	83.29	63.97	90.30	76.94	69.34	60.08	89.31	82.35	82.81	77.60***	10.85	0.7582	0.0003
ATCNet [39]	85.07	67.71	93.75	77.54	77.08	68.29	88.15	**86.15**	87.85	81.29*	9.14	0.7989	0.0209
M-FANet [55]	87.61	69.93	91.92	73.61	76.84	63.39	85.17	78.63	87.18	79.36**	9.39	0.7746	0.0014
ADFCNN [58]	87.15	61.45	93.75	75.69	75.34	65.27	88.54	82.29	85.06	79.39**	10.85	0.7751	0.0075
HCA-Net [59]	83.68	61.81	94.44	78.47	**79.86**	68.75	89.93	84.72	83.68	80.59	10.08	0.7915	0.0525
NexusNet [60]	82.29	60.42	**96.53**	71.18	70.49	67.71	89.24	85.42	85.76	78.78*	11.81	0.7743	0.0250
DPMS-Net	**88.86**	**74.61**	94.87	**83.58**	78.49	**70.04**	**91.71**	83.97	**89.26**	**83.93**	8.23	0.8246	-

TABLE IV
COMPARISON OF SUBJECT DEPENDENT CLASSIFICATION PERFORMANCE BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 2.

Methods	Subjects									Average accuracy	Standard deviation	F1-score	p value
	B1	B2	B3	B4	B5	B6	B7	B8	B9				
FBCSP [28]	70.00	60.36	60.94	97.50	93.12	80.63	78.13	92.50	86.88	80.00**	13.85	0.7964	0.0047
EEGNet [50]	70.31	70.36	78.44	95.33	93.44	82.18	91.88	87.19	71.65	82.37*	10.15	0.8187	0.0127
EEG-TCNet [51]	69.69	69.07	77.22	96.14	91.13	74.57	84.96	91.25	74.38	81.21**	10.14	0.8046	0.0024
FBCNet [35]	67.38	61.21	67.19	90.06	93.75	72.31	80.56	88.44	71.56	76.94***	11.64	0.7597	0.0002
Conformer [53]	**82.06**	67.14	63.18	97.12	87.03	88.91	87.81	93.29	**91.37**	84.22	11.49	0.8339	0.0962
FBMSNet [61]	72.81	53.21	80.31	97.19	95.00	85.63	84.38	90.00	88.75	83.03*	13.39	0.8201	0.0192
LightConvNet [62]	76.56	60.00	67.19	97.81	94.38	86.88	84.69	92.50	89.38	83.26*	12.83	0.8225	0.0173
M-FANet [55]	67.77	68.24	81.03	96.33	89.91	81.87	89.55	92.56	77.91	82.80**	10.25	0.8186	0.0039
SWCNet [56]	71.04	66.76	82.31	96.33	86.45	81.19	90.06	87.61	83.92	82.85**	9.16	0.8213	0.0016
SST-DPN [57]	77.50	68.93	**82.81**	96.88	96.25	87.50	89.06	93.44	87.50	86.65**	9.10	0.8577	0.0041
DPMS-Net	78.77	**71.81**	82.13	**98.16**	**96.58**	**89.63**	**92.12**	**95.60**	90.64	**88.38**	8.96	0.8801	-

expressed as follows:

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{17}$$

where $\mathrm{Precision} = TP/(TP+FP)$ and $\mathrm{Recall} = TP/(TP+FN)$. TP denotes the number of true positive samples, FP denotes the number of false positive samples, and FN signifies the number of false negative samples.

A. Subject dependent

The subject dependent experimental results of DPMS-Net alongside several SoTA methods on Dataset 1 are presented in Table III. Notably, DPMS-Net achieved the highest accuracy for the majority of subjects. In comparison to the classic EEGNet, DPMS-Net exhibits a remarkable advantage of over 10%. Furthermore, when compared to SoTA methods such as ATCNet, ADFCNN, and HCA-Net, DPMS-Net demonstrates a lead of more than 2.6%. Additionally, DPMS-Net exhibits the smallest standard deviation among all evaluated methods, recorded at merely 8.19, underscoring its excellent stability and commendable generalizability across different subjects. It is significant to note that DPMS-Net shows statistical significance when compared to all other methods, except for HCA-Net.

The subject dependent classification results of DPMS-Net and the other SoTA methods applied to Dataset 2 are illustrated in Table IV. Conformer achieved the highest accuracy for the first and ninth subjects, while DPMS-Net obtained the highest accuracy for all other subjects except for subject 3. It is particularly noteworthy that DPMS-Net's standard deviation is the only one below 10, measured at just 8.96, demonstrating its robustness.

We also compared DPMS-Net to SoTA methods on Dataset 3. The experimental results are presented in Table V. EEGNet, ADFCNN, EEG-TransNet and TMSA-Net all attained accuracies exceeding 60%; however, DPMS-Net achieved the highest accuracy, accompanied by a standard deviation that surpasses that of most SoTA models. Although Conformer's accuracy is relatively unsatisfactory, it recorded the smallest standard deviation.

B. Subject independent

The subject independent experimental results of DPMS-Net and the other SoTA methods across the three datasets are listed in Tables VI, VII, and VIII, respectively. In Dataset 1, DPMS-Net achieved an accuracy of 65.88%, surpassing all other SoTA methods in the table; but MIN2Net, ATCNet, and MBK-CNN also demonstrated commendable performance with notable accuracies. In Dataset 2, DPMS-Net emerged as the sole method to exceed 76% accuracy, while EEG-TCNet, ADFCNN and TMSA-Net also yielded robust results, each achieving accuracies above 72%. In Dataset 3, while DPMS-Net secured the highest accuracy along with the lowest standard deviation, the performance disparity among the various methods is not

TABLE V
COMPARISON OF SUBJECT DEPENDENT CLASSIFICATION PERFORMANCES BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 3.

Methods	Average accuracy	Standard deviation	F1-score	p value
EEGNet [50]	61.67*	12.72	0.6057	0.0211
DeepConvNet [63]	60.90**	12.62	0.5973	0.0091
EEG-TCNet [51]	57.08**	11.49	0.5613	0.0013
Conformer [53]	54.17***	6.4	0.5327	3.64e-5
ADFCNN [58]	62.08*	12.49	0.6125	0.0122
EEG-TransNet [64]	63.67	12.17	0.6276	0.0902
TMSA-Net [67]	63.73	11.93	0.6281	0.0664
DPMS-Net	**67.67**	11.07	0.6705	-

TABLE VI
COMPARISON OF SUBJECT INDEPENDENT CLASSIFICATION PERFORMANCE BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 1.

Methods	Average accuracy	Standard deviation	F1-score	p value
EEGNet [50]	53.39***	9.41	0.4824	0.0007
EEG-TCNet [51]	54.38**	10.49	0.4979	0.0036
MIN2Net [65]	60.03*	9.24	0.5589	0.0216
ATCNet [39]	61.85	12.67	0.5718	0.0797
MBK-CNN [66]	63.55	16.82	0.5903	0.3371
M-FANet [55]	55.86**	10.17	0.5126	0.0024
ADFCNN [58]	57.43**	13.26	0.5319	0.0071
TMSA-Net [67]	60.55*	13.90	0.5617	0.0327
DPMS-Net	**65.88**	12.72	0.6163	-

TABLE VII
COMPARISON OF SUBJECT INDEPENDENT CLASSIFICATION PERFORMANCE BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 2.

Methods	Average accuracy	Standard deviation	F1-score	p value
EEGNet [50]	71.07***	6.45	0.6624	7.06e-5
EEG-TCNet [51]	72.02***	6.36	0.6683	0.0005
FBCNet [35]	68.28**	5.71	0.6359	0.0060
Conformer [53]	69.24**	6.95	0.6452	0.0040
M-FANet [55]	69.84**	5.24	0.6581	0.0025
ADFCNN [58]	72.83*	7.71	0.6817	0.0452
TMSA-Net [67]	73.43*	7.21	0.6905	0.0188
DPMS-Net	**76.01**	6.41	0.7135	-

TABLE VIII
COMPARISON OF SUBJECT INDEPENDENT CLASSIFICATION PERFORMANCES BETWEEN DPMS-NET AND SOTA METHODS ON DATASET 3.

Methods	Average accuracy	Standard deviation	F1-score	p value
EEGNet [50]	53.25**	7.81	0.4988	0.0062
DeepConvNet [63]	52.58**	7.34	0.4832	0.0037
EEG-TCNet [51]	52.97*	6.84	0.4857	0.0125
Conformer [53]	54.73*	7.89	0.5019	0.0398
ADFCNN [58]	54.93	8.28	0.5028	0.0766
EEG-TransNet [64]	54.58	7.32	0.4996	0.0927
TMSA-Net [67]	53.28*	5.94	0.4911	0.0123
DPMS-Net	**57.58**	6.62	0.5328	-

statistically significant. This underscores the substantial potential for further exploration in the realm of MI recognition in stroke patients. In conclusion, DPMS-Net consistently exhibited superior accuracy across all three datasets, illustrating its remarkable robustness.

C. The analysis of experimental results

We validated DPMS-Net on three different feature datasets under both subject dependent and subject independent scenarios. From the perspective of EEG channel count, Dataset 1 contains data from 22 channels covering the primary motor areas, providing a relatively comprehensive acquisition of signals. Consequently, DPMS-Net achieved a notable accuracy of 83.93% on a four-class MI classification task. Conversely, Dataset 2 utilized only 3 EEG channels, positioned over the core electrodes of the left and right motor regions, effectively capturing the most salient signal features; consequently, DPMS-Net attained an accuracy of 88.38% on a two-class MI decoding task. Dataset 3 leveraged 16 electrodes covering the majority of motor areas for signal acquisition. Given that stroke patients exhibit varying extents of motor cortex impairment, all models struggled to surpass 70% accuracy even on binary MI tasks, with DPMS-Net achieving the highest accuracy of 67.67%.

Regarding inter-subject variability, the two publicly available datasets comprise data from 9 healthy subjects, displaying relatively low inter-subject variability compared to Dataset 3's 24 stroke patients. This reduced variability facilitates the model's capacity to extract more consistent and robust motor imagery features. In contrast, the pronounced physiological and pathological heterogeneity among stroke patients escalates the challenge of model generalization, as differences in signal amplitude and execution strategies across subjects can lead to substantial cross subject distribution shifts, a phenomenon corroborated by the subject independent experimental results.

Regarding the level of signal noise, the acquisition protocols and controlled environments of the two public datasets ensure high quality EEG recordings. Although real time monitoring of electrode impedance is conducted during the collection of our laboratory dataset, the inherent constraints of the clinical setting inevitably introduced extraneous noise artifacts, even after bandpass filtering, non-stationary noise may still persist, causing the feature distribution to vary across time and conditions, thereby affecting the model's feature extraction efficacy and decoding accuracy.

Although Dataset 1 and Dataset 2 each consist of only 9 subjects, a sufficient amount of data was collected per subject, with 576 trials and 720 trials respectively. This allows for minimizing the significant impact of low-quality trials on the decoding model. However, due to the challenges in acquiring EEG data, especially from stroke patients in Dataset 3, despite including 24 subjects, contains only 60 trials per subject. The relatively small amount of data per subject means that data quality has a greater influence on the model's decoding performance, weakening the model's ability to stably estimate discriminative features. This may be one of the reasons why the decoding accuracy of Dataset 3 is significantly lower than that of Datasets 1 and 2.

Future work will aim to scale up the dataset by broadening subject coverage and enriching task diversity, thereby

TABLE IX
ABLATION EXPERIMENTS OF DPMS-NET ON THREE DATASETS.

Module			Subject dependent accuracy			Subject independent accuracy		
HSCM	AFCAM +FIPM	NFMM	Dataset 1 (p value)	Dataset 2 (p value)	Dataset 3 (p value)	Dataset 1 (p value)	Dataset 2 (p value)	Dataset 3 (p value)
×	✓	✓	81.15 (0.0003)	86.85 (0.0037)	66.67 (0.3034)	65.08 (0.0059)	74.76 (0.0390)	56.53 (0.0307)
✓	×	✓	82.33 (0.0031)	86.19 (0.0054)	64.35 (0.0617)	64.97 (0.0050)	74.65 (0.0208)	55.49 (0.0498)
✓	✓	×	82.16 (0.0029)	87.39 (0.0016)	64.90 (0.0978)	63.99 (0.0046)	73.79 (0.0468)	56.32 (0.1014)
✓	✓	✓	**83.93**	**88.38**	**67.67**	**65.88**	**76.01**	**57.58**

strengthening robustness to inter-subject variability. More rigorous and standardized denoising and quality assurance protocols can be incorporated to mitigate the detrimental effects of noise during training. Moreover, adopting methodologies tailored to cross subject distribution shift could further bolster generalization to previously unseen individuals.

V. DISCUSSION

A. Ablation experiments

To investigate the significance and efficacy of each of the key modules within DPMS-Net, we conducted a series of ablation experiments in both the subject dependent and subject independent scenarios, with the results detailed in Table IX. HSCM refers to the horizontal spatial correlation module, while AFCAM+FIPM encompasses the adaptive fine-grained channel attention module and the frequency-related information perception module. NFMM denotes the neuron focus mechanism module.

In the subject dependent scenario, the removal of any individual module markedly affects the performance of DPMS-Net across all three datasets. Specifically, for Dataset 1, the HSCM exerts the most significant influence on DPMS-Net's accuracy. Conversely, in Datasets 2 and 3, the AFCAM+FIPM demonstrates the most pronounced impact on accuracy. In the subject independent scenario, the NFMM emerges as the most influential component affecting DPMS-Net's performance in Datasets 1 and 2, whereas in Dataset 3, the AFCAM+FIPM takes precedence in its contribution. In conclusion, HSCM, AFCAM+FIPM, and NFMM each play pivotal roles within DPMS-Net, substantially enhancing its performance and effectively increasing its upper performance thresholds.

B. The effectiveness of the frequency-related information perception module

To investigate the role of frequency-domain features within the DPMS-Net framework, we systematically removed the frequency-related information perception module and conducted rigorous evaluations in both subject dependent and subject independent scenarios. The experimental results are presented in Fig. 5. In the subject dependent scenario, the absence of frequency-related information perception module resulted in a marked decline in accuracy, of about 2% for both Dataset 1 and Dataset 2, while Dataset 3 also experienced a reduction of over 1%. Similarly, in the subject independent scenario, DPMS-Net demonstrated a decline of more than 1% in accuracy across all three datasets. These findings underscore the notion that frequency-domain features serve as a critical complement to temporal-domain features, thereby playing an indispensable role in enhancing the performance of DPMS-Net.

C. The effectiveness of the temporal variance insight module

The temporal variance insight module markedly enhances the capacity to discern the underlying patterns of EEG signal fluctuations by concentrating on the overall trends within the time series. We further validated this hypothesis, as illustrated in the experimental results presented in Fig. 6. In the subject dependent scenario, the removal of temporal variance insight module exerted the most significant influence on the performance of DPMS-Net in Datasets 1 and 3, with accuracy diminishing by over 1.8% in each case. Even in Dataset 2, where the effects are comparatively less pronounced, a reduction in accuracy exceeding 1% is observed. In the subject independent scenario, the removal of temporal variance insight module resulted in a decline of more than 1% in DPMS-Net's performance on Dataset 2; while the impact on Datasets 1 and 3 is relatively modest, there are still discernible decreases in accuracy. These experimental findings underscore the critical importance of focusing on the overall trends within the time series, which significantly influences the MI decoding accuracy.

D. The effectiveness of dynamic convolution

Dynamic convolution possesses the ability to dynamically adjust the weights of multiple sets of convolution kernels, thereby facilitating a more precise alignment with the distinctive features of each input sample. To ascertain its efficacy, we substituted the dynamic convolution in DPMS-Net with traditional two-dimensional convolution and conducted a series of experiments, the results of which are shown in Fig. 7. Evidently, this substitution resulted in a marked deterioration of DPMS-Net's performance across all three datasets. Notably, the most significant decline is observed in Dataset 2, where performance diminished by 2.44% in the subject dependent scenario and 1.89% in the subject independent scenario. In Dataset 1, DPMS-Net likewise demonstrated a decline, with reductions of 1.91% and 0.93% across the two respective scenarios. Although the performance decrement in Dataset 3 is less pronounced than in Datasets 1 and 2, it still exhibited reductions of 0.86% and 0.98% in the two scenarios. These experimental findings underscore the critical importance of dynamic convolution in enhancing the classification performance of DPMS-Net across diverse datasets.

E. The complementary relationship between the horizontal spatial correlation module and the adaptive fine-grained channel attention module

The horizontal spatial correlation module (HSCM) is devised

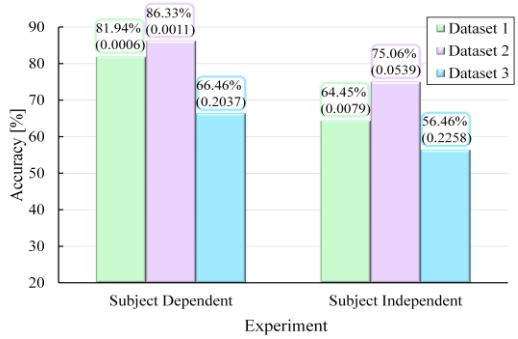


Fig. 5. The experimental results of DPMS-Net following the removal of the frequency-related information perception module across three datasets (The value in parentheses is the p value).

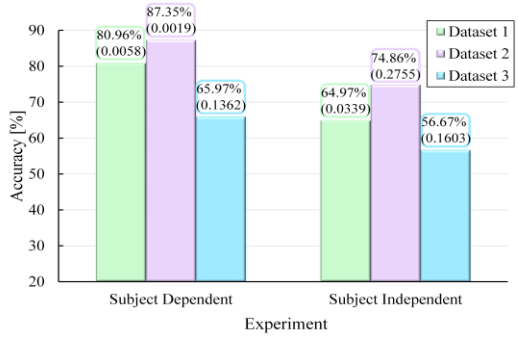


Fig. 6. The experimental results of DPMS-Net following the removal of the temporal variance insight module across three datasets (The value in parentheses is the p value).

to capture local spatial dependencies among adjacent electrodes. The adaptive fine-grained channel attention module (AFCAM) simultaneously extracts personalized local features while establishing global patterns across individuals, thereby significantly strengthening feature extraction capacity. To assess whether these two modules introduce informational redundancy, we conducted a dedicated set of validation experiments, as summarized in Table X.

As shown, removing both HSCM and AFCAM leads to a consistent performance degradation exceeding 2.4% across all three datasets; notably, in the subject dependent setting on Dataset 1, the drop surpasses 4.2%. Incorporating AFCAM into DPMS-Net yields significant improvements on all three datasets. Introducing HSCM alone also confers measurable gains, although these improvements are less pronounced than those achieved by AFCAM in isolation. When both HSCM and AFCAM are jointly integrated, DPMS-Net attains a further substantial boost in performance, indicating their highly complementarity and synergistic effect.

F. The selection of dynamic convolution parameters

In DPMS-Net, we set the parameter compression ratio in the dynamic convolution to 0.25 and the number of convolutional kernels K to 4. To substantiate the rationality of these hyperparameter choices, we conducted a set of corresponding experiments. The results are summarized in Table XI and Table XII. Table XI delineates the influence of the number of kernels in the dynamic convolution on DPMS-Net performance. With K=4, DPMS-Net attains the best performance across both

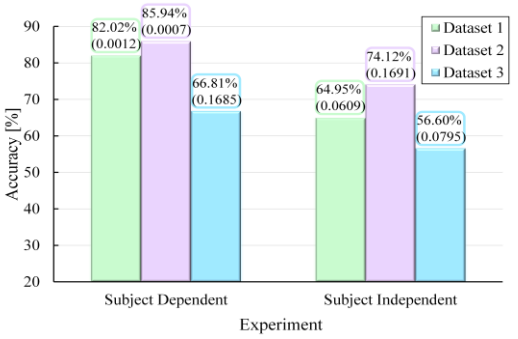


Fig. 7. The experimental results of DPMS-Net following the removal of the dynamic convolution across three datasets (The value in parentheses is the p value).

subject dependent and subject independent scenarios on all three datasets, with the sole exception of the subject independent setting on Dataset 2. Deviating from this configuration, either by increasing or decreasing K, leads to performance deterioration to varying degrees, indicating that K=4 yields a robust and well calibrated trade-off between representational expressiveness and model complexity. Table XII reports the effect of the compression ratio on DPMS-Net performance. When the compression ratio is increased to 0.5, a discernible reduction in performance is observed. Further departures from the selected value, either lowering it to 0.125 or raising it to 0.75, result in a more pronounced degradation. A plausible explanation is that an overly small compression ratio constrains the model's capacity to encode global information, thereby compromising feature discriminability, whereas an excessively large ratio increases parameter redundancy and computational burden, while also exacerbating overfitting susceptibility. These findings lend support to the chosen compression ratio of 0.25.

G. Real online inference speed test

To investigate the inference speed of DPMS-Net in real-world application scenarios, we integrated it into a binary MI-based online system (left vs. right hand). The system utilized 16 electrodes, specifically FC5, FC1, FCZ, FC2, FC6, C5, C3, C1, CZ, C2, C4, C6, CP5, CP1, CP2, and CP6, to acquire MI signals at a sampling rate of 1000 Hz. Five healthy subjects (3 males and 2 females, aged 22-29 years) were recruited for the experiment. Each subject participated in four sessions, with each session comprising 40 trials (20 trials each for left and right hand MI). Each trial lasted 9 seconds: from 0 to 2 s, a fixation cross appeared in the center of the screen to cue the subjects to prepare; from 2 to 6 s, a left or right arrow was randomly presented, instructing the subjects to perform the corresponding MI task; finally, from 6 to 9 s, the word "Rest" was displayed for relaxation. The experimental paradigm is depicted in Fig. 8. Data from the first three sessions were utilized for training DPMS-Net, while the final session was dedicated to evaluating the model in online testing environment, as illustrated in Fig. 9. The preprocessing procedures for both training and online testing data were consistent with those applied to Datasets 1, 2, and 3. The online experimental results, including the maximum, minimum, and average inference

TABLE X
THE EXPERIMENTS ON THE IMPACT OF HORIZONTAL SPATIAL CORRELATION MODULE AND ADAPTIVE FINE-GRAINED CHANNEL ATTENTION MODULE OVER THE THREE DATASETS.

Datasets	HSCM	AFCAM	Subject dependent (p value)	Subject independent (p value)
Dataset 1	×	×	79.71 (5.62e-5)	63.41 (0.0014)
	×	✓	81.15 (0.0003)	65.08 (0.0059)
	✓	×	80.17 (0.0001)	63.91 (0.0019)
	✓	✓	**83.93**	**65.88**
Dataset 2	×	×	85.55 (0.0012)	73.19 (0.0008)
	×	✓	86.85 (0.0037)	74.76 (0.0390)
	✓	×	85.80 (0.0018)	73.91 (0.0010)
	✓	✓	**88.38**	**76.01**
Dataset 3	×	×	64.79 (0.0492)	54.79 (0.0035)
	×	✓	66.67 (0.3034)	56.53 (0.0307)
	✓	×	65.42 (0.1736)	55.28 (0.0174)
	✓	✓	**67.67**	**57.58**

TABLE XI
THE EXPERIMENTS ON THE IMPACT OF DIFFERENT CONVOLUTION KERNELS WITHIN DYNAMIC CONVOLUTION OVER THE THREE DATASETS.

Datasets	K	Subject dependent (p value)	Subject independent (p value)
Dataset 1	2	83.47 (0.0418)	65.41 (0.0391)
	3	83.68 (0.1294)	65.72 (0.0627)
	4	**83.93**	**65.88**
	5	83.74 (0.0825)	65.64 (0.1098)
	6	83.56 (0.0353)	65.57 (0.0863)
Dataset 2	2	87.53 (0.0286)	75.72 (0.0497)
	3	87.91 (0.0364)	75.90 (0.0913)
	4	**88.38**	76.01
	5	88.02 (0.0953)	**76.04** (0.2409)
	6	87.89 (0.1081)	75.83 (0.0985)
Dataset 3	2	67.08 (0.0289)	57.25 (0.1054)
	3	67.42 (0.0852)	57.33 (0.1965)
	4	**67.67**	**57.58**
	5	67.33 (0.0641)	57.08 (0.0768)
	6	67.17 (0.0301)	56.92 (0.0479)

latencies per trial, are summarized in Table XIII. The mobile platform employed for this study is a laptop configured with AMD Ryzen 7 5800H CPU, NVIDIA GeForce RTX 3070 Laptop GPU, and 32 GB of RAM.

The results indicate that only one subject's online accuracy fell below 80%, with the highest reaching 92.50%. Despite individual variations in accuracy, DPMS-Net demonstrated remarkable stability in inference efficiency. Experimental results show that the processing time exhibits minimal fluctuation, with the peak latency of 0.1139 s, the minimum of 0.1071 s, and the overall average is 0.10872 s These findings validate the robustness of the model's decoding process and its capacity for stable operation on mobile platforms.

Compared to the dynamic thresholding strategies utilized in [68] and the multi-modal fusion approaches in [69], DPMS-Net demonstrates superior decoding fidelity (achieving a peak accuracy of 92.5%) and enhanced inference stability within a unimodal MI context. Notably, the average inference latency of 0.1087 s, characterized by negligible fluctuation, underscores the model's robustness and its reliability for real-time deployment on mobile platforms. While this study realizes

TABLE XII
THE EXPERIMENTS ON THE IMPACT OF DIFFERENT COMPRESSION RATE WITHIN DYNAMIC CONVOLUTION OVER THE THREE DATASETS.

Datasets	Compression rate	Subject dependent (p value)	Subject independent (p value)
Dataset 1	0.125	82.72 (0.0025)	64.39 (0.0219)
	0.25	**83.93**	**65.88**
	0.5	83.10 (0.0764)	64.91 (0.0471)
	0.75	82.60 (0.0483)	64.41 (0.0328)
Dataset 2	0.125	86.98 (0.0084)	74.60 (0.0029)
	0.25	**88.38**	**76.01**
	0.5	87.85 (0.0133)	75.51 (0.0765)
	0.75	87.06 (0.0962)	74.69 (0.0042)
Dataset 3	0.125	66.25 (0.0639)	55.90 (0.0267)
	0.25	**67.67**	**57.58**
	0.5	66.46 (0.1096)	56.46 (0.1253)
	0.75	65.83 (0.0358)	55.69 (0.0084)

Fig. 8. The time configuration for each trial.

Fig. 9. Online testing scenarios for DPMS-Net.

TABLE XIII
THE ONLINE TEST ACCURACY AND INFERENCE TIME OF DPMS-NET.

Subject	Online test accuracy	Maximum inference time (s)	Minimum inference time (s)	Average inference time (s)
1	77.50	0.1135	0.1078	0.1089
2	92.50	0.1124	0.1071	0.1085
3	87.50	0.1120	0.1076	0.1087
4	82.50	0.1125	0.1069	0.1082
5	85.00	0.1139	0.1081	0.1093

high-precision online decoding, the current experimental paradigm is constrained by a fixed-duration task structure, lacking the adaptive thresholding mechanisms employed in [68] to further optimize the information transfer rate. Furthermore, as our validation cohort primarily consists of healthy volunteers, the clinical adaptability and long-term therapeutic efficacy of the model remain subjects for future longitudinal investigation, particularly when compared to the clinical focus on stroke rehabilitation presented in [70].

VI. CONCLUSION

This paper introduces a deep neural network-powered multifaceted strategy model (DPMS-Net) for the decoding of MI. The fundamental contribution of DPMS-Net is its ability to adaptively refine feature extraction for each input sample through dynamic convolution and multi-dimensional perception modules. By amalgamating local and global spatial

features and discerning temporal fluctuations, the model overcomes the information loss prevalent in existing SoTA methods. Additionally, the multidimensional temporal perception module serves to mitigate the risk of temporal information loss, while the temporal variance insight module is employed to discern patterns in the fluctuations of EEG signals. To comprehensively extract both temporal and frequency domain features from EEG signals, we have incorporated the adaptive fine-grained channel attention module along with the frequency-related information perception module within the architecture of DPMS-Net. Finally, the neuron focus mechanism module adjusts the weights of individual neurons. We conducted both subject dependent and subject independent evaluations of DPMS-Net across two public datasets and one laboratory self-collected dataset, with experimental outcomes illustrating that DPMS-Net surpasses other SoTA methods. In conclusion, DPMS-Net significantly enhances robustness and classification accuracy across diverse scenarios through an array of sophisticated local and global feature extraction techniques, thereby demonstrating its efficacy and broad applicability in various scenarios. In future, we intend to further enhance the DPMS-Net architecture, such as refining dynamic convolution. Moreover, we will focus on reducing the computational complexity of DPMS-Net to better align with the demands of real-world BCI systems.

VII. Ethics Statement

The Ethics Committee of East China University of Science and Technology reviewed and formally approved the research involving human subjects (Document number: ECUST-2024-146). In accordance with the Declaration of Helsinki, all subjects have signed written consent documents after being fully informed of the study's purpose and potential risks. For any data and experimental results that involve identifiable characteristics of individuals, specific authorization and written informed consent from the subjects have been obtained.

References

[1] M. Zhang, Z. Luo, L. Xie, T. Liu, Y. Yan, D. Yao, S. Zhao and E. Yin, "Multimodal vigilance estimation with modality-pairwise contrastive loss," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 4, pp. 1139-1150, 2024.

[2] W. Ko, S. Jeong, S.-K. Song and H.-I. Suk, "EEG-Oriented Self-Supervised Learning With Triple Information Pathways Network," *IEEE Trans. Cybern.*, vol. 54, no. 11, pp. 6495-6508, 2024.

[3] 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, 2021.

[4] J. Jin, R. Xu, I. Daly, X. Zhao, X. W ang, and A. Cichocki, "MOCNN: A multi-scale deep convolutional neural network for ERP-based brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 54, no. 9, pp. 5565-5576, 2024.

[5] X. He, B. Z. Allison, K. Qin; W. Liang, X. Wang, A. Cichocki and J. Jin, "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, 2024.

[6] 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, 2023.

[7] 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. Inf.*, vol. 27, no. 8, pp. 3818-3829, 2023.

[8] 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, 2015.

[9] 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 Networks*, vol. 118, pp. 262-270, 2019.

[10] 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, 2016.

[11] F. Qi, W. Wu, Z.L. Yu, Z. Gu, Z. Wen, T. Yu and Y. Li "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558-567, 2021.

[12] Z. Wang, L. Yang, M. Wang, Y. Zhou, L. Chen, B. Gu, S. Liu, M. Xu, F. He and D. Ming, "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.

[13] 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 Networks Learn. Syst.*, vol. 32, no. 11, pp. 4814-4825, 2021.

[14] J.-H. Cho, J.-H. Jeong and S.-W. Lee, "NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13279-13292, 2022.

[15] X. Xu, J. Chen, Q. He, Y. Wu and Y. Zhou, "EEG emotion recognition with uncertainty-aware contrastive learning and frequency-aware self-attention," *IEEE Trans. Cybern.*, doi: 10.1109/TCYB.2026.3665442.

[16] C. L. Philip Chen, B. Chen and T. Zhang, "AdamGraph: Adaptive attention-modulated graph network for EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2038-2051, 2025.

[17] 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, 2015.

[18] 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.

[19] 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, 2014.

[20] 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, 105530, 2024.

[21] X. Yu and X. Yu, "Brain-controlled wheeled mobile robots: A framework combining probabilistic brain-computer interface and model predictive control," *IEEE Trans. Cybern.*, vol. 55, no. 9, pp. 4311-4321, 2025.

[22] 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, 2017.

[23] 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, 2022.

[24] 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, 2024.

[25] 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.

[26] B. Wang, C.M. Wong, Z. Kang, F. Liu, C. Shui, F. Wan and C.L. Phiilip Chen, "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 5008-5020, 2021.

[27] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K. -r. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process Mag.*, vol. 25, no. 1, pp. 41-56, 2008.

[28] KK. Ang, ZY. Chin, C. Wang, C. Guan and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, vol. 6, 21002, 2012.

[29] 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, 104397, 2023.

[30]P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Trans. Instrum. Meas.*, vol. 70, no. 4002709, pp. 1-9, 2021.

[31]Q. Huang, Z. Cai and T. Luo, "Multiple regularized knowledge transfer learning for cross-subject motor imagery electroencephalogram classification," *Eng. Appl. Artif. Intell.*, vol. 165, 113397, 2026.

[32]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, 2023.

[33]W. Dang, D. Lv, M. Tang, X. Sun, Y. Liu, C. Grebogi and Z. Gao, "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, 2024.

[34]B. Lu, F. Wang, J. Chen, G. Wen, C. Hua and R. Fu, "Dynamic Hierarchical Convolutional Attention Network for Recognizing Motor Imagery Intention," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2202-2212, 2025.

[35]R. Mane, N. Robinson, A. P. Vinod, S.-W. Lee, and C. Guan, "A multiview CNN with novel variance layer for motor imagery brain computer interface," *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 2950-2953, 2020.

[36]X. Wang, W. Yang, W. Qi, Y. Wan, X.. Ma and W. Wang, "STaRNet: A spatio-temporal and Riemannian network for high-performance motor imagery decoding," *Neural Networks*, vol. 178, 106471, 2024.

[37]X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang and L. Sun, "A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164-2177, 2019.

[38]S. Ke, B. Yang, Y. Qin, F. Rong, J. Zhang and Y. Zheng, "FACT-Net: A frequency adapter CNN with temporal-periodicity inception for fast and accurate MI-EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 4131-4142, 2024.

[39]H. Altaheri, G. Muhammad and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Inf.*, vol. 19, no. 2, pp. 2249-2258, 2023.

[40]R. Liu, Y. Hu, J. Wu, K-C Wong, Z-A Huang, Y-A Huang and K. C. Tan, "Dynamic Graph Representation Learning for Spatio-Temporal Neuroimaging Analysis," *IEEE Trans. Cybern.*, vol. 55, no. 3, pp. 1121-1134, 2025.

[41]J. Fumanal-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernández, J. A. Sanz and H. Bustince, "Motor-imagery-based brain-computer interface using signal derivation and aggregation functions," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7944-7955, 2022.

[42]C. L. Philip Chen, B. Chen and T. Zhang, "AdamGraph: Adaptive Attention-Modulated Graph Network for EEG Emotion Recognition," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2038-2051, 2025.

[43]S. Lin, C. Fan, D. Han, Z. Jia, Y. Peng and S. Kwong, "HATNet: EEG-Based Hybrid Attention Transfer Learning Network for Train Driver State Detection," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2437-2450, 2025.

[44]X. Niu, N. Lu, J. Kang and Z. Cu, "Knowledge-driven feature component interpretable network for motor imagery classification," *J. Neural Eng.*, vol. 19, 016032, 2022.

[45]X. Niu, N. Lu, R. Yan and H. Luo, "Model and data dual-driven double-point observation network for ultra-short MI EEG classification," *IEEE J. Biomed. Health. Inf.*, vol. 28, no. 6, pp. 3434-3445, 2024.

[46]Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan and Z. Liu, "Dynamic convolution: Attention over convolution kernels," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 11030-11039, 2020.

[47]H. Sun, Y. Wen, H. Feng, Y. Zheng, Q. Mei, D. Ren and M. Yu, "Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing," *Neural Networks*, vol. 176, 106314, 2024.

[48]L. Yang, R. Zhang, L. Li and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," *Proc. 38th Int. Conf. Mach. Learn. (PMLR)*, vol. 139, 11863-11874, 2021.

[49]X. Liang, L. Wu, J. Li, Y. Wang, Q. Meng, T. Qin, W. Chen, M. Zhang, T. Liu, "R-Drop: Regularized dropout for neural networks," *Proc. Neural Inf. Process. Syst.*, vol. 34, pp. 10890-10905, 2021.

[50]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, 056013, 2018.

[51]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," *2020 IEEE Int. Conf. Syst. Man Cybern. (SMC)*, Toronto, ON, Canada, pp. 2958-2965, 2020.

[52]H. Zhao, Q. Zheng, K. Ma, H. Li and Y. Zheng, "Deep representation-based domain adaptation for nonstationary EEG classification," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 535-545, 2021.

[53]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.

[54]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.

[55]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.

[56]K. Singh, N. Singha, G. Jaswal and S. Bhalaik, "A novel CNN with sliding window technique for enhanced classification of MI-EEG sensor data," *IEEE Sens. J.*, vol. 25, no. 3, pp. 4777-4786, 2025.

[57]C. Han, C. Liu, J. Wang, Y. Wang, C. Cai and D. Qian, "A spatial-spectral and temporal dual prototype network for motor imagery brain-computer interface," *Knowledge-Based Syst.*, vol. 315, 113315, 2025.

[58]W. Tao, Z. Wang, C. Wong, Z. Jia, C. Li, X. Chen, C. L. P. Chen and F. Wan, "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, 2024.

[59]Q. Ai, Y. Liu, Q. Liu, L. Ma and K Chen, "Holographic convolutional attention neural network for motor imagery decoding based on EEG temporal-spatial frequency features," *Biomed. Signal Process. Control*, vol. 104, 107526, 2025.

[60]Z. Wang, Y. Si, Z. Wang, T. Zhou, T. Xu and H. Hu, "NexusNet: Lightweight graph modeling for motor imagery-based brain-computer interfaces," *IEEE Internet Things J.*, vol. 12, no. 15, pp. 31408-31419, 2025.

[61]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, 2023.

[62]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.

[63]R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391-5420, 2017.

[64]X. Ma, W. Chen, Z. Pei, Y. Zhang and J. Chen, "Attention-based convolutional neural network with multi-modal temporal information fusion for motor imagery EEG decoding," *Comput. Biol. Med.*, vol.175, 108504, 2024.

[65]P. Autthasan, R. Chaisaen, T. Sudhawiyangkul, P. Rangpong, S. Kiatthaveephong, N. Dilokthanakul, G. Bhakdisongkhram, H. Phan, C. Guan and T. Wilaiprasitporn, "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, 2022.

[66]J. Shin and W. Chung, "Multi-band CNN with band-dependent kernels and amalgamated cross entropy loss for motor imagery classification," *IEEE J. Biomed. Health. Inf.*, vol. 27, no. 9, pp. 4466-4477, 2023.

[67]Q. Zhao and W. Zhu, "TMSA-Net: A novel attention mechanism for improved motor imagery EEG signal processing," *Biomed. Signal Process. Control*, vol. 102, 107189, 2025.

[68]K. Wang, Y. Liu, F. Tian, W. Yi, Y. Zhang, T.-P. Jung, M. Xu and D. Ming, "Adaptive neurofeedback training using a virtual reality game enhances motor imagery performance in brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 33, pp. 2956-2966, 2025.

[69]X. Tang, W. Zhu, X. Liu, T. Wang, J. Guan and X. Dang, "Motor imagery brain-computer interaction using different modalities of action observation," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1-11, 2025.

[70]Z. Rao, R. Zhang, S. He, Y. Zhou, Z. Lu, K. Li and Y. Li, "A once-calibration brain-computer interface to enhance convenience for continuous BCI interventions in stroke patients," *IEEE Sens. J.*, vol. 25, no. 2, pp. 3949-3963, 2025.