

# EEG-Channel-Temporal-Spectral-Attention Correlation for Motor Imagery EEG Classification

Wei-Yen Hsu<sup>ID</sup>, Senior Member, IEEE, and Ya-Wen Cheng

**Abstract**—In brain-computer interface (BCI) work, how correctly identifying various features and their corresponding actions from complex Electroencephalography (EEG) signals is a challenging technology. However, most current methods do not consider EEG feature information in spatial, temporal and spectral domains, and the structure of these models cannot effectively extract discriminative features, resulting in limited classification performance. To address this issue, we propose a novel motor-imagery EEG discrimination method, namely wavelet-based temporal-spectral-attention correlation coefficient (WTS-CC), to simultaneously consider the features and their weighting in spatial, EEG-channel, temporal and spectral domains in this study. The initial Temporal Feature Extraction (ITFE) module extracts the initial important temporal features of MI EEG signals. The Deep EEG-Channel-attention (DEC) module is then proposed to automatically adjust the weight of each EEG channel according to its importance, thereby effectively enhancing more important EEG channels and suppressing less important EEG channels. Next, the Wavelet-based Temporal-Spectral-attention (WTS) module is proposed to obtain more significant discriminative features between different MI tasks by weighting features on two-dimensional time-frequency maps. Finally, a simple discrimination module is used for MI EEG discrimination. The experimental results indicate that the proposed WTS-CC method can achieve promising discrimination performance that outperforms the state-of-the-art methods in terms of classification accuracy, Kappa coefficient, F1 score, and AUC on three public datasets.

**Index Terms**—Brain-computer interface (BCI), motor-imagery electroencephalography (MI EEG), EEG-channel attention, temporal-spectral attention, wavelet transform, correlation coefficient.

## I. INTRODUCTION

BCI is an advanced technology that establishes a system that does not need to go through peripheral nerves and muscles and does not need any body part to move [1], [2], [3].

Manuscript received 25 October 2022; revised 30 January 2023; accepted 5 March 2023. Date of publication 10 March 2023; date of current version 15 March 2023. This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST110-2221-E-194-027-MY3 and Grant MOST111-2410-H-194-038-MY3. (Corresponding author: Wei-Yen Hsu.)

Wei-Yen Hsu is with the Center for Innovative Research on Aging Society (CIRAS), Department of Information Management, Advanced Institute of Manufacturing with High-Tech Innovations, National Chung Cheng University, Chiayi 62102, Taiwan (e-mail: shenswy@gmail.com).

Ya-Wen Cheng is with the Department of Information Management, National Chung Cheng University, Chiayi 62102, Taiwan (e-mail: yawen0213@gmail.com).

Digital Object Identifier 10.1109/TNSRE.2023.3255233

Analyzing EEG can infer the subject's mind, enabling people to directly communicate with external devices through brain consciousness and be a communication channel [4], [5], [6]. With the rapid development of technology, MI-EEG is widely used in BCI research. Suppose these neural activities based on MI-EEG can be mastered. In that case, it is possible to use "imagination" to control devices such as computers or robotic arms, helping improve patients' quality of life with physical disabilities or neuromuscular degeneration [7], [8], [9]. However, physiological signals can produce different results due to individual differences in subjects, coupled with the instability of brain activity and low signal-to-noise ratio (SNR) [10], [11]. These factors may limit the performance of classification. Therefore, how to extract useful features from complex EEG to improve the classification accuracy of MI-based BCI systems is still an important challenge [12], [13].

In various MI-EEG applications, many machine learning algorithms have been developed. For example, Common Spatial Modeling (CSP) is the most widely used method for identifying MI-EEG signals, which builds spatial filters and effectively extracts feature information. Novi et al. [14] proposed Sub-Band Common Spatial Pattern (SBCSP), which extracts CSP features for each frequency band separately, and uses Linear Discriminant Analysis (LDA) to extract features from multiple frequency bands in a fractional fusion manner for classification. Ang et al. [15] proposed Filter Bank Common Spatial Pattern (FBCSP), which divides a wide frequency band into multiple smaller frequency bands, and calculates the CSP features of these frequency bands to select the most discriminative features. However, these methods focus on the energy features of EEG and cannot obtain features with high discriminative power from the original EEG signal, thus limiting the decoding performance of MI-EEG [16]. The advantage of deep learning is that the model can automatically extract features. Compared with traditional machine learning methods, its nonlinear characteristics can learn more feature information, which solves the above problems to a certain extent. Chen et al. [17] proposed a filter-bank spatial filtering and temporal-spatial convolutional neural network (FBSF-TSCNN) and designed the FBSF module based on the traditional FBCSP algorithm to further explore the temporal information of EEG. Zhang et al. [18] designed a hybrid network based on spatial and temporal feature extraction. However, these methods still have some limitations in MI-EEG discrimination. First, incomplete temporal or spatial analysis of end-to-end CNN frameworks may destroy EEG non-stationarity. Second, most methods do not simultaneously

consider the characteristics of EEG in the frequency, time, and space domains and lack deeper information.

In the application of BCI, the performance of EEG classification depends to a large extent on feature extraction and representation of MI data. An attention mechanism named Squeeze-and-Excitation (SE) [19], [28] was proposed to emphasize channel-wise features. Inspired by its advantages, this research exploits the performance of SE to predict the importance of each EEG channel, thereby enhancing the extracted features directly. Studies have shown that MI tasks induce neural activity in corresponding regions of the brain, resulting in enhancement or inhibition of mu and beta frequencies in the sensorimotor cortex, a phenomenon known as event-related synchronization (ERS) and event-related desynchronization (ERD). If the features are expressed only in the time domain, ignoring the information in the frequency domain, the classification accuracy may degrade. In order to extract more important and valuable features, it is necessary to span the representation of features into the two-dimensional time-frequency domain. Wavelet transform constructs a time-frequency domain signal with good time and frequency localization. The frequency components contained in the signal and their corresponding time segments can be observed to identify the occurrence and localization of the ERD/ERS phenomenon.

In addition to feature extraction, another important issue is how to design a method that can effectively classify MI-EEG. Classification methods currently widely used in BCI research: linear discriminant analysis (LDA), support vector machine (SVM), neural networks, etc. LDA provides good performance for classification under the assumption that the sample covariance matrices between different classes are similar. However, there is often potential noise interference in EEG signals, so overfitting problems may occur, leading to the degradation of the classification performance [20], [21], [22]. To solve this problem, more and more regularization techniques and classification algorithms are applied to classify MI tasks. Among them, SVM is a well-known classification method that maximizes the interval to achieve better generalization ability. Combined with the CSP method, SVM provides state-of-the-art performance in related research on MI task classification [23], [24], [25]. Another method developed in recent years is sparse representation-based classification (SRC), which judges the classification results by representing the CSP features of test samples as a linear combination of training samples and then detecting the minimum residual norm [26]. SRC has been successfully applied to the study of BCI and has also been shown to outperform traditional LDA in SMR classification [27]. Although these methods are proven effective under certain conditions, their classification accuracy depends on the manually extracted features. Since EEG signals have individual differences in subjects, a small training set, and a low signal-to-noise ratio, these problems will affect the classification accuracy of MI-EEG, resulting in poor classification performance. In this study, the correlation coefficient, which is relatively simple and can effectively demonstrate the performance of the proposed method, is used to compare the feature similarity in different MI tasks to achieve the effect of accurate classification.

Therefore, in this study, we propose a novel wavelet-based temporal-spectral-attention correlation coefficient (WTS-CC) for MI EEG discrimination by simultaneously taking the features and their weighting into account in spatial, EEG-channel, temporal and spectral domains. The iTFE module extracts preliminary important temporal features of original MI-EEG signals by means of convolution operations of different sizes. The DEC module is then proposed to automatically adjust the weight of each EEG channel according to its importance, thereby effectively enhancing more important EEG channels and suppressing less important EEG channels at the same time. Next, to effectively extract the temporal-spectral features, the WTS module is proposed to obtain more significant discriminative temporal-spectral-attention features on two-dimensional time-frequency maps between different MI tasks. Finally, a simple discrimination module is used for MI EEG discrimination. The contributions of this study are summarized as follows:

- 1) We propose a novel wavelet-based temporal-spectral-attention correlation coefficient (WTS-CC) to achieve more accurate MI EEG discrimination by simultaneously considering the features and their weighting in spatial, EEG-channel, temporal, and spectral domains.
- 2) The Deep EEG-channel-attention (DEC) module is proposed to automatically adjust the weight of each EEG channel according to its importance, thereby effectively enhancing more important EEG channels and suppressing less important EEG channels.
- 3) The wavelet-based temporal-spectral-attention (WTS) module is proposed to obtain more significant discriminative temporal-spectral-attention features on two-dimensional time-frequency maps between different MI tasks.
- 4) The experimental results indicate that the proposed WTS-CC method achieves promising performance in comparison with the state-of-the-art methods in terms of classification accuracy, Kappa coefficient, F1 score and AUC on three public datasets.

## II. METHODOLOGY

We provide a detailed explanation of the proposed WTS-CC and introduce each module of our method, including the iTFE module, DEC module, WTS module and Discrimination module.

### A. Initial Temporal Feature Extraction Module (iTFE Module)

In MI EEG classification, extracting more feature information from EEG signals is important to improving classification accuracy. EEG signals contain a large number of temporal, spatial and spectral features that are difficult to define manually [9]. Therefore, the iTFE module extracts preliminary features by means of convolution operations directly from raw EEG signals. Fig. 1(a) shows the structure of the iTFE module, including the shape transformation layer and the temporal convolution layer. First, the raw EEG signals are converted from the raw time representation to a 2-dimensional form. In previous studies, most models use a single-size convolution

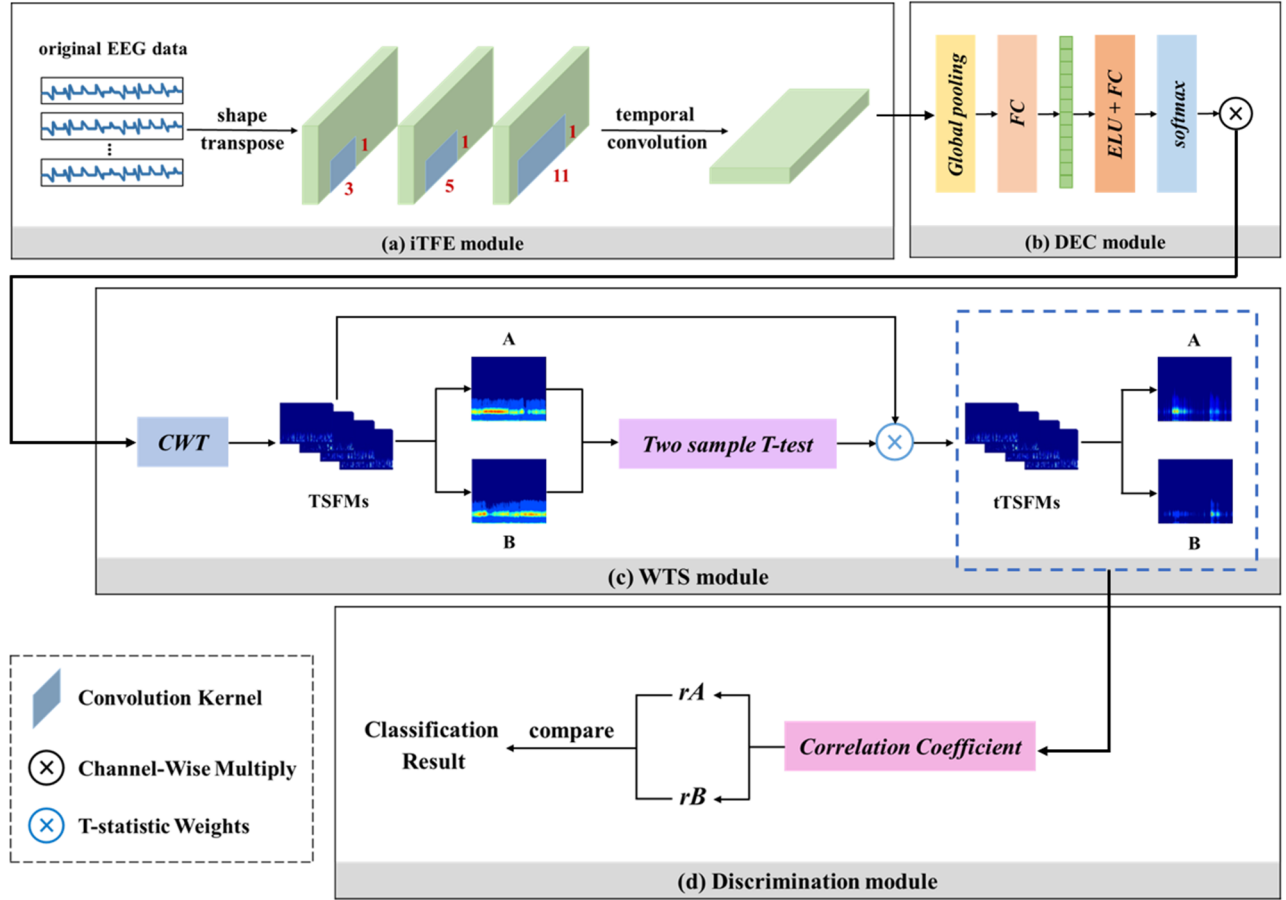


Fig. 1. Network architecture of the proposed WTS-CC method.

kernel for convolution operations, limiting the performance of partial feature extraction and classification of the model. Therefore, we design three convolution kernels of different sizes,  $1 \times 3$ ,  $1 \times 5$ , and  $1 \times 11$ , respectively, and perform convolution operations over time to extract richer feature information of different sizes.

### B. Deep EEG-Channel-Attention Module (DEC Module)

Although extracting richer features helps classify MI tasks, these feature messages are usually mixed with many irrelevant or redundant messages. Considering that the quality of feature extraction is an important key to the success of EEG classification, a variant of the squeeze excitation (SE) mechanism is proposed to emphasize more discriminative feature information in EEG channels. More specifically, the Deep EEG-channel-attention (DEC) module is proposed to automatically adjust the weight of each EEG channel according to its importance, thereby effectively enhancing more important EEG channels and suppressing less important EEG channels. Fig. 1(b) shows the structure of the DEC module. SEC recalibrates the feature for the input feature map through the ‘‘Squeeze’’ operation. This part performs Global Average Pooling on the input feature map of each EEG channel so that the two-dimensional feature map of each channel is compressed into a global feature to be represented to achieve the purpose of channel statistical features. Formally, the statistical output computed by global

average pooling is defined as:

$$Z_C = F_{sq}(u_C) = \frac{1}{T} \sum_{t=1}^T u_C(i, c, t) \quad (1)$$

where  $u_C \in R^{i \times c \times t}$  represents the input feature map,  $T$  is the time, and  $Z_C$  is the compressed result. After extracting channel information from the ‘‘Squeeze’’ operation, the ‘‘Excitation’’ operation is then used to predict the importance of each EEG channel. Here, two fully connected layers (FC) and nonlinear functions, ELU and Softmax, are used. Then, the generated channel information is applied to the input features by weighting between the learned channel weights and feature maps. The output of the SE operation is represented as:

$$s = F_{se}(Z_C, W) = \sigma(W_2 \delta(W_1 Z_C)) \quad (2)$$

where  $W_1$  represents the parameters of the first fully connected layer (squeeze),  $W_2$  represents the parameters of the second fully connected layer (restoration),  $\sigma$  represents the Softmax function, and  $\delta$  represents the ELU function. Compared with the original SE operation [19], we choose to use the Softmax function instead of the Sigmoid function because Softmax preserves the size information between the input vectors. Finally, the weight output by the Excitation operation is multiplied back to the original features to further strengthen the more important features as the final output of the SE module.

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (3)$$

### C. Wavelet-Based Temporal-Spectral-Attention Module (WTS Module)

Electroencephalography (EEG) signals have the characteristics of time series, and their frequency components will change over time. If only temporal features are considered, and frequency information is ignored, the classification accuracy may be degraded. Wavelet transform is a typical time-frequency domain method, which can effectively extract features from transient physiological signals through the local transformation of time and frequency. To further explore the information on time and frequency domains, we propose a novel WTS module to obtain more significant discriminative features between different MI tasks by weighting features on two-dimensional time-frequency maps. Fig. 1(c) shows the structure of the WTS module, including continuous wavelet transform and independent sample T statistics. In the feature extraction process, the EEG signal output from the above steps is first converted into the time-frequency domain by CWT, then drawn into two-dimensional temporal-spectral feature maps (TSFMs). In this way, we can easily observe the change in energy magnitude at a specific frequency or time and estimate the relative position of the ERS/ERD that is beneficial for MI analysis by observing the subject's multiple trials. The wavelet transform is as follows:

$$W_s^C(a, b) = \int f_s^C(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \quad (4)$$

where  $\frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$  is the wavelet basis function  $\psi(t)$  generated by the mother wavelet by stretching  $a$  and translating  $b$  coefficients.  $f_s^C(t)$  represents the input electroencephalography (EEG) signal,  $W_s^C(a, b)$  represents the result after continuous wavelet transform,  $s$  represents the motor imagery state of different tasks, and  $C$  represents the channel. In addition, when selecting the mother wavelet, when analyzing the non-stationarity of time series, it is generally hoped to obtain smooth and continuous wavelet amplitude, so a non-orthogonal wave function is suitable. In this study, we use the Morlet wavelet, a single-frequency sinusoidal function under a Gaussian envelope without a scaling function, and a non-orthogonal decomposition that makes the energy more concentrated [29]. After completing the CWT, the EEG signals are converted into representations of TSFMs. In order to capture more critical features between different MI tasks, we first merged the EEG channels of the same type of MI tasks and used the independent sample T statistic to judge the degree of feature difference between the two groups of different MI tasks.

The independent sample T statistic is a statistical method used to judge whether there is a significant difference between two data groups. Therefore, in this study, the training data is divided into two groups (group A and group B), and a two-stage T-statistical experiment is conducted. In the first stage, the left hand and right hand are divided into group A, and the feet and tongue are divided into group B. After independent sample T statistics, a set of T statistics will be obtained to judge the characteristic differences between the two groups of MI tasks. When the preliminary results of the first stage are obtained, the second stage divides the MI tasks of the same

group in the first stage into two groups (group A is the left hand, group B is the right hand; group A is the feet, and group B is the tongue), which then leads to the second stage of the T statistics experiment. The detailed procedure of the T statistic will be described below.

After dividing the training data into MI tasks of group A and MI tasks of group B, the average (such as Eq. (5)) and standard deviation (such as Eq. (6)) of the two groups of MI tasks are calculated, respectively,

$$\overline{\pi_s^C(b)} = \frac{1}{N_s} \sum_{n=1}^{N_s} \pi_s^{C,n}(b) \quad (5)$$

$$\sigma^2(b) = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} \left( \pi_s^{C,n} - \overline{\pi_s^C(b)} \right)^2 \quad (6)$$

where  $s$  represents the state of different MI tasks,  $N_s$  represents the total number of trials of different MI tasks, and  $C$  represents the channel. The result is then used to calculate the values of the T statistic,

$$t_{s_1, s_2}^C(b) = \frac{\left| \overline{\pi_{s_1}^C(b)} - \overline{\pi_{s_2}^C(b)} \right|}{\sqrt{\frac{(N_{s_1}-1)\sigma_{s_1}^{2C}(b) + (N_{s_2}-1)\sigma_{s_2}^{2C}(b)}{N_{s_1} + N_{s_2}}} \quad (7)$$

where  $s_1$  and  $s_2$  represent motor imagery in group A and motor imagery in group B, respectively,  $N_{s_1}$  and  $N_{s_2}$  represent the total number of motor imagery trials in group A, and the total number of motor imagery trials in group B, respectively, and  $C$  here is the state after the channel is merged. The obtained T statistic  $t_{s_1, s_2}^C(b)$  provides the degree of difference between the two groups of MI tasks, as well as represents the difference between the two groups of MI tasks at different time points  $b$  and different frequencies. Specifically, the point where  $t_{s_1, s_2}^C(b)$  is a local maximum indicates that there will be the most significant feature difference between the two sets of MI tasks.  $t_{s_1, s_2}^C(b)$  is taken as the weight of the time-frequency feature, and the TSFMs of each trial are multiplied by their corresponding weight to obtain t-statistic-weighted temporal-spectral feature maps (tTSFMs), and the t-statistic-weighted average temporal-spectral feature maps of the MI groups A and B are plotted, as a benchmark for subsequent evaluation of different MI task classifications.

### D. Discrimination Module

After the WTS module is performed, we obtain the average tTSFMs of two different MI tasks (i.e., A and B), which are used as classification benchmarks for these two groups. Unlike the state-of-the-art approaches and deep learning models, the Discrimination module uses the correlation coefficient for MI EEG discrimination by evaluating the correlation between the tTSFMs of each test trial and the average tTSFMs of these two MI tasks. Fig. 1(d) shows the structure of the correlation coefficient-based discrimination module. The correlation coefficient is a statistical indicator used to reflect the closeness of the correlation between the data. This research compared the correlation coefficients between the tTSFMs of each EEG test and the average tTSFMs of the A and B groups, respectively.



The larger the value of the correlation coefficient, the closer the relationship between the data. The correlation coefficient is as follows:

$$r_s = \frac{\sum_{i=1}^{N_s} (s_i - \bar{s})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N_s} (s_i - \bar{s})^2 \sum_{i=1}^{N_s} (y_i - \bar{y})^2}} \quad (8)$$

where  $N_s$  represents the total number of trials,  $s$  represents the t-statistic-weighted average temporal-spectral feature maps for different motor imagery tasks, and  $y$  represents the t-statistic-weighted temporal-spectral feature maps for each trial. The  $r_s$  obtained above represents the degree of correlation between the tTSFMs of a trial and the average temporal-spectral feature maps of the two groups of MI tasks weighted by T statistics.

More specifically, the tTSFMs of a certain experiment were calculated with the average tTSFMs of group A, and the correlation coefficient obtained was  $r_{s1}$ . The correlation coefficient calculated with the average tTSFMs of group B was  $r_{s2}$ . Then,  $r_{s1}$  and  $r_{s2}$  are compared with each other, and the MI task with a greater correlation between the two indicates that the trial is more similar to this type of MI task and is regarded as the predicted classification result. In other words, if  $r_{s1} > r_{s2}$ , the trial is predicted to be a group A motor imagery, and conversely, if  $r_{s2} > r_{s1}$ , the trial is predicted to be a group B motor imagery.

### III. EXPERIMENTAL RESULTS

In the experiments, three public BCI Competition datasets were used to validate the effectiveness of the proposed WTS-CC method for MI EEG discrimination.

#### A. EEG Datasets

In this study, the same time interval  $[-0.5s, 5s]$  (relative to the onset of motor imagery cues) is used for the EEG signals of each trial. Before training and testing, the raw signals are filtered to 8-30Hz by a third-order Butterworth low-pass filter, which can minimize less-correlated artifacts. In addition, the original training set for each subject is divided into ten equal parts (nine for training and one for validation).

1) *BCI Competition IV Dataset 2a*: The BCI IV 2a [24] dataset was derived from EEG signals obtained from a trial of 9 subjects, covering four different MI tasks (left hand, right hand, feet, and tongue). Each subject will have two sessions, each session consists of 6 runs, and each run will have 48 trials, for a total of 288 trials (12 for each MI task, for a total of 72 trials). There are a total of 25 Ag/AgCl electrodes in the experiment, of which 22 electrodes are used to record EEG signals, and the other three electrodes are responsible for recording eye movement signals (not used in the experiment). All collected signals will be processed by a 0.5 to 100Hz bandpass filter and a 50Hz notch filter, and the sampling frequency is 250Hz.

2) *BCI Competition IV Dataset 2b*: The BCI IV 2b [24] dataset was derived from EEG signals obtained from a trial of 9 subjects, including two different MI tasks (left-hand and right-hand). The third training set of the dataset was used, with 80 trials for each MI task for each subject, for a total of 160 trials. In the experiments, three electrodes were used to

record EEG signals (C3, Cz, and C4). All the collected signals were processed by bandpass filters of 0.5 to 100 Hz and notch filters of 50 Hz, and the sampling frequency was 250 Hz.

3) *2020 International BCI Competition Dataset Track#1*: Dataset “Track#1 Few-shot EEG learning” [32] aimed to classify subject-specific MI data using minimal training data based on few-shot learning. The Track#1 dataset came from the EEG signals obtained from experiments conducted by 20 subjects, including two different MI tasks (left hand and right hand). Each subject performed ten trials of each MI task for 20 trials. 62 Ag/AgCl electrodes were used in the experiment, and the sampling frequency was 1000 Hz.

#### B. Evaluation Metrics

MI-EEG data is tested for classification using ten-fold cross-validation in all experiments. The data of each subject is divided into ten equal disjoint subsamples of the same size. One is used as a test sample, and the other nine are used as a training sample. This process is repeated until each aliquot of subsamples is tested (i.e., ten sets of results are obtained), and the average is used as the final average classification rate. In the experiments, we use classification accuracy (ACC) [25], Cohen’s Kappa coefficient (K) [17], F1-score (F1) [27], and area under the curve (AUC) [9] to evaluate the performance of the proposed WTS-CC. In addition, two-way ANOVA and multiple comparison tests are also performed to verify whether the results of different methods are significantly different. Here, Cohen’s kappa coefficient is represented as:

$$\begin{aligned} \text{kappa} &= \frac{\text{Accuracy} - p_e}{1 - p_e} \quad (9) \\ p_e &= \frac{(TP + FN)(TP + FP) + (FP + TN)(FN + TN)}{\text{Total}^2} \quad (10) \end{aligned}$$

where the numerator is the sum of the row and column elements of the confusion matrix, and the denominator represents the square of the sum of all elements in the confusion matrix. That is, the sum of the “products of actual and predicted numbers” for all types, divided by the square of the total number of samples.

#### C. Comparisons With the State-of-the-Arts

To verify the performance of the proposed WTS-CC method, the state-of-the-art models are compared in three public BCI Competition datasets. These compared models are described briefly below,

- 1) Shallow ConvNet [30]: A deep learning model consisting of two convolutional layers and an average pooling layer.
- 2) Deep ConvNet [30]: This model is more complex than Shallow ConvNet and uses three convolutional layers in the temporal dimension.
- 3) CP-MixedNet [31]: This model uses multi-scale EEG features generated from multiple convolutional layers, each layer extracting EEG temporal representations from different scales.

TABLE I

MI CLASSIFICATION ACCURACIES (ACC (%)), COHEN'S KAPPA COEFFICIENT (K), F1-SCORE (F1) AND AUC ON BCI COMPETITION IV DATASET 2A. FOR EACH SUBJECT, THE BEST RESULT IS MARKED IN BOLDFACE

Subject		S01	S02	S03	S04	S05	S06	S07	S08	S09	Avg±Std
Shallow ConvNet [30]	ACC(%)	80.20	54.86	<b>92.01</b>	59.03	66.67	54.17	<b>87.50</b>	79.51	<b>82.29</b>	72.92±14.5
	K	0.736	0.398	<b>0.894</b>	0.454	0.556	0.389	<b>0.833</b>	0.727	<b>0.764</b>	0.639±0.19
	F1	0.800	0.549	<b>0.920</b>	0.590	0.664	0.541	<b>0.875</b>	0.796	0.821	0.728±0.14
	AUC	<b>0.963</b>	0.803	<b>0.992</b>	0.859	0.887	0.803	<b>0.988</b>	<b>0.955</b>	0.956	0.912±0.08
Deep ConvNet [30]	ACC(%)	80.90	52.08	84.72	71.18	70.49	55.56	69.10	81.94	81.94	71.99±11.8
	K	0.745	0.361	0.796	0.616	0.606	0.407	0.588	0.759	0.759	0.627±0.16
	F1	0.809	0.523	0.846	0.712	0.704	0.552	0.685	0.819	0.819	0.719±0.12
	AUC	0.955	0.791	0.981	0.888	0.922	0.825	0.922	0.953	0.951	0.910±0.06
CP-MixedNet [31]	ACC(%)	74.65	53.47	73.26	70.14	67.36	48.96	74.31	72.92	69.44	67.17±9.4
	K	0.622	0.380	0.644	0.602	0.565	0.319	0.657	0.639	0.593	0.562±0.12
	F1	0.744	0.535	0.731	0.694	0.675	0.480	0.742	0.730	0.694	0.670±0.10
	AUC	0.907	0.785	0.916	0.909	0.893	0.793	0.925	0.926	0.903	0.884±0.05
TS-SEFFNet [28]	ACC(%)	82.29	49.79	87.57	71.74	70.83	63.75	82.92	81.53	81.94	74.71±12.0
	K	0.764	0.331	0.834	0.623	0.611	0.517	0.772	0.754	0.759	0.663±0.16
	F1	0.830	0.450	0.899	0.729	0.729	0.652	0.874	0.808	<b>0.838</b>	0.757±0.14
	AUC	0.959	0.748	0.988	0.916	0.921	0.864	0.986	0.952	<b>0.964</b>	0.922±0.08
Proposed Method	ACC(%)	<b>82.67</b>	<b>74.69</b>	86.31	<b>82.79</b>	<b>84.19</b>	<b>79.37</b>	79.38	<b>82.11</b>	81.56	<b>81.45</b> ±3.3
	K	<b>0.769</b>	<b>0.663</b>	0.817	<b>0.770</b>	<b>0.789</b>	<b>0.725</b>	0.725	<b>0.761</b>	0.753	<b>0.752</b> ±0.04
	F1	<b>0.827</b>	<b>0.747</b>	0.863	<b>0.828</b>	<b>0.841</b>	<b>0.793</b>	0.793	<b>0.821</b>	0.815	<b>0.814</b> ±0.03
	AUC	0.948	<b>0.922</b>	0.945	<b>0.926</b>	<b>0.955</b>	<b>0.939</b>	0.946	0.940	0.937	<b>0.939</b> ±0.01

- 4) TS-SEFFNet [28]: This model consists of a deep-temporal convolution block, a multi-spectral convolution block, and a squeeze-and-excitation feature fusion block.
- 5) SS-MEMDBF [33]: The multichannel EEG signals are decomposed into a set of multivariate intrinsic mode functions (MIMFs) to extract the filter range for subject-specific MEMDs.
- 6) MEMDBF-CSP [34]: The MEMDBF method is used as preprocessing, and the CSP function is implemented to enhance the performance of two-class MI BCI tasks.
- 7) TS-TL [35]: Combining SS-MEMDBF preprocessing and TS-TL classification, the sample covariance is used as feature sets to enhance the performance of two-class MI BCI tasks.
- 8) LR-TSTL [36]: The features of MI-EEG are determined by a tangent space method, and the derived features are used for classification as input to a logistic regression model.
- 9) SGRM [20]: The model uses inter-subject information to improve the effect of MI classification through intra-group sparsity and group sparsity constraints.
- 10) clsSRC [22]: The model proposes a clustered-group sparse representation to overcome using only a limited amount of EEG trials.
- 11) clsSRC2 [21]: The model proposes a sparse representation classification scheme that extends current sparse representation schemes by exploiting the group sparsity of relevant features.
- 12) This model is based on power spectral density (PSD) feature extraction and discriminates MI tasks with the SVM classifier.
- 13) This model is derived from the filter bank CSP features using the regularized linear discriminant analysis (RLDA).

Considering deep EEG-channel attention and wavelet-based temporal-spectral attention, the proposed WTS-CC is an effective method for MI-EEG classification. To evaluate the effectiveness, we compare the proposed WTS-CC with the state-of-the-art models. TABLE I lists the comparison results of MI classification accuracy (ACC (%)), Cohen's kappa coefficient (K), F1-score (F1), and AUC on BCI competition IV dataset 2a. The experimental results indicate that the proposed WTS-CC method achieves the best average classification accuracy, Kappa coefficient, F1-score, and AUC, which is better than all state-of-the-art models.

In addition, TABLE II lists the comparison results between the current state-of-the-art methods and ours on BCI competition IV dataset 2a in terms of classification accuracy (ACC (%)) and Cohen's kappa coefficient (K). SS-MEMDBF [33] and MEMDBF-CSP [34] are filtering methods based on decomposition. This decomposition extracts multi-channel EEG signals and transforms them into intrinsic mode functions (IMFs) to reduce the within-subject and subject-specific effects of EEG signals. The average accuracies are 79.93% and 79.19%, respectively, due to the influence of non-stationarity among the subjects. TS-TL [35] and LR-TSTL [36] are transfer learning classification models based on tangent space, with average accuracy rates of 75.52% and 78.95%, respectively. In contrast, our method, which simultaneously considers the features and their weighting in spatial, EEG-channel, temporal, and spectral domains, achieves an average classification accuracy of 81.45% and an average kappa coefficient of 0.752. Therefore, the results indicate that the proposed WTS-CC

TABLE II

MI CLASSIFICATION ACCURACIES (ACC (%)), COHEN'S KAPPA COEFFICIENT (K), F1-SCORE (F1) AND AUC ON BCI COMPETITION IV DATASET 2A. FOR EACH SUBJECT, THE BEST RESULT IS MARKED IN BOLDFACE

Subject		S01	S02	S03	S04	S05	S06	S07	S08	S09	Avg±Std
SS-MEMDBF [33]	ACC(%)	<b>91.49</b>	60.56	94.16	76.72	58.52	68.52	78.57	97.01	93.85	79.93±14.99
	K	<b>0.860</b>	0.240	0.700	0.680	0.360	0.340	0.660	0.75	0.820	0.600±0.23
MEMDBF-CSP [34]	ACC(%)	90.78	57.75	<b>97.08</b>	70.69	61.48	70.37	72.14	<b>97.76</b>	<b>94.62</b>	79.19±15.85
	K	-	-	-	-	-	-	-	-	-	-
TS-TL [35]	ACC(%)	88.65	61.27	91.24	74.14	57.04	69.44	60.00	94.03	83.85	75.52±14.39
	K	-	-	-	-	-	-	-	-	-	-
LR-TSTL [36]	ACC(%)	89.36	63.38	92.70	76.72	63.70	74.07	73.57	94.78	82.31	78.95±11.68
	K	-	-	-	-	-	-	-	-	-	-
Proposed Method	ACC(%)	82.67	<b>74.69</b>	86.31	<b>82.79</b>	<b>84.19</b>	<b>79.37</b>	<b>79.38</b>	82.11	81.56	<b>81.45±3.3</b>
	K	0.769	<b>0.663</b>	<b>0.817</b>	<b>0.770</b>	<b>0.789</b>	<b>0.725</b>	<b>0.725</b>	0.761	0.753	<b>0.752±0.04</b>

TABLE III

MI CLASSIFICATION ACCURACIES (%) ON BCI COMPETITION IV DATASET 2B. FOR EACH SUBJECT, THE BEST RESULT IS MARKED IN BOLDFACE

Classification Accuracy (%)	B0103T	B0203T	B0303T	B0403T	B0503T	B0603T	B0703T	B0803T	B0903T	Avg±Std
SGRM [20]	76.30	56.00	49.00	<b>98.20</b>	91.10	74.80	<b>88.30</b>	85.40	84.90	78.78±16.3
CSP with clsSRC2 [21]	71.88	59.64	57.19	95.63	<b>92.50</b>	82.50	73.75	90.00	85.94	78.20±14.0
RCSP with clsSRC [22]	71.25	58.21	56.56	95.94	91.56	82.19	74.06	90.94	85.62	78.48±14.4
RCSP with clsSRC2 [21]	70.63	56.79	58.44	96.25	91.25	81.25	73.44	90.63	86.56	78.36±14.4
PSD-SVM [24]	73.12	61.78	58.43	93.75	87.50	82.81	76.87	<b>92.19</b>	<b>87.50</b>	79.33
FBCSP- RLDA [37]	80.70	67.00	57.00	88.50	83.10	75.60	78.90	73.50	73.60	75.36±9.26
Proposed Method	<b>81.11</b>	<b>78.14</b>	<b>76.88</b>	91.11	71.66	<b>91.51</b>	86.25	74.61	84.79	<b>81.78±7.1</b>

performs better than all the state-of-art methods in all metrics on BCI competition IV dataset 2a.

The comparison results of MI classification accuracy (%) on BCI competition IV dataset 2b are listed in TABLE III. The experimental results indicate that the proposed WTS-CC method performs better than all state-of-the-art approaches in average classification. Therefore, it demonstrates the capability of our method in MI-EEG discrimination.

Insufficient training data and individual differences among subjects are a major challenge in MI-EEG classification research. Few-shot learning aims to develop models using less training data, and it is a method that allows models to learn to adapt quickly to new tasks. To evaluate the effectiveness of the proposed WTS-CC, we conduct experiments on the ‘‘Track#1 Few-shot EEG learning’’ dataset in 2020 International BCI Competition. TABLE IV lists the experimental results of WTS-CC for MI-EEG classification accuracy (ACC (%)), Cohen’s kappa coefficient (K), F1-score (F1) and the AUC. The results indicate that the proposed WTS-CC performs well even on the few-shot learning dataset with an average accuracy of 83.31%.

Moreover, to further investigate the importance of the components of the proposed WTS-CC, we performed ablation studies.

#### D. Performance Evaluation of the DEC Module

The DEC module mainly learns the correlation between each EEG channel, filters out the attention for the EEG channel, and obtains a one-dimensional vector with the same number of channels as the weight to represent the correlation between the channel and the important feature information,

TABLE IV

MI CLASSIFICATION ACCURACIES (ACC (%)), COHEN'S KAPPA COEFFICIENT (K), F1-SCORE (F1) AND AUC ON 2020 INTERNATIONAL BCI COMPETITION DATASET #TRACK1

Subject	Proposed Method			
	ACC(%)	K	F1	AUC
S01	82.78	0.662	0.828	0.844
S02	84.61	0.683	0.84	0.833
S03	85.71	0.714	0.857	0.857
S04	78.57	0.571	0.787	0.792
S05	76.92	0.541	0.769	0.774
S06	85.71	0.714	0.86	0.889
S07	84.62	0.698	0.844	0.857
S08	85.71	0.708	0.857	0.854
S09	71.43	0.44	0.714	0.729
S10	64.29	0.285	0.645	0.646
S11	87.91	0.755	0.879	0.879
S12	78.57	0.553	0.782	0.771
S13	92.31	0.847	0.923	0.929
S14	89.93	0.794	0.898	0.891
S15	84.62	0.691	0.846	0.845
S16	85.35	0.704	0.849	0.849
S17	84.79	0.691	0.848	0.849
S18	89.93	0.799	0.9	0.914
S19	92.31	0.843	0.922	0.917
S20	80.04	0.596	0.801	0.801
Avg	83.31±6.87	0.664±0.13	0.832±0.06	0.808±0.06

which will affect the final classification result. First, we consider the adopted activation functions in the DEC module. TABLE V lists the comparison of classification accuracy between four different activation functions, including ReLU,

TABLE V

COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT ACTIVATION FUNCTIONS. THE BEST RESULT IS MARKED IN BOLDFACE

Activation Function	ACC(%)
ReLU	79.78±3.9
Tanh	78.65±3.8
Sigmoid	81.02±3.5
Softmax	<b>81.45±3.3</b>

TABLE VI

ABLATION STUDIES ON DEC MODULE. THE BEST RESULT IS MARKED IN BOLDFACE

Methods	ACC(%)	K	F1	AUC
w/o DEC module	80.46±4.1	0.739±0.07	0.804±0.05	0.939±0.01
w/ DEC module	<b>81.45±3.3</b>	<b>0.752±0.04</b>	<b>0.814±0.03</b>	<b>0.939±0.01</b>

Tanh, Sigmoid, and Softmax. The results denote that using Tanh as an activation function gives the worst average accuracy while using Softmax as an activation function achieves the highest average accuracy, which indicates that choosing an appropriate activation function is very important.

To evaluate the performance of the DEC module, we compare WTS-CC with and without the DEC module, as listed in TABLE VI. The results indicate that the proposed WTS-CC achieves an average accuracy of 81.45%, which is higher than that without the DEC module. The DEC module can help alleviate the heterogeneity of redundant features and enhance the more discriminative features. Therefore, only the feature information extracted from convolution operations of different sizes is insufficient for MI-EEG discrimination. The DEC module can judge the importance of each EEG channel and adjust the weighting according to its importance, which can effectively improve classification performance.

### E. Performance Evaluation of the WTS Module

In addition to feature information analysis with the DEC module, the WTS module also greatly influences the discrimination results. The concept of the WTS module is to capture significant time-frequency features. We use CWT and independent sample t-statistics to judge whether the difference between the time-frequency maps of two-group MI tasks is significant or not and find the most discriminative features to further improve the classification accuracy. Next, t-statistic-weighted temporal-spectral feature maps (tTSFMs) are established by multiplying their corresponding MI averaged TSFMs.

To evaluate the effectiveness of the WTS module, the results of WTS-CC with and without the WTS module are shown in TABLE VII. The results indicate that compared with WTS-CC with the DEC module, WTS-CC without the WTS module leads to a more obvious drop in classification accuracy, with only an average accuracy of 55.49%, which is much lower than the proposed WTS-CC. The WTS module indeed helps capture the features that are significantly different between the two groups of MI-EEG signals and enhances the feature information with large differences. Therefore, the WTS module can effectively enhance the characteristics of different

TABLE VII

ABLATION STUDIES ON WTS MODULE. THE BEST RESULT IS MARKED IN BOLDFACE

Methods	ACC(%)	K	F1	AUC
w/o WTS module	55.49±21.6	0.413±0.25	0.430±0.21	0.854±0.12
w/ WTS module	<b>81.45±3.3</b>	<b>0.752±0.04</b>	<b>0.814±0.03</b>	<b>0.939±0.01</b>

MI tasks, thereby significantly improving the classification performance by distinguishing the state of different MI tasks.

### F. Performance Evaluation of Subject-Dependent and Subject-Independent

In many BCI studies, EEG signal single-subject identification models based on various machine learning algorithms can achieve high accuracy, but this requires a large amount of EEG data for training, and the obtained models are only effective for specific subjects after training. They cannot be directly applied to cross-subject MI-EEG classification tasks. Therefore, many current studies have tried to create a more effective general model for multiple subjects, called the subject-independent model [38].

The subject-dependent results are shown in TABLE I. We use the proposed WTS-CC for the MI EEG classification of each subject and perform model training based on the experimental data of each subject, achieving an average classification accuracy of 81.45%. In order to evaluate the impact of individual differences among subjects on WTS-CC, we also conduct subject-independent experiments. We first combine the experimental data of multiple subjects as the input for the training model and evaluate the MI-EEG classification for different subjects. The average classification accuracy obtained is 62.05%.

The comparison results of subject-dependent and subject-independent for our method are listed in TABLE IX. As we can observe, compared to subject-dependent, subject-independent significantly leads to a decrease in classification accuracy, with only an average accuracy of 62.05%. This may be due to the fact that MI-EEG signals have characteristics of inter-subject variability and brain activity instability, which may limit the classification performance when dealing with the distribution of data from different subjects.

### G. Number of Parameters and Inference Time

The number of parameters and inference time between the state-of-the-art models and the proposed method are shown in TABLE X. In general, it is desirable for a model to have as few learnable parameters as possible to ensure its strong generalization [9]. To evaluate the computational cost of the proposed method during model training, we calculate our number of parameters and then compare it with other models. The results indicate that the proposed method is compact (i.e., has less number of parameters), which is efficient and sufficient to perform the MI-EEG classification task. Moreover, we also evaluate the model inference time, which is the time for a trained model to classify an MI-EEG trial. The average result is provided over all subjects for each dataset. The results



TABLE VIII

COMPARISON OF UNWEIGHTED AND WEIGHTED T-STATISTIC TIME-FREQUENCY PLOTS FOR CORRELATION COEFFICIENT. THE BEST RESULT IS MARKED IN BOLDFACE

Average Time-Frequency Diagram	Correlation Coefficient			
	Group A Unweighted	Group A Weighted	Group B Unweighted	Group B Weighted
Group A	0.6461	<b>0.8500</b>	0.6106	0.8243
Group B	0.5818	0.8427	0.6311	<b>0.8659</b>

denote that the proposed method needs the least inference time.

#### IV. DISCUSSIONS

##### A. Comparisons With the State-of-the-Arts

The comparison results of the state-of-the-art models and the proposed WTS-CC on BCI competition IV dataset 2a are listed in TABLE I in terms of classification accuracy (ACC (%)), Cohen’s kappa coefficient (K), F1-score (F1), and AUC. The experimental results denote that the proposed WTS-CC method achieves the best average classification accuracy, which is significantly better than all state-of-the-art models (i.e., Shallow ConvNet ( $p < 0.05$ ), Deep ConvNet ( $p < 0.01$ ), CP-MixedNet ( $p < 0.01$ ), and TS-SEFFNet ( $p < 0.05$ )). In addition, WTS-CC also outperforms all the state-of-the-art approaches in average Kappa coefficient, F1-score, and AUC. Therefore, the proposed WTS-CC is an effective MI-EEG classification method considering both deep EEG-channel attention and wavelet-based temporal-spectral attention.

Moreover, TABLE II lists the comparison results among the current state-of-the-art approaches and the proposed WTS-CC on BCI competition IV dataset 2a in terms of classification accuracy (ACC (%)) and Cohen’s kappa coefficient (K). SS-MEMDBF [33] and MEMDBF-CSP [34] are decomposition-based filtering methods, which extract multi-channel EEG signals and transform them into IMFs to reduce within-subject and subject-specific effects of EEG signals, while TS-TL [35] and LR-TSTL [36] are tangent space-based transfer learning classification models. The experimental results indicate that the proposed WTS-CC method achieves the best average classification accuracy (81.45%), which is significantly better than all state-of-the-art models (i.e., TS-TL [35] ( $p < 0.05$ ) and LR-TSTL [36] ( $p < 0.05$ )), and the best average Kappa coefficient (0.752), which is significantly better than all state-of-the-art models (i.e., SS-MEMDBF [33] ( $p < 0.01$ )). Accordingly, the proposed WTS-CC method achieves the best performance because it simultaneously considers the features and their weighting in spatial, EEG-channel, temporal and spectral domains.

TABLE III lists the comparison results of the state-of-the-art models and the proposed WTS-CC on BCI competition IV dataset 2b in terms of MI classification accuracy (%). The experimental results show that the proposed WTS-CC method significantly outperforms all state-of-the-art approaches in terms of average classification accuracy (SGRM [20] ( $p <$

TABLE IX

COMPARISON OF SUBJECT-DEPENDENT AND SUBJECT-INDEPENDENT FOR OUR METHOD. THE BEST RESULT IS MARKED IN BOLDFACE

Our Method	ACC(%)	K	F1	AUC
Subject-Dependent	<b>81.45±3.3</b>	<b>0.752±0.04</b>	<b>0.814±0.03</b>	<b>0.939±0.01</b>
Subject-Independent	62.05	0.241	0.620	0.379

TABLE X

COMPARISON OF NUMBER OF PARAMETERS AND INFERENCE TIME. THE BEST RESULT IS MARKED IN BOLDFACE

Dataset	Methods	Number of Parameters (million)	Inference time (ms)
	Shallow ConvNet	0.047	36.20
	Deep ConvNet	0.284	36.28
BCI IV 2a	CP-MixedNet	0.836	36.60
	TS-SEFFNet	0.282	36.22
	Proposed Method	<b>0.004</b>	<b>21.00</b>

0.05), clsSRC2 [21] ( $p < 0.05$ ), and clsSRC [22] ( $p < 0.05$ )). Thus, it also demonstrates the capability and power of deep EEG-channel attention and wavelet-based temporal-spectral attention in our MI-EEG discrimination method.

TABLE IV lists the experimental results of the proposed WTS-CC method for each subject on the “Track#1 Few-shot EEG learning” dataset in 2020 International BCI Competition in terms of classification accuracy (ACC (%)), Cohen’s kappa coefficient (K), F1-score (F1) and the AUC. Since insufficient training data is a challenge in MI-EEG classification studies, few-shot learning is important, aiming to develop models using less training data, which is a way for models to learn to quickly adapt to new tasks. The experimental results indicate that the proposed WTS-CC method performs well even on the four metrics in the few-shot learning dataset (“Track#1 Few-shot EEG learning” dataset in 2020 International BCI Competition).

##### B. Discussion of the DEC and WTS Modules

To evaluate the effectiveness of the DEC module, TABLE VI lists the results of WTS-CC with and without the DEC modules. The results denote that the proposed WTS-CC with the DEC module is better than that without the DEC module. It is because the DEC module can help alleviate the heterogeneity of redundant features and enhance the more discriminative features. Accordingly, only the feature information extracted from convolution operations of different sizes is insufficient for MI-EEG discrimination. The DEC module can effectively judge the importance of each EEG channel and adjust the weighting according to its importance, which can further improve classification performance.

In addition to the DEC module, the WTS module also greatly influences the discrimination results. To evaluate the effectiveness of the WTS module, TABLE VII lists the results of WTS-CC with and without the WTS modules. The results indicate that compared with WTS-CC with the DEC module,

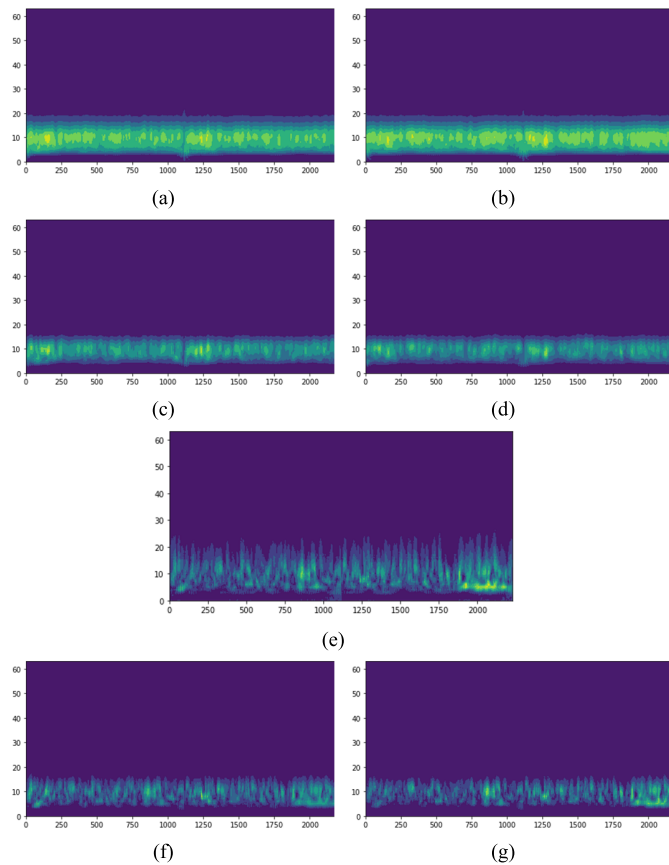


Fig. 2. The first stage (left hand and right hand in group A; feet and tongue in group B) (a) Average TSMs of Group A, (b) Average TSMs of Group B, (c) Standard deviation TSMs of Group A, (d) Standard deviation TSMs of Group B, (e) t-statistics TSMs, (f) tTSMs of Group A, (g) tTSMs of Group B.

WTS-CC without the WTS module leads to a more obvious drop in classification accuracy (average 55.49%), which is much lower than the proposed WTS-CC. The WTS module indeed helps capture the features that are significantly different between the two groups of MI-EEG signals and enhances the feature information with large differences. Accordingly, the WTS module can effectively enhance the characteristics of different MI tasks, thereby significantly improving the classification performance by distinguishing the state of different MI tasks.

### C. Space and Time Complexity

For the training data of each subject in all the datasets, the initial important temporal features of MI EEG signals are extracted by the iTFE module. The DEC module is then proposed to automatically adjust the weight of each EEG channel according to its importance. Next, Next, the Wavelet-based Temporal-Spectral-attention (WTS) module is proposed to obtain more significant discriminative features between different MI tasks. The two-dimensional time-frequency maps of different MI tasks (such as left and right MI tasks) have been obtained, respectively. It does not undergo any training process. Therefore, the proposed WTS-CC method is fast and effective, which has significantly less number of parameters and needs less inference time, as listed in TABLE X.

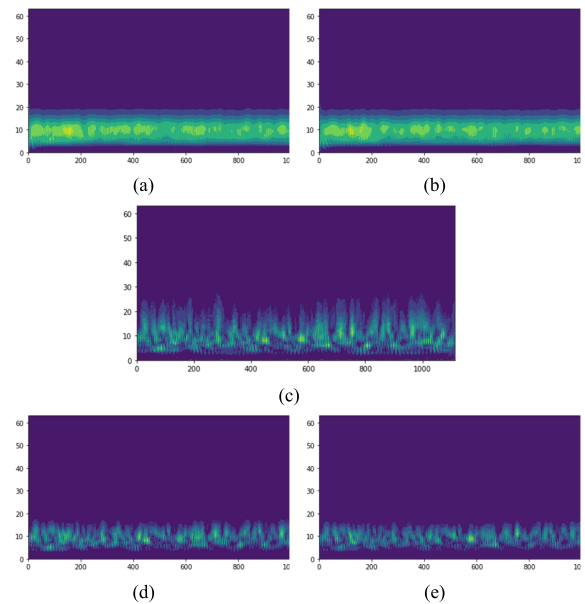


Fig. 3. The Second stage (left hand of group A, right hand of group B) (a) Average TSMs of group A, (b) Average TSMs of group B, (c) t-statistics TSMs, (d) tTSMs of group A, (e) tTSMs of group B.

### D. Advantages, Limitations, and Future Studies

The advantages of the proposed method are that WTS-CC can achieve fast, effective, and more accurate MI EEG discrimination by simultaneously considering the features and their weighting in spatial, EEG-channel, temporal, and spectral domains.

Although our proposed WTS-CC can effectively classify MI EEG signals, our method is still subject to some limitations. Our method uses independent sample T statistics to evaluate the degree of feature differences between two classes of different MI tasks. Independent sample T statistics is a statistical method used to judge whether there is a significant difference between two classes of data. Specifically, each experiment must be carried out with two classes of MI data in order to correctly calculate the T statistic and obtain the weight of the time-frequency feature. This method can be used to judge two or four classes of MI EEG studies, but it cannot be directly applied to distinguish three classes of studies, which can be further solved by the one-versus-one method.

Finally, our method uses the correlation coefficient to assess the degree of correlation between the test data and the mean time-frequency maps of different MI tasks, which is a simple and convenient classification method. In future studies, we can conduct with the classifiers, which may further improve the performance of MI EEG signal classification.

## V. CONCLUSION

In this study, a novel WTS-CC model is proposed to effectively improve MI-EEG discrimination by simultaneously taking into account the features and their weighting of spatial, EEG-channel, temporal and spectral domains. The initial important temporal features are extracted by means of the iTFE module. The DEC module effectively enhances more important EEG channels and suppresses less important ones simultaneously by automatically adjusting each EEG

channel's weight according to its importance. The WTS module obtains more significant discriminative temporal-spectral features between different MI tasks, which can further help the discrimination module enhance the classification. The experimental results indicate that the proposed WTS-CC method achieves promising discrimination performance outperforming the state-of-the-art methods in terms of classification accuracy, Kappa coefficient, F1 score, and AUC on three public datasets. In future work, we intend to study how the WTS-CC model can be modified to develop and apply to adaptive discrimination. The model is helpful in BCI applications due to the nature of time-varying EEG signals. In addition, we will extend our method with transfer learning to design an effective BCI system.

## REFERENCES

- [1] 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.
- [2] W.-Y. Hsu and Y.-N. Sun, "EEG-based motor imagery analysis using weighted wavelet transform features," *J. Neurosci. Methods*, vol. 176, no. 2, pp. 310–318, 2009.
- [3] S. Ren, W. Wang, Z.-G. Hou, X. Liang, J. Wang, and W. Shi, "Enhanced motor imagery based brain-computer interface via FES and VR for lower limbs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 8, pp. 1846–1855, Aug. 2020.
- [4] A. Jiang, J. Shang, X. Liu, Y. Tang, H. K. Kwan, and Y. Zhu, "Efficient CSP algorithm with spatio-temporal filtering for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 1006–1016, Apr. 2020.
- [5] W.-Y. Hsu, C.-C. Lin, M.-S. Ju, and Y.-N. Sun, "Wavelet-based fractal features with active segment selection: Application to single-trial EEG data," *J. Neurosci. Methods*, vol. 163, no. 1, pp. 145–160, Jun. 2007.
- [6] W. Y. Hsu, "Embedded prediction in feature extraction: Application to single-trial EEG discrimination," *Clin. EEG Neurosci.*, vol. 44, no. 1, pp. 31–38, 2013.
- [7] W.-Y. Hsu, "Continuous EEG signal analysis for asynchronous BCI application," *Int. J. Neural Syst.*, vol. 21, no. 4, pp. 335–350, 2011.
- [8] W.-Y. Hsu, "Clustering-based compression connected to cloud databases in telemedicine and long-term care applications," *Telematics Informat.*, vol. 34, no. 1, pp. 299–310, Feb. 2017.
- [9] R. T. Schirmer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [10] W.-Y. Hsu, "Enhanced active segment selection for single-trial EEG classification," *Clin. EEG Neurosci.*, vol. 43, no. 2, pp. 87–96, Apr. 2012.
- [11] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2818–2830, Sep. 2020.
- [12] W. Y. Hsu, Y. C. Li, C. Y. Hsu, C. T. Liu, and H. W. Chiu, "Application of multiscale amplitude modulation features and fuzzy C-means to brain-computer interface," *Clin. EEG Neurosci.*, vol. 43, no. 1, pp. 32–38, 2012.
- [13] W.-Y. Hsu, "Application of competitive Hopfield neural network to brain-computer interface systems," *Int. J. Neural Syst.*, vol. 22, no. 1, pp. 51–62, Feb. 2012.
- [14] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 204–207, doi: 10.1109/CNE.2007.369647.
- [15] 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, no. 1, p. 39, 2012.
- [16] L. Yang, Y. Song, K. Ma, and L. Xie, "Motor imagery EEG decoding method based on a discriminative feature learning strategy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 368–379, 2021.
- [17] J. Chen, Z. Yu, Z. Gu, and Y. Li, "Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2356–2366, Nov. 2020.
- [18] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *J. Neural Eng.*, vol. 16, no. 6, Oct. 2019, Art. no. 066004.
- [19] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze- and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
- [20] Y. Jiao et al., "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.
- [21] V. P. Oikonomou, S. Nikolopoulos, and I. Kompatsiaris, "Robust motor imagery classification using sparse representations and grouping structures," *IEEE Access*, vol. 8, pp. 98572–98583, 2020, doi: 10.1109/ACCESS.2020.2997116.
- [22] V. P. Oikonomou, S. Nikolopoulos, and I. Kompatsiaris, "Motor imagery classification via clustered-group sparse representation," in *Proc. IEEE 19th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2019, pp. 1–5.
- [23] 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.
- [24] V. P. Oikonomou, K. Georgiadis, G. Liaros, S. Nikolopoulos, and I. Kompatsiaris, "A comparison study on EEG signal processing techniques using motor imagery EEG data," in *Proc. IEEE 30th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2017, pp. 781–786.
- [25] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 317–326, Aug. 2008.
- [26] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [27] Y. Shin, S. Lee, J. Lee, and H. N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems," *J. Neural Eng.*, vol. 9, no. 5, 2012, Art. no. 056002.
- [28] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and- excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021.
- [29] B. Xu et al., "Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification," *IEEE Access*, vol. 7, pp. 6084–6093, 2018, doi: 10.1109/ACCESS.2018.2889093.
- [30] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.
- [31] D. Nath, M. Singh, D. Sethia, D. Kalra, and S. Indu, "A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network," in *Proc. 4th Int. Conf. Comput. Data Anal.*, Mar. 2020, pp. 142–147.
- [32] J. H. Jeong et al., "2020 International brain-computer interface competition: A review," *Frontiers Hum. Neurosci.*, vol. 16, Jul. 2022, Art. no. 898300.
- [33] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Syst. Appl.*, vol. 95, pp. 201–211, Apr. 2018.
- [34] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "An automatic subject specific intrinsic mode function selection for enhancing two-class EEG-based motor imagery-brain computer interface," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6938–6947, Aug. 2019.
- [35] P. Gaur, K. McCreddie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent space features-based transfer learning classification model for two-class motor imagery brain-computer interface," *Int. J. Neural Syst.*, vol. 29, no. 10, 2019, Art. no. 1950025.
- [36] P. Gaur, A. Chowdhury, K. McCreddie, R. B. Pachori, and H. Wang, "Logistic regression with tangent space-based cross-subject learning for enhancing motor imagery classification," *IEEE Trans. Cogn. Develop. Syst.*, vol. 14, no. 3, pp. 1188–1197, Sep. 2022.
- [37] S. Muthong, P. Vateekul, and M. Sriyudthsak, "Stacked probabilistic regularized LDA on partitioning non-stationary EEG data for left/right hand imagery classification," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2016, pp. 301–306.
- [38] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network," *Neural Netw.*, vol. 136, pp. 1–10, Apr. 2021.