Learning Common Time-Frequency-Spatial Patterns for Motor Imagery Classification

Yangyang Miao, Jing Jin, *Senior Member, IEEE*, Ian Daly, Cili Zuo, Xingyu Wang,
Andrzej Cichocki, *Fellow, IEEE*, and Tzyy-Ping Jung, *Fellow, IEEE*

Abstract— **The common spatial patterns (CSP) algorithm is the most popular spatial filtering method applied to extract electroencephalogram (EEG) features for motor imagery (MI) based brain-computer interface (BCI) systems. The effectiveness of the CSP algorithm depends on optimal selection of the frequency band and time window from the EEG. Many algorithms have been designed to optimize frequency band selection for CSP, while few algorithms seek to optimize the time window. This study proposes a novel framework, termed common time-frequency-spatial patterns (CTFSP), to extract sparse CSP features from multi-band filtered EEG data in multiple time windows. Specifically, the whole MI period is first segmented into multiple subseries using a sliding time window approach. Then, sparse CSP features are extracted from multiple frequency bands in each time window. Finally, multiple support vector machine (SVM) classifiers with the Radial Basis Function (RBF) kernel are trained to identify the MI tasks and the voting result of these classifiers determines the final output of the BCI. This study applies the proposed CTFSP algorithm to three public EEG datasets (BCI competition III dataset IVa, BCI competition III dataset IIIa, and BCI competition IV dataset 1) to validate its effectiveness, compared against several other state-of-the-art methods. The experimental results demonstrate that the proposed algorithm is a promising candidate for improving the performance of MI-BCI systems.**

Index Terms— **Common spatial patterns (CSP), motor imagery (MI), electroencephalogram (EEG), brain-computer interface (BCI).**

I. INTRODUCTION

BRAIN-COMPUTER interfaces (BCIs) establish a direct connection link between the brain and the external world, which is independent of peripheral nerves and muscles [1], [2]. BCIs not only provide an alternative communication channel for disabled patients [3]–[6], but also have many promising applications for healthy users such as environment control [7], [8] or fatigue detection [9], [10]. To acquire brain signals for controlling a BCI the electroencephalography (EEG) is commonly used due to its non-invasive nature, high temporal resolution, and low cost [11], [12]. However, EEG signals suffer from external noise and are prone to physiological artifacts, which poses challenges when using them for BCI control [13].

Among BCI systems, motor imagery (MI)-based BCIs can be more flexible in their applications than many other types of BCI because they can be operated asynchronously and can be more intuitive to control [14]–[17]. During MI, the rhythmic EEG activity is suppressed on the contralateral side of the brain to the limb the individual is attempting to control. This is the so-called event-related desynchronization (ERD) [18]. However, the spatial location in the brain, temporal onset, relative decrease in EEG power, and stability of the ERD are highly variable across trials, sessions, and individuals [19]. This poses a considerable challenge for designing MI-BCI systems.

To optimally extraction of EEG features that describe the ERD, researchers proposed the common spatial patterns (CSP) algorithm, which seeks spatial filters to extract the features that optimally discriminate different motor control attempts [20]. Due to its success, a large number of variants of CSP have emerged. For instance, regularization was introduced to CSP to solve the problem of overfitting [21]–[24].

The effectiveness of CSP depends on identifying the optimal EEG frequency band. However, the optimal filter frequency band is participant-specific, meaning a general solution with a fixed frequency band is not possible. Thus, several variants of CSP use different frequency bands, including methods such as sub-band CSP (SBCSP) [25], filter bank CSP (FBCSP) [26], discriminative FBCSP (DFBCSP) [27], and sliding window discriminative CSP (SWDCSP) [28].
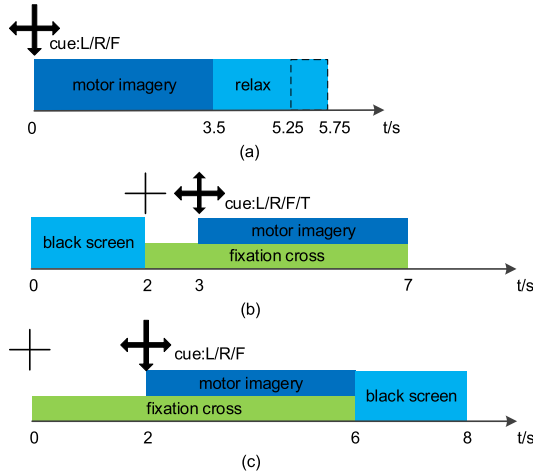
Fig. 1.   Timeline of one trial for (a) dataset IVa from BCI Competition III, (b) dataset IIIa from BCI Competition III, (c) dataset 1 from BCI Competition IV.

However, in the above literature, the importance of the temporal onset of the ERD is often overlooked. The time course of the ERD following a movement cue varies over time and across participants [29]. Therefore, an optimal CSP method should account for this variability when training the spatial filter. However, only a few studies have tried to tackle this problem by employing approaches for automatic selection of the optimal time interval [30], [31].

This study proposed a novel framework, termed common time-frequency-spatial patterns (CTFSP), to learn sparse CSP features from multi-band filtered EEG data over multiple candidate time windows. The major innovations and contributions of this work can be summarized as follows: 1) We adopted a sliding time window approach to decode MI tasks. 2) In each sub time window, we extracted sparse CSP features from multiple candidate frequency bands. 3) We developed a novel classification method that makes use of a classifier fusion strategy. The final decision output is based on the results from multi classifiers to reduce the risk of misclassification. Finally, we validated the efficiency of the proposed framework using three public EEG datasets.

The remainder of this paper is organized as follows. Section 2 introduces the three BCI Competition Datasets used in this study and describes the details of our proposed framework. Section 3 presents the experimental results, which are then discussed in Section 4. Section 5 concludes this work.

II. METHODS

A. Data Description

Dataset 1 (DS1): The first dataset we use is Dataset IVa from the BCI Competition III [32]. It was recorded from five healthy participants labelled aa, al, av, aw, and ay. Visual cues were displayed for a period of 3.5 s, during which the participants were instructed to perform one of the corresponding MI tasks: left hand, right hand, and right foot imagination. Then, the participants were instructed to relax for a period of between 1.75 to 2.25 s. Each participant was asked to complete 280 trials. The timeline of this experiment is shown in Fig. 1(a). The original dataset includes

S : number of samples
M : number of channels
N : number of all training trials
N_c : number of the training trials in class c
$\mathbf{E}_{i,c} \in \mathbb{R}^{M \times S}$: EEG data matrix of i-th trial in class c
$\mathbf{R}_c, \mathbf{R}_{\bar{c}} \in \mathbb{R}^{M \times M}$: average spatial covariance matrix of class c or class $\bar{c}$
$\mathbf{W}_{\mathrm{csp}}$: spatial filter of CSP method

signals from 118 EEG channels, which are down-sampled to 100 Hz. More details can be found on the following website: http://www.bbci.de/competition/iii/.

Dataset 2 (DS2): The second dataset is Dataset IIIa from the BCI Competition III [32]. It was recorded from 60 EEG electrodes with a sample rate of 250 Hz from three participants labelled k3, k6, and l1. The participants were instructed to rest for 2 s at the beginning of each trial. At t = 2 s, an acoustic stimulus and fixation cross were presented. At t = 3 s, an arrow pointing left, right, up, or down was displayed for 1 s and the participant was asked to imagine a left-hand, right-hand, foot, or tongue movement until the cross disappeared at t = 7 s. The numbers of trials per class are 90, 60, and 60 for participants 'k3', 'k6' and 'l1', respectively. In this study, we only select the trials from two MI classes: left- and right-hand movements for Participants 'k3' and 'l1'; right-hand and tongue movements for Participant 'k6'. Fig. 1(b) shows the timeline of the experiment.

Dataset 3 (DS3): The third dataset is Dataset 1 from the BCI Competition IV [33]. This dataset was recorded from seven participants performing MI. At the beginning of each trial, a fixation cross was first displayed at the center of the computer screen for a period of 2 s. At t = 2 s, an arrow pointing left, right, or down was displayed and the participant was asked to perform the corresponding MI task (left / right hand and foot) until the cross disappeared at t = 6 s. Then, the participant was instructed to rest for 2 s. Fig. 1(c) shows the timeline of the experiment. Each participant was asked to complete 200 trials. We do not use the data from three participants labelled c, d, and e, because they are artificially generated. The dataset includes signals from 59 EEG channels, which are down-sampled to 100 Hz. More details can be found on the following website: http://www.bbci.de/competition/iv/.

B. Preprocessing

Table I shows a list of the notations and definitions that are used later. A fifth-order Butterworth band-pass filter of 8-30 Hz was first used to filter out components unrelated to sensorimotor rhythms. Then, the filtered EEG signals are divided into seven frequency bands: mu band (8-13 Hz), two sub-bands of the mu band (8-10 Hz; 10-13 Hz), beta band

(13-30 Hz), and three sub-bands of the beta band (13-18 Hz; 18-23 Hz; 23-30 Hz).

C. Feature Extraction

CSP is a widely used feature-extraction method in MI-based BCIs. It optimizes a set of spatial filters to maximize the variance of one class while minimizing the variance of the other class. The average spatial covariance matrix can be computed as

$$\mathbf{R}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{\mathbf{E}_{i,c}\mathbf{E}_{i,c}^T}{\mathrm{tr}(\mathbf{E}_{i,c}\mathbf{E}_{i,c}^T)} \quad (1)$$

We can find a spatial filter that maximizes the variance of one class and minimizes the other class by solving

$$\arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R}_c \mathbf{w}}{\mathbf{w}^T \mathbf{R}_{\bar{c}} \mathbf{w}} \quad (2)$$

The above Rayleigh quotient can be transformed into the generalized eigenvalue problem

$$\mathbf{R}_c \mathbf{w} = \lambda \mathbf{R}_{\bar{c}} \mathbf{w} \quad (3)$$

where λ and $\mathbf{w}$ are the generalized eigenvalue and eigenvector, respectively. The spatial filter $\mathbf{W}_{\mathrm{csp}} = [\mathbf{w}_1, \ldots, \mathbf{w}_{2m}]$ is formed by the eigenvector corresponding to the m maximum and minimum eigenvalues.

D. Feature Selection

By applying CSP on each of the filtered signals, we derive the following feature set

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,D} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \cdots & \mathbf{x}_{N,D} \end{bmatrix} \quad (4)$$

where $\mathbf{x}_{i,j}$ denotes the j-th feature extracted from EEG in the i-th trial and $D = 2m \times 7$ is the dimensionality of the feature set.

When CSP is applied to multiple frequency bands multiple features are generated. We used feature selection to reduce the dimensionality of these features and simplify the subsequent classification model. The least absolute shrinkage and selection operator (LASSO) is a penalized least squares method imposing an L_1-penalty on the regression coefficients [34], [35]. The LASSO estimates are defined as

$$\arg\min_{\boldsymbol{\beta},\beta_0} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{D} |\beta_j| \right) \quad (5)$$

where y_i is the class label for trial i, $\mathbf{x}_i$ is the D-dimensional feature vector for trial i, λ is a positive regularization parameter, and $\boldsymbol{\beta}$ and β_0 are regression parameters (β is a D-dimensional vector; β_0 is a scalar).

The LASSO method does not depend on any classifier. The features are automatically discarded corresponding to coefficients that are exactly 0. Thus, the most significant features are selected from multiple frequency bands.

E. SVM Classification

The support vector machine (SVM) has broad applications in BCI systems. It finds an optimal hyperplane with the largest possible margin to separate the samples from two classes [36], [37].

One variant of the algorithm is to solve the following optimization problem

$$\min_{\omega,b,\xi} \ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N} \xi_i$$
$$s.t. \ \ y_i(\omega^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \ \xi_i \geq 0, \ i = (1, \ldots, \mathrm{n}) \quad (6)$$

where $\phi(x_i)$ maps $\mathbf{x}_i$ into a higher-dimensional space, C is the penalty parameter of the error term, and ξ is the slack variable.

F. Sliding Time Windows

The time latency during the MI period varies across trials for each participant [30]. To address this issue, we use three temporal segments with 0.5 s intervals instead of time window selection. A time window of length 2.5 s is used with DS1, while a time window of length 3 s used is with DS2 and DS3. Specifically, the three sub time windows used with DS1 are [0-2.5], [0.5-3], and [1-3.5]. The three sub time windows used with DS2 are [3-6], [3.5-6.5], and [4-7]. Finally, the three sub time windows used with DS3 are [2-5], [2.5-5.5], and [3-6] (please also refer to the timeline in Fig. 1).

G. Decision Output

Three SVM classifiers are formed because three sub time windows are used. The decision output is based on the "Max Wins" voting strategy [38], in which three binary SVM classifiers will vote for each class, and the winner class will be the class receiving the maximum votes. The decision function F_j of each SVM classifier is

$$F_j = \mathrm{sgn}(\omega_j^T \phi_j(\mathbf{x}) + b_j). \ j = 1, 2, 3 \quad (7)$$

The final decision output F is

$$F = \mathrm{sgn}(\sum_{j=1}^{3} F_j) \quad (8)$$

III. RESULTS

A. Whole Framework

We first give an overview of the architecture of our method. Our goal is to learn a robust model that can achieve high classification accuracy. To achieve this goal, we propose a novel method, CTFSP, to learn the sparse features from multi-band filtered EEG data across multiple time windows. To be specific, CTFSP is formulated as a combination of multi-band filtering, feature extraction, sparse feature selection, and classifier fusion, resulting in the overall architecture illustrated in Fig. 2. First, the whole MI period is segmented into multiple time windows. In each time window, multi-band filtering is used to increase the signal-to-noise ratio (SNR) of
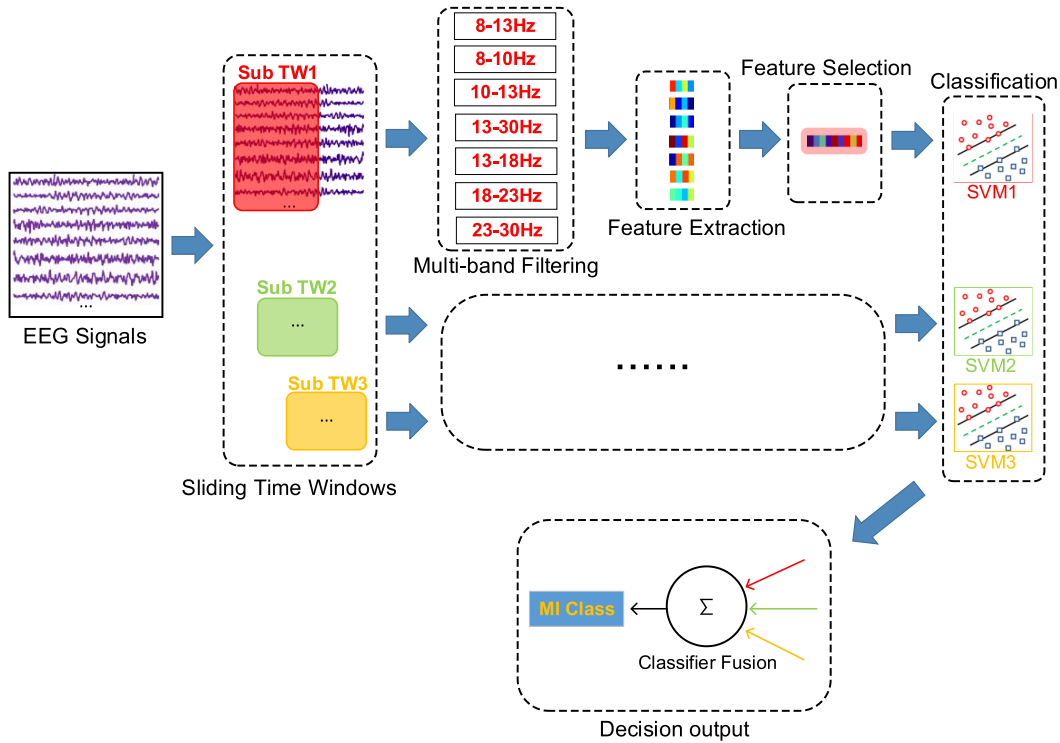
Fig. 2. Architecture of the proposed CTFSP method. 'TW' represents time window.

the EEG signals by extracting MI-related features from the mu band, beta band, and their sub-bands. Feature extraction and selection may learn sparse spatial features by using the CSP algorithm and the LASSO method. The final decision output is determined by the voting result of three SVM classifiers.

B. Experimental Results

An extensive experimental comparison is implemented to compare the performance of our proposed method to other competing algorithms.

(1) CSP: Features are extracted by CSP from EEG data within the whole time window ([0 3.5] with DS1, [3 7] with DS2, and [2 6] with DS3) in the frequency band (8-30 Hz).

(2) FBCSP [39]: CSP features are extracted from EEG data within the whole time window at multiple frequency bands (4-8, 8-12, …, 36-40 Hz). Discriminative pairs of frequency bands and corresponding CSP features are automatically selected based on the Mutual Information Best Individual Feature (MIBIF) selection algorithm.

(3) SCSP [40]: Features are extracted by CSP from EEG data within the whole time window in the frequency band (8-30 Hz) and then optimized to obtain sparse CSP features by introducing an L_1-norm regularization term in the optimization problem of CSP.

(4) SMFCSP: CSP features are extracted from EEG data within the whole time window at multiple frequency bands (8-13, 8-10, …, 23-30 Hz). The LASSO method is then used to obtain sparse features from multi-frequency filtered CSP features.

(5) CTFSP: The whole MI period is segmented into multiple sub time windows. In each sub time window, the SMFCSP algorithm is implemented.

We then used the LIBSVM classifier to classify the filtered EEG data after applying each of the above comparative methods [41]. The Radial Basis Function (RBF) is chosen as the kernel function, and the default classifier parameters are used.[1] Note that, in our proposed CTFSP algorithm, three SVM classifiers are formed and the final decision is determined by their voting results.

For each of the three datasets, a 10-fold cross-validation approach is used to evaluate the classification performance. The number of CSP filters used is set to 4 (i.e., $m = 2$).

Table II summarizes classification accuracies for all participants. For three datasets, the proposed CTFSP algorithm achieved the highest classification accuracy among the five algorithms. The classification performance of the CTFSP algorithm was significantly better than the performance achieved with CSP ($p < 0.005$), SCSP ($p < 0.005$), and SMFCSP ($p < 0.005$), as assessed with a Wilcoxon signed-rank test.

Specifically, for five participants from DS1, the mean accuracies were 78.57% (with CSP), 82.36% (with FBCSP), 78.43% (with SCSP), 82.57% (with SMFCSP), and 85.00% (with CTFSP), respectively. For three participants from DS2, the mean accuracies were 92.32% (with CSP), 87.32% (with FBCSP), 90.36% (with SCSP), 93.43% (with SMFCSP), and 96.20% (with CTFSP), respectively. For four participants from DS3, the mean accuracies were 56.38% (with CSP), 60.50% (with FBCSP), 56.25% (with SCSP), 68.50% (with SMFCSP), and 75.00% (with CTFSP), respectively.

C. Comparison of Feature Distributions

To facilitate the comparison of feature distributions, Fig. 3 shows the distributions of the two features obtained by CSP,

[1] www.csie.ntu.edu.tw/~cjlin/libsvm/

| Participant	Methods				
	CSP	FBCSP	SCSP	SMFCSP	CTFSP
aa	80.36	83.57	80.71	80.71	**86.07**
al	96.07	94.64	96.07	96.79	**98.57**
av	46.79	51.43	46.07	50.00	**52.14**
aw	84.64	93.93	84.29	95.00	**96.07**
ay	85.00	88.21	85.00	90.36	**92.14**
Mean±std	78.57±18.69	82.36±17.87	78.43±18.98	82.57±19.25	**85.00±18.96**
k3(L vs R)	96.11	97.78	94.44	97.78	**99.44**
k6(R vs T)	89.17	95.00	88.33	94.17	**95.83**
l1(L vs R)	91.67	69.17	88.33	88.33	**93.33**
Mean±std	92.32±3.51	87.32±15.78	90.36±3.53	93.43±4.77	**96.20±3.07**
a	55.00	64.50	54.50	82.00	**86.50**
b	45.50	48.00	46.00	53.00	**53.50**
f	48.50	50.00	48.50	53.50	**66.50**
g	76.50	79.50	76.00	85.50	**93.50**
Mean±std	56.38±13.99	60.50±14.65	56.25±13.64	68.50±17.67	**75.00±18.34**
p-value	0.002	0.002	0.002	0.002	-
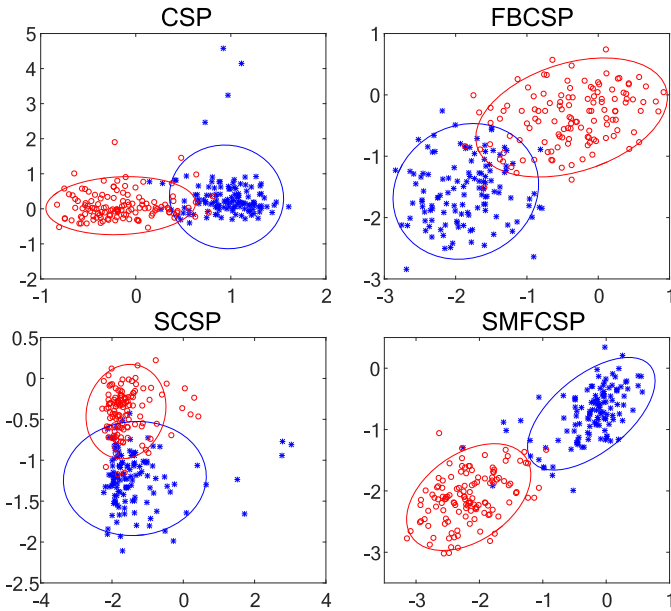
Fig. 3. Distributions of the two features obtained by CSP, FBCSP, SCSP and SMFCSP, respectively, for exemplar participant aw. The two large circles reflect the distribution range of 90% of the features in each class.

FBCSP, SCSP, and SMFCSP for exemplar participant aw. The SMFCSP algorithm provided more easily separated feature distributions in comparison with the other three algorithms. This is consistent with the classification performance of the four algorithms. Fig. 4 shows the distributions of the two features with different sub time windows obtained by CTFSP, for exemplar participant aw. The two features with sub time window 1 are the least separable.

D. Classification Performance With Different Time Windows

Fig. 5 presents the classification accuracies obtained by applying the CSP algorithm to EEG data with different time windows for all the participants. The results show that the whole time window is not the best choice for all participants except participant aa. The classification accuracies with sub time window 2 are higher for participants aa, ay, k3, l1, a, and g. The classification accuracies with sub time window 3 are higher for participants al, aw, b, and f. The classification accuracy with sub time window 1 is higher for participant k6. Finally, the classification accuracies with sub time window 1 and with sub time window 3 are the same for participant av.

This shows that the response times to the cues for MI tasks are not the same for all participants. This, in turn, suggests that our multiple time window approach may reduce the risk of misclassification caused by selecting the wrong time window.

E. Classification Performance With Different Fusion Strategies

Fig. 6 describes the process of the frameworks with two different fusion strategies. We carried out the proposed algorithms with classifier fusion (Fig. 6(a)) and feature fusion (Fig. 6(b)), respectively. Table III lists the individual classification accuracies for all participants with different fusion strategies. We found that the feature fusion approach obtained a slightly higher mean accuracy of 84.57%. However, there was no significant difference between the two fusion strategies in terms of accuracy ($p > 0.05$), as measured with a Wilcoxon signed-rank test.
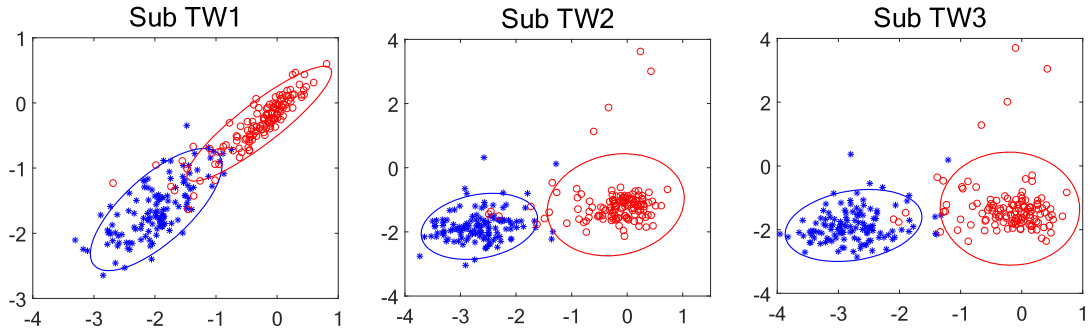
Fig. 4. Distributions of the two features with different sub time windows obtained by CTFSP, for exemplar participant aw.
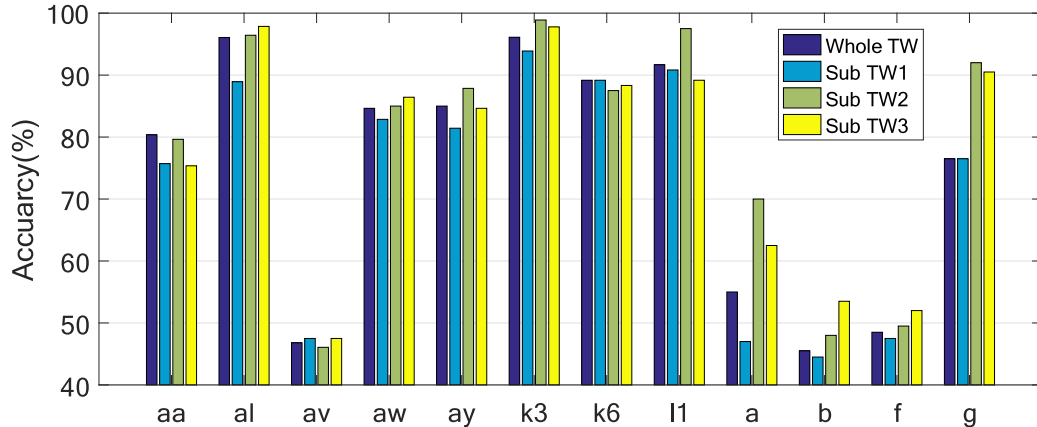

Fig. 5. Classification accuracies for all participants from datasets DS1, DS2, and DS3 with different time windows.
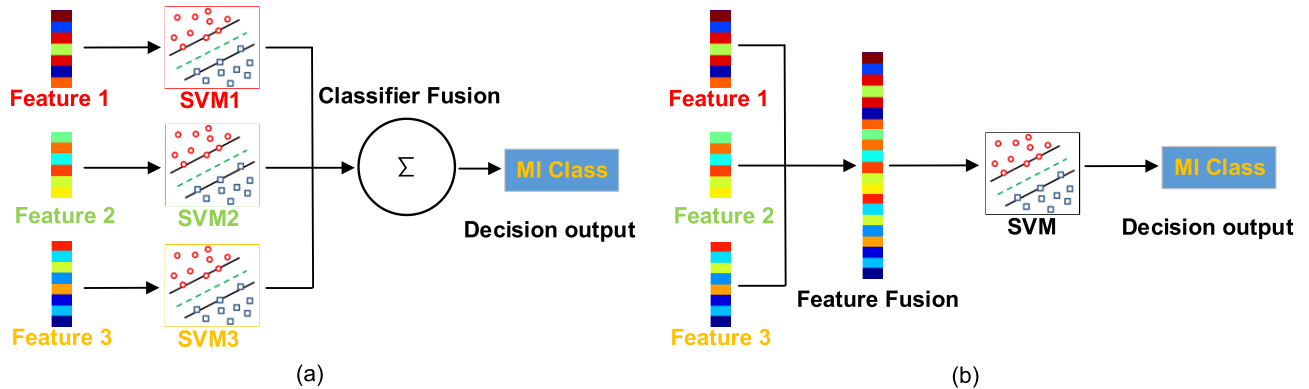

Fig. 6. The frameworks with different fusion strategies. (a) The fusion strategy is applied in the decision space. (b) The fusion strategy is applied in the feature space.

F. Computational Efficiency

We also investigated the computational efficiency for each of the compared algorithms on DS1. Fig. 7 shows the computational time evaluated with 10-fold cross validation using MATLAB R2016b on a PC with a i5-9600K 3.7GHz CPU, and 16GB of RAM. The results indicate that all of these algorithms can be implemented efficiently. Beside this, Table IV lists the testing time needed for each of these algorithms for one trial for participant aa. Although our algorithm took a longer time to compute than other methods, it could meet the requirement of real-time processing (it took less time to compute than the length of the trial). Most of time cost of our method

is spent on training multiple models. In other words, our proposed CTFSP algorithm achieved classification accuracy improvements without sacrificing computational efficiency for BCI applications.

IV. DISCUSSION

CSP is a spatial feature extraction method that has become the most commonly used algorithm in MI-BCIs [31]. However, it is also sensitive to noise and has low generalization capacity. Thus, many modifications to the CSP algorithm have been proposed to extend the generalizability of the CSP algorithm across the frequency domain to compensate for the

TABLE III
ACCURACIES (%) FOR ALL PARTICIPANTS USING CTFSP WITH DIFFERENT FUSION STRATEGIES

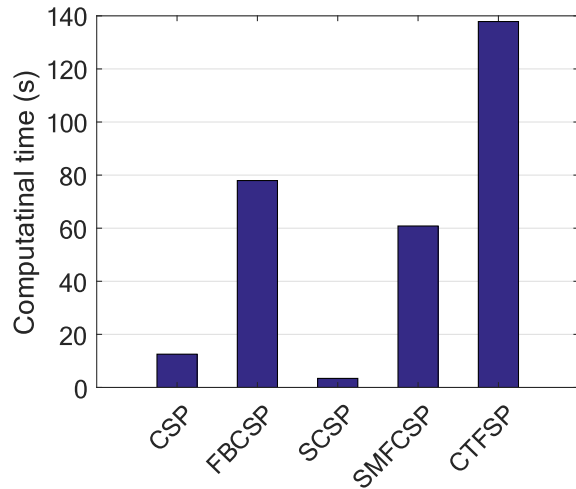| Participant	Fusion strategies	
	Classifier fusion	Feature fusion
aa	**86.07**	80.36
al	**98.57**	97.50
av	**52.14**	51.43
aw	**96.07**	95.71
ay	**92.14**	91.43
k3	**99.44**	99.44
k6	**95.83**	95.00
l1	93.33	**95.00**
a	**86.50**	84.00
b	53.50	**60.00**
f	66.50	**71.50**
g	**93.50**	93.50
Mean±std	84.47±17.16	**84.57±15.80**
p-value	>0.05	

Fig. 7. Computational time taken by each of the compared algorithms on DS1 with 10-fold cross validation.

TABLE IV
COMPARISON OF COMPUTATIONAL TIME FOR TESTING ONE TRIAL WITH DIFFERENT ALGORITHMS. THE TRIAL IS MORE THAN 5.25 S LONG

Methods	Testing time (ms)
CSP	8.75
FBCSP	78.86
SCSP	13.91
SMFCSP	60.48
CTFSP	143.19

shortcomings of conventional CSP. Extensions to CSP have included efforts such as SFBCSP [42] and SBLFB [43]. In the current study, we used multiple frequency bands to extract CSP features, including the mu band, beta band, and their sub-bands. Then, our proposed method selected the most useful features from the set of extracted features by means of sparse regression.

Previous studies only considered the influence of the frequency domain and ignored the differences in response times to the cues during MI tasks. As shown in Fig. 5, the response latency of the MI tasks varied for each participant. Therefore, this study used a sliding time window approach and segmented the entire MI period into three sub time windows. We also compared the performance of the frameworks with two fusion strategies. The results show that the classification accuracy with the classifier fusion was not significantly different from that achieved with the feature fusion approach.

Despite the proposed CTFSP algorithm obtaining highly encouraging performance in comparison with other algorithms, this work can further be improved along the following lines. First, the feature selection method based on LASSO does not depend on any classifier, so it may also omit some useful features for classification. Similarly, a previous study [44] showed that the sparse CSP algorithm generally gave lower performance than CSP. Therefore, there may be more effective feature-selection methods that can further improve classification performance. Second, in this study we only used the "Max Wins" voting strategy in the classifier fusion step. This might be the easiest classifier fusion strategy, but is not necessarily the best. Other approaches such as minimum / maximum fusion or Dempster-Shafer fusion could be considered in future work [45], [46]. Third, in fact motor imagery produces not only sensory motor rhythm (SMR) signals but also movement-related cortical potential (MRCP) signals [47]–[50]. Several signal-processing and classification methods have been used in low-frequency MRCP detection including Locality Preserving Projection (LPP) [51], Discriminative Canonical Pattern Matching (DCPM) [52] and CSP [53]. Therefore in future work, we will try to investigate the feasibility of the proposed CTFSP algorithm for simultaneous decoding of SMR and MRCP to reduce BCI inefficiency. Of course, transfer learning and deep learning have also receive increasing attention at the current BCI research [54], [55]. The combination of our work and these approaches may induce new vitality in the BCI field, and is worthy of further study.

V. Conclusion

This study proposed an integrated framework of multi-band filtering, CSP feature extraction, sparse feature selection, and classifier fusion, termed CTFSP. We evaluated our proposed algorithm on three public EEG datasets. Our results demonstrate that the CTFSP algorithm outperformed other competing algorithms, indicating that our proposed method is a promising framework to enhance the decoding of MI patterns, which is significant for the development of high-performance BCIs.

Conflict of Interest

The authors confirm that there are no known conflicts of interest.

References

[1] J. R. Wolpaw *et al.*, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.

[2] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, Jan. 2008.

[3] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *J. Neurosci. Methods*, vol. 195, no. 2, pp. 270–281, Feb. 2011.

[4] A. Geronimo, Z. Simmons, and S. J. Schiff, "Performance predictors of brain–computer interfaces in patients with amyotrophic lateral sclerosis," *J. Neural Eng.*, vol. 13, no. 2, Apr. 2016, Art. no. 026002.

[5] Y. Miao *et al.*, "An ERP-based BCI with peripheral stimuli: Validation with ALS patients," *Cognit. Neurodynamics*, vol. 14, no. 1, pp. 21–33, Feb. 2020.

[6] Y. Miao *et al.*, "BCI-based rehabilitation on the stroke in sequela stage," *Neural Plasticity*, vol. 2020, pp. 1–10, Dec. 2020.

[7] J. Li *et al.*, "Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom," *Int. J. Neural Syst.*, vol. 24, no. 4, Jun. 2014, Art. no. 1450014.

[8] Y. Yu *et al.*, "Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface," *Comput. Biol. Med.*, vol. 77, pp. 148–155, Oct. 2016.

[9] C. Zhang, H. Wang, and R. Fu, "Automated detection of driver fatigue based on entropy and complexity measures," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 168–177, Feb. 2014.

[10] R. Chai *et al.*, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 3, pp. 715–724, May 2017.

[11] I. Daly, R. Scherer, M. Billinger, and G. Muller-Putz, "FORCe: Fully online and automated artifact removal for brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 725–736, Sep. 2015.

[12] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and dempster-shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 24, 2020 , doi: 10.1109/TNNLS.2020.3015505.

[13] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, Mar. 2019.

[14] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[15] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.

[16] Y. Jiao *et al.*, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.

[17] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain-computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 25, 2020, doi: 10.1109/TNNLS.2020.3016666.

[18] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[19] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[20] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[21] R. Tomioka and K.-R. Müller, "A regularized discriminative framework for EEG analysis with application to brain–computer interface," *NeuroImage*, vol. 49, no. 1, pp. 415–432, 2010.

[22] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[23] H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 653–662, Mar. 2012.

[24] Y. Miao, F. Yin, C. Zuo, X. Wang, and J. Jin, "Improved RCSP and AdaBoost-based classification for motor-imagery BCI," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environments Meas. Syst. Appl. (CIVEMSA)*, Jun. 2019, pp. 1–5.

[25] 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.

[26] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.

[27] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov. 2009.

[28] G. Sun, J. Hu, and G. Wu, "A novel frequency band selection method for common spatial pattern in motor imagery based brain computer interface," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–6.

[29] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP_AM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.

[30] J. Feng *et al.*, "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Netw.*, vol. 102, pp. 87–95, Jun. 2018.

[31] Y. Zhang, C. S. Nam, G. X. Zhou, J. Jin, X. Y. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.

[32] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[33] M. Tangermann *et al.*, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[35] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

[36] T. N. Lal *et al.*, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[37] J. J. Meng, G. Huang, D. G. Zhang, and X. Y. Zhu, "Optimizing spatial spectral patterns jointly with channel configuration for brain–computer interface," *Neurocomputing*, vol. 104, pp. 115–126, Mar. 2013.

[38] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Yang Bang, "Constructing support vector machine ensemble," *Pattern Recognit.*, vol. 36, no. 12, pp. 2757–2767, Dec. 2003.

[39] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.

[40] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2008, pp. 417–420.

[41] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.

[42] Y. Zhang, G. X. Zhou, J. Jin, X. Y. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *J. Neurosci. Methods*, vol. 255, pp. 85–91, Nov. 2015.

[43] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 02, Mar. 2017, Art. no. 1650032.

[44] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[45] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 281–286, Feb. 2002.

[46] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognit.*, vol. 34, no. 2, pp. 299–314, Feb. 2001.

[47] M. Hallett, "Movement-related cortical potentials," *Electromyography Clin. Neurophysiol.*, vol. 34, no. 1, pp. 5–13, 1994.

[48] M. Jahanshahi and M. Hallett, *The Bereitschaftspotential: Movement-Related Cortical Potentials*. New York, NY, USA: Springer, 2003.

[49] T. Liu, G. Huang, N. Jiang, L. Yao, and Z. Zhang, "Reduce brain computer interface inefficiency by combining sensory motor rhythm and movement-related cortical potential features," *J. Neural Eng.*, vol. 17, no. 3, Jun. 2020, Art. no. 035003.

[50] R. Xu *et al.*, "Endogenous sensory discrimination and selection by a fast brain switch for a high transfer rate brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 8, pp. 901–910, Aug. 2016.

[51] R. Xu, N. Jiang, C. Lin, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "Enhanced low-latency detection of motor intention from EEG for closed-loop brain-computer interface applications," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 288–296, Feb. 2014.

[52] K. Wang, M. Xu, Y. Wang, S. Zhang, L. Chen, and D. Ming, "Enhance decoding of pre-movement EEG patterns for brain–computer interfaces," *J. Neural Eng.*, vol. 17, no. 1, 2020, Art. no. 016033.

[53] L. Yao *et al.*, "Common spatial pattern with polarity check for reducing delay latency in detection of MRCP based BCI system," in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 544–547.

[54] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[55] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 051001.