
Survey/review study

Deep Common Spatial Pattern Based Motor Imagery Classification with Improved Objective Function

Nanxi Yu^{1,2}, Rui Yang¹, and Mengjie Huang^{1,*}

¹ School of Electrical Engineering, Electronics & Computer Science, University of Liverpool, Liverpool, L69 3BX, United Kingdom

² Department of Biostatistics, Graduate School of Arts and Sciences, Yale University, New Haven, CT 06511, United States

* Correspondence: Mengjie.Huang@liverpool.ac.uk

Received: 12 October 2022

Accepted: 28 November 2022

Published: 22 December 2022

Abstract: Common spatial pattern (CSP) technique has been very popular in terms of electroencephalogram (EEG) features extraction in motor imagery (MI)-based brain-computer interface (BCI). Through the simultaneous diagonalization of the covariance matrices, CSP intends to transform data into another mapping with data of different categories having maximal differences in their measures of dispersion. This paper shows the objective function realized by original CSP method could be inaccurate by regularizing the estimated spatial covariance matrix from EEG data by trace, leading to some flaws in the features to be extracted. In order to deal with this problem, a novel deep CSP (DCSP) model with optimal objective function is proposed in this paper. The benefits of the proposed DCSP method over original CSP method are verified with experiments on two EEG based MI datasets where the classification accuracy is effectively improved.

Keywords: brain-computer interface; common spatial pattern; electroencephalogram; motor imagery

1. Introduction

Brain-computer interface (BCI) systems provide an alternative communication approach between people's brain and the external world, where brain neurobiological signals are directly translated into devices' control commands that do not rely on the traditional pathways using nerves or muscles [1–4]. Since the first successful record of electroencephalogram (EEG) signals from human scalps in 1920s, it has become one of the most popular input forms of BCI systems [5–8]. The non-invasive EEG recording is relatively low-cost and easy-to-use, and can avoid the risks of invasive methods such as triggering immune responses which results in scar tissues [9]. However, brain signals recorded from the surfaces of the scalps can be vague and weak due to the obstruction of human skulls [10]. Moreover, the quality of EEGs significantly suffers from brain signals' non-stationary and time-varying nature [11–13]. In these circumstances, original EEG signals could be hard to interpret, and suitable feature extraction methods are essential to extract representative and discriminative features from raw collected signals in successful BCI systems [14–16].

Due to the characteristics of EEG signals mentioned above, the conventional time-domain, frequency-domain, and time-frequency-domain analysis to extract EEG features could have some limitations [17]. These methods are generally unsupervised, while there is no definite rule or paradigm to decide the parameters used in these methods [18]. Therefore, the choices of the parameters could be made mostly depending on experts' experience or the results of repetitive experiments [19, 20]. These unsupervised methods may have some limitations in terms of universality, which means independent analysis is required for each specific subject and task [21, 22]. On the contrary, as a supervised feature extraction method, the common spatial pattern (CSP) designs a set of spatial filters based on the given labeled data to maximize the differences in the variances between classes.

CSP was first introduced in 1990 [23] to extract spatial patterns in EEGs from two populations. Researchers [24] further extended the application of this technique to motor imagery (MI) classification problems (MI is a process of rehearsing a motor act internally without actually performing it externally [25]). The physiological basis to classify different MI tasks in BCI systems is the correlation between human's brain electrical activities and human's acts, thoughts, and states [26–28]. By finding the optimal spatial filters to ensure the data from one class having the

maximal variance while the data from the other class having the minimal variance after projection [24], CSP can automatically detect and extract the most discriminative features for the desired tasks based on the labels provided, and has shown stable performances in various experiments.

Due to the satisfactory performance and highly interpretable nature of CSP, many studies have been carried out on the variants of the CSP method to seek possible improvements. Typical approaches to further improve the performance of CSP include selecting data channels, choosing frequency bands, and adding regularization term, with typical literature summarized in Table 1. According to Takens' Embedding Theorem, time delay embedding could help recover the hidden information of a dynamical system [29]. Therefore, Lemm et al. [30] proposed the common spatio-spectral pattern (CSSP) algorithm as a variant of CSP to the state space by embedding a time delay to the original data and concatenating the delayed data to the original data. The experiment results suggested that CSSP algorithm could obtain higher classification accuracy and better generalization ability compared with the original CSP method. Onaran and Ince [31] further improved CSSP by choosing only a subset of the available channels to embed several temporal delays through recursive weight elimination (RWE). This proposed spatially sparse CSSP alleviated the problem of overfitting caused by the increased number of parameters in CSSP. Correlation-based channel selection (CCS) is another approach for channel selection, as the authors of the CCS-CSP method assumed that the channels related to the specific MI tasks should have common features and information while noise and artifacts would appear randomly [32]. Based on this assumption, the channels that had a higher correlation index were selected to compute the CSP features, with experiment results showing that an appropriate channel selection could improve the performance of the CSP method.

Table 1 Variants of CSP Algorithms

Categories	Type of Modification	Typical Literature
Modification of the Data	Channel Selection	Spatially Sparse Common Spatio-Spectral Pattern [31]
		Correlation-based Channel Selection CSP [32]
	Frequency Band Selection	Sub-band CSP (SBCSP) [34]
		Filter Bank CSP (FBCSP) [35]
Regularization	Time Window Selection	Discriminative Filter Bank CSP (DFBCSP) [36]
	Regularized CSP	Wavelet CSP (WCSP) [37]
		Temporally Constrained Sparse Group Spatial Patterns [38]
		Review of Regularized CSP Methods [41]

Apart from selecting data channels [33], some researchers sought to identify the most significant or sensitive frequency sub-bands from raw EEG data. Sub-band CSP (SBCSP) [34] and filter bank CSP (FBCSP) [35] are two typical examples, in which SBCSP analyzed the EEG data on different sub-bands and FBCSP chose features extracted from various frequency sub-bands according to the maximal mutual information criterion. Some researchers concentrated on the study of significant frequency sub-bands at subject level. Discriminative filter bank CSP (DFBCSP) [36] analyzed subject-specific frequency bands to extract more discriminative features. Besides, wavelet common spatial pattern (WCSP) [37] was proposed to choose the most active frequency and resolution for each subject. Zhang et al. [38] suggested that an appropriate selection of time windows for EEG segmentation was also essential to extract discriminative features for each trial. Since EEG signals are extraordinarily non-stationary and sensitive to noise, it is considered that the raw EEG data could be regularized to alleviate such negative effects. A regularization term can be added either in the spatial covariance matrix estimation or CSP's objective function [39, 40]. Lotte and Guan [41] reviewed six existing regularized CSP (RCSP) methods and also proposed four new ones. It was suggested that RCSP was also a practical approach to carry out the subject-to-subject transfer, in addition to improving the classification accuracy.

In recent years, the studies on the CSP method mainly focused on combining CSP with other algorithms or applying to other areas. As mentioned above, the objective of CSP is to find the optimal filters and discriminate data from two classes by maximizing their differences in variances [24]. Based on the analysis of the objective function optimized in the original CSP algorithm in this paper, it is found that the current objective function design is inaccurate through analysis. In certain circumstances, the variance differences are not indeed maximized, and the computed variances fail to represent the most discriminative features of a class, leading to some limitations in feature extraction. However, to the best of author's knowledge, no research has investigated this defect of basic CSP in its principles [17]. Therefore, the objective of this paper is to cope with the problem of inaccurate objective functions of the CSP algorithm. In particular, an improved objective function is presented and a novel deep CSP (DCSP) method is proposed in this paper to extract discriminative MI-EEG features based on the improved objective function. The performance superiority of the proposed DCSP method over the conventional CSP method is verified with two MI based

EEG datasets.

The contributions of this paper are listed below: (1) an improved objective function is proposed in this paper by investigating and analyzing the theoretic limitation of the original CSP algorithm; (2) a novel DCSP method is proposed to extract sensitive and discriminative features from raw EEG signals, aiming to solve the problem of inaccurate objective functions in the original CSP algorithm. The rest of this paper is organized as follows. The objective function inaccuracy problem is explained in detail in Section 2. In Section 3, the methodology adopted in this research is presented, and a novel DCSP method is proposed to deal with the objective function inaccuracy problem. Section 4 demonstrates the experiment and result analysis. A conclusion is drawn in Section 5.

2. Problem Formulation

The CSP method intends to learn spatial filters from the raw EEG data to ensure the projected data from two categories could achieve the maximal variance differences. Let X_1 and X_2 represent the multi-channel EEG epochs of two classes, and both of them have the dimensions of $N * T$, where T is the number of collected samples and N is the number of data collection channels. The estimated spatial covariance matrices of X_1 and X_2 are defined as:

$$R_1 = \frac{X_1 X_1^T}{tr(X_1 X_1^T)}, \quad R_2 = \frac{X_2 X_2^T}{tr(X_2 X_2^T)} \quad (1)$$

where $tr(X)$ is the trace of X and X^T represents the transpose of X .

Then, \overline{R}_1 and \overline{R}_2 are computed by averaging the spatial covariance matrices of all the epochs from each category. The composite covariance matrix is defined as:

$$R = \overline{R}_1 + \overline{R}_2 \quad (2)$$

The composite covariance matrix R can then be factored as:

$$R = U \lambda U^T \quad (3)$$

where U is the eigenvectors matrix and λ is the diagonal matrix of the corresponding eigenvalues. Since the composite covariance matrix R is symmetric, the eigenvectors of R are orthogonal and thus, $U^{-1} = U^T$.

The whitening transformation P equalizing the variances in the eigenvector space can be computed as:

$$P = \sqrt{\lambda^{-1}} U^T \quad (4)$$

Hence, the following relationship between P and R can be obtained:

$$P R P^T = I \quad (5)$$

where I is the identity matrix.

The original CSP transforms \overline{R}_1 and \overline{R}_2 with P and the transformed matrices are:

$$S_1 = P \overline{R}_1 P^T, \quad S_2 = P \overline{R}_2 P^T \quad (6)$$

S_1 and S_2 can be further eigen-decomposed as follows:

$$S_1 = B_1 \lambda_1 B_1^T, \quad S_2 = B_2 \lambda_2 B_2^T \quad (7)$$

Either $B = B_1$ or B_2 could be chosen and the desired spatial filter W is:

$$W = B P \quad (8)$$

Without loss of generality, $W = B_1 P$ is chosen here. Then, the following equation can be obtained:

$$\begin{aligned} I &= B_1^T I B_1 \\ &= B_1^T P R P^T B_1 \\ &= B_1^T P (\overline{R}_1 + \overline{R}_2) P^T B_1 \\ &= B_1^T P \overline{R}_1 P^T B_1 + B_1^T P \overline{R}_2 P^T B_1 \\ &= B_1^T S_1 B_1 + B_1^T S_2 B_1 \\ &= \lambda_1 + B_1^T S_2 B_1 \end{aligned} \quad (9)$$

From (9), it has been shown that W can diagonalize \overline{R}_1 and \overline{R}_2 simultaneously. Moreover, since the summation of two diagonal matrices equals to I , if the order of the eigenvectors in B_1 is rearranged such that the corresponding eigenvalues in λ_1 decreases, the entries on the diagonal of $B_1^T S_2 B_1$ will show an increasing order. Therefore, in the space spanned by the eigenvectors in B_1 , the dimension that accounts for the maximal variance in S_1 will account for the minimal variance in S_2 , which is useful to discriminate between EEG data from two classes. Simi-

larly, discriminative features can also be extracted from the dimension that S_1 has the maximum variance and S_2 has the minimum. Lotte and Guan [41] redefined the CSP method as an optimization problem, and suggested that the objective of CSP method was actually to design the spatial filter W that extremizes:

$$J(W) = \frac{\overline{WR_1W^T}}{\overline{WR_2W^T}} \quad (10)$$

This objective function can be optimized through the Lagrange multiplier method, and the solutions are the eigenvectors of $M = \overline{R_2}^{-1} \overline{R_1}$ corresponding to the smallest and largest eigenvalues. Researchers [42] reviewed these two CSP approaches (the original algebraic operation method and the objective function method) and showed that both approaches were equivalent despite the different procedures. In other words, $J(W)$ is the optimized objective function in the original CSP method based on algebraic operations. However, from the initial design in literature [24], the aim of CSP is to find the optimal spatial filters to ensure the projected data of two categories have the maximal variance differences. Therefore, the objective function in (10) is inaccurate in representing the goal of CSP. By translating the goal directly into mathematical languages, the proper objective function should be:

$$J'(W) = \frac{\overline{R(WX_1)}}{\overline{R(WX_2)}} \quad (11)$$

where $R(\cdot)$ is the operation to compute the spatial covariance matrices. In other words, the EEG data should first be projected before measuring the variances. Using (1) to compute the spatial covariance matrices with same notations as defined above, $J(W)$ and $J'(W)$ can be transformed as follows:

$$J(W) = \frac{\frac{\overline{X_1X_1^T}}{\overline{tr(X_1X_1^T)}^{W^T}}}{\frac{\overline{X_2X_2^T}}{\overline{tr(X_2X_2^T)}^{W^T}}} \quad (12)$$

$$J'(W) = \frac{\frac{\overline{X_1X_1^T}}{\overline{tr(WX_1X_1^TW^T)}^{W^T}}}{\frac{\overline{X_2X_2^T}}{\overline{tr(WX_2X_2^TW^T)}^{W^T}}} \quad (13)$$

The deducing procedures of $\overline{tr(WX_1X_1^TW^T)} = \overline{tr(\lambda^{-1}X_1X_1^T)}$ are shown as follows:

$$\begin{aligned} \overline{tr(WX_1X_1^TW^T)} &= \overline{tr(X_1WW^TX_1^T)} \\ &= \overline{tr(X_1P^TBB^TPIX_1^T)} \\ &= \overline{tr(X_1P^TIPX_1^T)} \\ &= \overline{tr(X_1\lambda^{-1}UU^TX_1^T)} \\ &= \overline{tr(X_1\lambda^{-1}IX_1^T)} \\ &= \overline{tr(\lambda^{-1}X_1X_1^T)} \end{aligned} \quad (14)$$

Therefore, $J'(W)$ can be further transformed as:

$$J'(W) = \frac{\frac{\overline{X_1X_1^T}}{\overline{tr(\lambda^{-1}X_1X_1^T)}^{W^T}}}{\frac{\overline{X_2X_2^T}}{\overline{tr(\lambda^{-1}X_2X_2^T)}^{W^T}}} \quad (15)$$

As shown in (12) and (15), $J(W)$ and $J'(W)$ are not equivalent, and the main difference between them is the existence of variable λ^{-1} within the trace operation. Apparently, such discrepancy between the objective functions is caused by the normalization of traces of spatial covariance matrices. If this trace-normalized estimation is replaced by the unbiased estimation of the covariance matrix, $J(W)$ and $J'(W)$ will become equivalent. However, the normalization based on the matrix trace plays a significant role in balancing the amplitude differences among different trials, and cannot be omitted in order to extract features representing the common characteristics of one class. Under these circumstances, the function $J(W)$, which is optimized in the original CSP method, does not accurately represent the goal of maximizing the variance differences between classes of the projected signals. Aiming to extract more discriminative features raw data, the revised objective function $J'(W)$ considered in this paper is required to maximize

the variance differences in the projected data.

3. Methodology

In this section, the methodologies adopted in the experiments to process raw EEG data and extract EEG features are introduced. In particular, a novel DCSP method is proposed in this paper to solve the problem of inaccurate objective functions and extract sensitive and discriminative EEG features from the preprocessed raw data.

3.1. Signals Pre-Processing

The EEG data are pre-processed to remove the task-unrelated artifacts and enhance the signal-to-noise ratio (SNR). The main steps include the data re-reference, artifacts removal with independent components analysis (ICA) and band-pass filtering. The raw EEG data is first re-referenced with the reference electrodes selected relatively far from the center of the scalp, and the consideration of these locations may receive limited MI-related brain signals.

Based on the assumption of ICA, the recorded EEG data can be modeled as a combination of independent components related to various physiological activities [43], then the task-unrelated artifacts can be removed and the signals of interest are retrieved through ICA to increase the SNR. In this paper, the convenient ICA tool embedded in the Matlab toolbox "EEGLAB" [44] is adopted, with which the artifacts can be conveniently discovered with both numerical coefficients and graphical scalp diagrams to illustrate the decomposed independent components. Figure 1 shows a typical example of the decomposed EEG signals in the dataset utilized in this paper where the first component, as one of the artifacts, is removed.

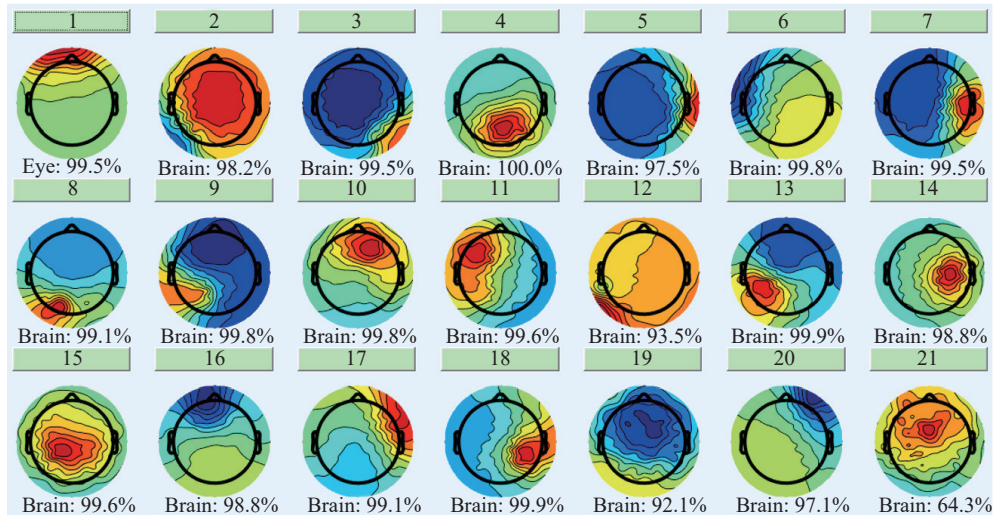


Figure 1. An example of ICA components.

After removing artifacts, the MI-EEG data is filtered precisely to retrieve only the MI tasks-related signals. As the MI event-related desynchronization (ERD) and event-related synchronization (ERS) exist in the alpha and beta rhythm of 8-12Hz and 13-30Hz, respectively [26, 27]. The raw EEG signals are band-pass filtered with a frequency band of 8-30Hz. Considering the reaction time of subjects to start motor imagery after receiving the cue, the last step in signals pre-processing is the selection of analysis time windows to ensure the middle part of each MI epoch which is selected for further analysis. The general procedures to conduct MI-EEG signal pre-processing in this project are summarized in Table 2.

Table 2 Steps of MI-EEG signal pre-processing

Step	Description
1	Re-reference EEG data.
2	Extract MI epochs.
3	Conduct ICA and remove task-unrelated artifacts.
4	Filter the data at 8-30Hz.
5	Select analysis time window.

3.2. Proposed Method

In order to resolve the problem of inaccurate objective functions discussed in Section 2, a novel DCSP method

is designed in this study. The procedures of the DCSP method are explained in detail as follows. After pre-processing, the EEG signals are separated into two sets for training and testing purpose. The traditional CSP filter W can be learned from MI-EEG signals in different trials of the training data. For the multi-channel time-series EEG signals X_{test} of each trial in the testing data, the projected data using CSP can be computed:

$$Z_{test} = WX_{test} \quad (16)$$

The variances of the row values in Z_{test} from two classes monotonously increase and decrease, respectively. Therefore, the projected data from different categories can have the maximal differences in variances in the first and last few rows, leaving the variances in middle rows as indistinguishable. In these cases, the middle parts of Z_{test} can be removed to reduce the computational complexity and alleviate the adverse effects caused by the redundant data.

In the traditional single-layer CSP models, the projected data Z_{test} is usually considered as the feature matrix and the features to be classified. For each trial, y_i can be computed from the feature matrix as follows:

$$y_i = \log(\text{Var}(z_i)) \quad (17)$$

where z_i represents the i^{th} row of Z_{test} .

Compared with traditional CSP models, the proposed DCSP model in this paper does not compute features at this step. The learned CSP filters are applied to the training data X_{train} and the projected data can be computed:

$$Z_{train} = WX_{train} \quad (18)$$

As the input of the following layer, the projected data Z_{train} is further filtered with the CSP method. More specifically, the CSP filters W' are learned on Z_{train} and the projected data in a new space can be computed as follows:

$$Z'_{train} = W'Z_{train} \quad (19)$$

Similar steps can then be repeated for several times. Figure 2 illustrates the framework of three-layer DCSP, while the number of the layers is not limited to three indeed. The extracted feature is obtained by CSP via maximizing the discriminability of two categories, and the support vector machine (SVM) is utilized as the classifier in this method. When the data is satisfactorily filtered, the features for each trial in the training set (for the establishment of SVM classifiers) can be computed as follows:

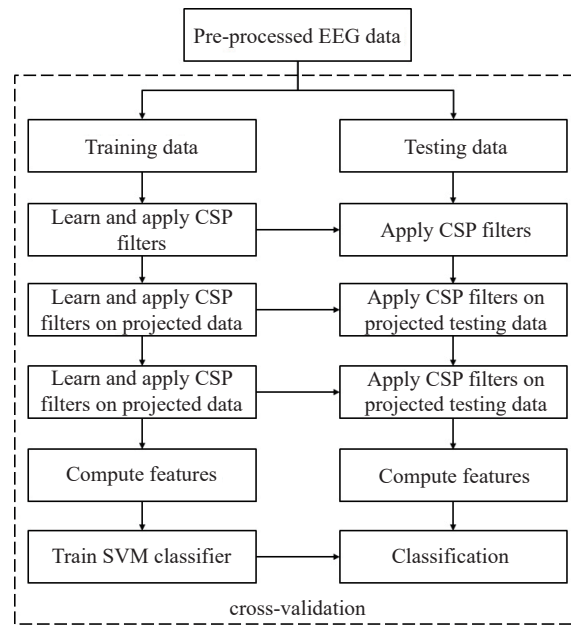


Figure 2. Framework of the DCSP method.

$$y_i = \log(\text{Var}(z_i^{final})) \quad (20)$$

where z_i^{final} represents the i^{th} row of the final feature matrix obtained in the last layer.

The features for the trials in the test set can be computed as follows:

$$y_i = \log(\text{Var}((W^{final} \dots W'WX_{test})_i)) \quad (21)$$

where W^{final} are the filters learned in the last layer, and the subscript i represents the i^{th} row of the projected test data.

The extracted features for each trial in the test set can then be classified by the trained SVM classifier. The framework of the DCSP method is shown in Figure 2.

To compare the feature extraction performance of the CSP and DCSP models, the same number of features to be classified is extracted by CSP and DCSP eventually in each experiment in this paper. In the DCSP structure, twice the output dimension of the next layer are kept in each layer. In other words, if $2m$ features are to be selected in the output of the current layer, the output data of the previous layer would have $4m$ rows. As explained above, the $2m$ features can be calculated from the first m and last m rows in the feature matrix, while the middle $2m$ are removed. The logic behind this structure is to keep twice candidates each time to choose discriminative data. Figure 3 illustrates the dimensions of the data and the spatial filters of each layer in the traditional single-layer CSP and 3-layer DCSP settings.

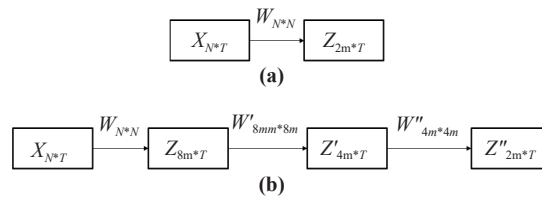


Figure 3. (a) Dimensions of the data in each layer in single-layer CSP; (b) Dimensions of the data in each layer 3-layer DCSP.

The algorithm of the proposed method is summarized in Table 3. Instead of optimizing the desired objective function $J'(W)$ directly, the proposed DCSP repeatedly applies conventional CSP to make the solution of $J(W)$ eventually converge to $J'(W)$. Comparing the objective functions $J(W)$ and $J'(W)$ as shown in (12) and (13), the main difference can be found as the existence of W and W^T in the matrix trace computation.

Table 3 Algorithm of the proposed DCSP method

Step	Algorithm operation
1	Pre-process the collected raw data to enhance SNR;
2	Separate pre-processed data into two datasets for training and testing, respectively;
3	Compute CSP spatial filters W using pre-processed data X_{train} following (1)-(8);
4	Compute projected data using (18) based on training dataset;
5	Keep the first and last rows and remove the middle rows of Z_{train} ;
6	Compute CSP spatial filters W' using data from step 5 based on (1)-(8);
7	Compute projected data using (19);
8	Keep the first and last rows and remove the middle rows of Z' ;
9	Repeat the process from step 6 to step 8 until the data is properly filtered;
10	Extract features from the final feature matrix using (20);
11	Establish SVM classifier with the extracted features from the training data;
12	Apply the filters learned from training data to test data X_{test} and remove middle rows;
13	Compute the features for the test data through (21);
14	Conduct EEG signal classification based on the extracted features from test data using the established SVM classifier.

The principles of DCSP method are explained intuitively in Figure 4. It is intended that the original data could be projected to the objective straightly based on the conventional single-layer CSP algorithm. However, due to the reasons (such as objective function inaccuracy) discussed in the previous section, there would be some deviations between the desired direction and the realized direction. After one time application of the CSP method, the data would be projected to the space labeled as "Step 1" in Figure 4, leaving space for applications of subsequent CSP filters. Then, the data projection can be improved further to "Step 2" by subsequent CSP filtering with minor deviations after several steps. Due to the nature of the CSP method, the deviations could decrease by repeating the steps (such as two or three CSP layers) to eventually reach the desired objective function.

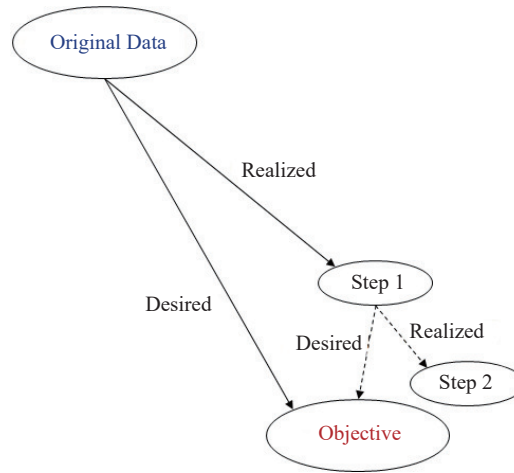


Figure 4. Principles of the DCSP method.

3.3. Multi-Class CSP

CSP method is originally designed to solve binary classification problems. To extend the application to multi-class classification problems, two techniques, namely "one vs. rest" and "one vs. one", are adopted to investigate the four-class MI classification experiment in this paper.

[1] "one vs. rest"

For each class C_i , the idea of the "one vs. rest" method is to design the CSP filters W_i between "this class" and "all the other classes", where the subscript i is the index of the class. Then, the features F_i can be extracted for each trial to represent the discriminative features that class C_i is different from the other classes. The same steps are carried out for all the classes, and the combined features $F = [F_1 F_2 \dots]$ are utilized to classify every single trial.

[2] "one vs. one"

The "one vs. one" method is another approach of multi-class CSP. For each pair C_i and C_j from two categories, the CSP filters $W_{i,j}$ can be learned between these two classes. Despite that the filters are learned only on the two classes, $W_{i,j}$ are applied to all the trials, and the features $F_{i,j}$ are extracted for each trial. In this case, $F_{i,j}$ might be less meaningful for data from a class other than C_i and C_j . However, these features are designed to be kept in order to align the dimensions of the features for each trial. Similarly, the combined features $F = [F_{1,2} F_{1,3} \dots]$ are to be classified. In a four-class classification problem, the combined features is $F = [F_{1,2} F_{1,3} F_{1,4} F_{2,3} F_{2,4} F_{3,4}]$ for each single trial.

Generally, the "one vs. one" method can give higher dimension features compared with the "one vs. rest" method, therefore improving the classification accuracy. However, high dimension features may also lead to over-fitting and negatively affect the model performance in practice.

4. Experiment and Result Analysis

In this paper, to examine the performance of the proposed MI-EEG classification model, two open-access MI datasets are utilized for performance evaluation and comparison between the proposed DCSP method and the conventional CSP method. The information of the datasets is briefly described first, then the experiment results and discussions are presented in this section.

4.1. Dataset Description

Experiments are carried out with two open-access datasets from BCI Competition III (<http://www.bbc.de/competition/iii/>). These two datasets are initially designed for competition purpose to test the performance of EEG classification algorithms, and such datasets are very commonly used by researchers in literature [10, 21].

[1] Dataset IVa [45]

This dataset consists of EEG signals of three MI tasks (left hand, right hand, and right foot) recorded from 118 EEG channels, with only cues of the right hand and foot tasks provided for competition purpose. Five healthy subjects (aa, al, av, aw, and ay) participate in the experiments, and cues of 280 trials are provided for each subject. The timeline of one trial is shown in Figure 5(a), where L, R and F stand for left hand, right hand and right foot respectively. In a single trial, a visual cue is presented for 3.5 seconds, during which the subject is asked to perform the corresponding MI task. Two trials are intermitted by a relaxing period of 1.75 to 2.25 seconds. The original signals are recorded at a sampling rate of 1000Hz and down-sampled to 100Hz by the dataset provider for subsequent analysis.

More details of this dataset can be found on the competition website (http://www.bbc.de/competition/iii/desc_IVa.html).

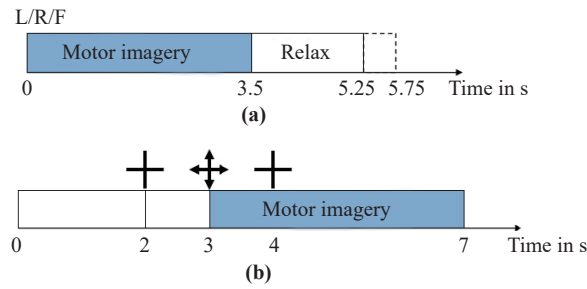


Figure 5. (a) Timeline of one trial in Dataset IVa; (b) Timeline of one trial in Dataset IIIa.

[2] Dataset IIIa [45]

This dataset is recorded over 60 EEG channels at 250Hz from three subjects (k3, k6, and I1). Every subject is instructed to imagine four kinds of movement, namely, MI tasks of the left hand, right hand, tongue, and foot. A complete trial lasts 7 seconds as shown in Figure 5(b). The participants of the experiments can rest in the first 2 seconds. At $t=2s$, an acoustic stimulus and a fixed cross are presented, indicating the beginning of the trial. Then, an arrow pointing to up, down, left, or right is displayed at $t=3s$ over the fixed cross as the cue of the MI task of the tongue, foot, left hand, or right hand, respectively. The arrow disappears at $t=4s$, while the subject should perform the MI task for four seconds until the end of the trial. The numbers of trials for subjects k3, k6, and I1 are 360, 240, and 240, respectively. Detailed descriptions of this dataset are available on the competition website (http://www.bbc.de/competition/iii/desc_IIIa.pdf).

4.2. Results and Discussion

The performance of DCSP models with different layers is examined, where the DCSP model with one layer actually degenerates to the traditional CSP model. Three indices (mean, maximal, and minimal accuracies) are computed based on different dimensions of the features to reflect model performances. For the binary classification dataset IVa, the dimension of the extracted features of a single trial is selected from 2 to 20 with a step size of 2. For each feature dimension in dataset IVa, 10-fold cross-validation is adopted in the classification accuracy computation and verification. For the experiments with dataset IIIa, the 5-fold cross-validation is used to compute and verify the classification accuracy. In the multi-class CSP, 2 to 10 (step size 2) features are extracted by a single CSP filter, resulting in at most 40-dimension and 60-dimension features for the "one vs. rest" and "one vs. one" method, respectively. The SVM classifier is adopted to make final classification decisions using the extracted features. The experiment results of both datasets are provided in Table 4 and Table 5, respectively.

Table 4 Classification performance in experiment on dataset IVa

Subject	aa	al	av	aw	ay	Average
One Layer (mean)	85.11%	98.21%	73.64%	92.89%	94.64%	88.90%
One Layer (max)	86.79%	98.93%	76.07%	94.64%	96.43%	90.57%
One Layer (min)	83.21%	97.86%	72.14%	91.07%	93.21%	87.50%
Two Layers (mean)	85.89%	98.54%	74.64%	94.75%	95.93%	89.95%
Two Layers (max)	87.50%	98.93%	76.79%	96.79%	96.79%	91.36%
Two Layers (min)	84.29%	98.21%	72.50%	93.57%	95%	88.71%
Three Layers (mean)	85.79%	98.61%	74.50%	94.64%	96.14%	89.94%
Three Layers (max)	87.14%	98.93%	76.07%	96.43%	97.14%	91.14%
Three Layers (min)	83.57%	98.21%	72.50%	93.21%	95%	88.50%

Table 5 Classification performance in experiment on dataset IIIa

Subject	1 vs else method				1 vs 1 method			
	K3	K6	L1	Average	K3	K6	L1	Average
One Layer (mean)	93.72%	73.67%	80.67%	82.69%	93.94%	74.25%	80.92%	83.04%
One Layer (max)	94.44%	77.08%	82.5%	84.67%	94.72%	78.33%	81.67%	84.91%
One Layer (min)	92.78%	66.67%	78.33%	79.26%	91.94%	69.17%	80.42%	80.51%
Two Layers (mean)	94.39%	75%	81.92%	83.77%	95.11%	75.67%	81.75%	84.18%
Two Layers (max)	95.83%	79.17%	83.75%	86.25%	96.11%	78.33%	83.33%	85.92%
Two Layers (min)	93.61%	70.42%	80.42%	81.48%	94.17%	69.17%	79.58%	80.97%
Three Layers (mean)	94.61%	75.5%	82.17%	84.09%	95.11%	76.75%	82.08%	84.65%
Three Layers (max)	96.67%	79.17%	84.17%	86.67%	96.11%	79.58%	83.75%	86.48%
Three Layers (min)	93.33%	71.67%	80.83%	81.94%	94.17%	72.5%	80.41%	82.36%

Table 4 demonstrates the performance of the models on the binary classification dataset. Compared with the traditional single-layer CSP model, almost all of the three indices (mean, maximal, and minimal accuracies) of the classification accuracy increase in the multi-layer (including two layers and three layers) designed CSP models. Similar results of the three indices are illustrated in the performances of the models in the four-class dataset as shown in Table 5, where multi-layer designed CSP models generally achieve higher classification accuracy than the original one-layer CSP model. The three-layer DCSP further improves the performance of the two-layer model with dataset IIIa and the performances of the two-layer and three-layer model with dataset IVa are similar. Both of two-layer and three-layer models outperform the single-layer model. In some scenarios, adding an additional layer to the two-layer model does not improve the accuracy further. This situation may owe to the fact that the spatial information contained in the EEG signals is limited. For a simple binary classification, the application of two-layer CSP may extract all the useful spatial features.

In Figure 6, the classification accuracies of two subjects with models of different number of layers and features in dataset IVa are illustrated, with blue, orange, and grey lines representing the one-layer, two-layer, and three-layer CSP models respectively, and the horizontal and vertical axes representing the feature dimension and classification accuracy, respectively. As shown in Figure 6, both the grey lines generally outperform the blue and orange lines when dealing with practical data. By adding extra layers in the conventional CSP models and observing the cross-validation over experiments with different feature dimensions, the statistics in Table 4 and Table 5 illustrate that the averages of the mean classification accuracy improve by at least 1% increment in both datasets. Therefore, it can be concluded that the proposed multi-layer design is beneficial to extract discriminative features and achieve high classification accuracy for EEG signals.

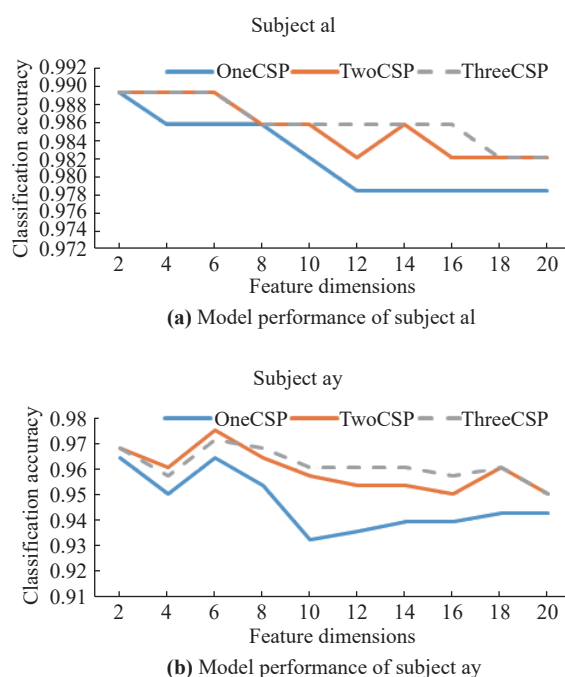


Figure 6. Comparisons of the classification accuracy in dataset IVa.

5. Conclusion

Nowadays, the research of BCI is attracting worldwide attention and is developing rapidly. CSP is currently considered as the most popular method in EEG feature extraction, which aims to extract optimal features for signal classification by maximizing variances of one type signal and minimizing the variance of other type signals via filter design. In this paper, the theoretic limitation of the objective function in original CSP method is discussed and investigated. In order to solve the problem of inaccurate objective functions, an improved objective function is presented and a novel DCSP method is proposed. The experiment results on two MI datasets illustrate that the proposed method can improve the classification performance compared with the original CSP method. The further studies could seek to find other algorithms and learning mechanisms to optimize the revised objective function of CSP in order to enhance the feature extraction in BCI [46, 47, 48].

Data Availability Statement: Restrictions apply to the availability of these data. The data in this study are downloaded from the open access dataset for BCI Competition III (www.bbci.de/competition/iii/, accessed on Nov 21st 2022) with the permission.

Author Contributions: Nanxi Yu: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, writing - original draft, writing - review and editing, validation, visualization. Rui Yang: conceptualization, formal analysis, funding acquisition, investigation, methodology, supervision, writing - original draft, writing - review and editing, validation. Mengjie Huang (Corresponding Author): conceptualization, funding acquisition, project administration, resources, supervision, writing - original draft, writing - review and editing.

Funding: This work is partially supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (20KJB520034), the Jiangsu Provincial Qinglan Project (2021), the Research Development Fund of XJTU (RDF-18-02-30, RDF-20-01-18), the Suzhou Science and Technology Programme (SYG202106) and the Key Program Special Fund in XJTU (KSF-E-34).

Conflicts of Interest: The authors declare no conflict of interest.

Informed Consent Statement: The study is conducted according to the guidelines of the Declaration of Helsinki, and was approved by University Ethics Committee of Xi'an Jiaotong-Liverpool University with proposal number EXT20-01-07 on 31 March 2020.

References

1. Wolpaw, J.R.; Birbaumer, N.; McFarland, D.J.; *et al.* Brain–computer interfaces for communication and control. *Clin. Neurophysiol.*, **2002**, *113*: 767–791.
2. Krepki, R.; Blankertz, B.; Curio, G.; *et al.* The berlin brain-computer interface (BBCI)—towards a new communication channel for online control in gaming applications. *Multimed. Tools Appl.*, **2007**, *33*: 73–90.
3. Huang, M.J.; Zheng, Y.T.; Zhang, J.J.; *et al.* Design of a hybrid brain-computer interface and virtual reality system for post-stroke rehabilitation. *IFAC-Papersonline*, **2020**, *53*: 16010–16015.
4. Wang, L.; Huang, M.J.; Yang, R.; *et al.* Survey of movement reproduction in immersive virtual rehabilitation. *IEEE Trans. Vis. Comput. Graph.* **2022**, in press. doi: [10.1109/TVCG.2022.3142198](https://doi.org/10.1109/TVCG.2022.3142198)
5. Haas, L.F. FHans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *J. Neurol. Neurosurg. Psychiatry*, **2003**, *74*: 9.
6. da Silva, F. L. EEG: Origin and measurement. In EEG-fMRI; Mulert, C.; Lemieux, L., Eds.; Springer: Berlin, 2009; pp. 19–38. doi: [10.1007/978-3-540-87919-0_2](https://doi.org/10.1007/978-3-540-87919-0_2)
7. Lim, S.B.; Louie, D.R.; Peters, S.; *et al.* Brain activity during real-time walking and with walking interventions after stroke: A systematic review. *J. Neuroeng. Rehabil.*, **2021**, *18*: 8.
8. Kline, A.; Ghrooga, C.G.; Pittman, D.; *et al.* EEG differentiates left and right imagined lower limb movement. *Gait Posture*, **2021**, *84*: 148–154.
9. Ahn, M.; Ahn, S.; Hong, J.H.; *et al.* Gamma band activity associated with BCI performance: Simultaneous MEG/EEG study. *Front. Human Neurosci.*, **2013**, *7*: 848.
10. Chen, Y.; Yang, R.; Huang, M.J.; *et al.* Single-source to single-target cross-subject motor imagery classification based on multisub-domain adaptation network. *IEEE Trans. Neural Syst. Rehabil. Eng.*, **2022**, *30*: 1992–2002.
11. Wan, Z.T.; Yang, R.; Huang, M.J.; *et al.* EEG fading data classification based on improved manifold learning with adaptive neighborhood selection. *Neurocomputing*, **2022**, *482*: 186–196.
12. Mathur, P.; Chakka, V.K. Graph signal processing based cross-subject mental task classification using multi-channel EEG signals. *IEEE Sens. J.*, **2022**, *22*: 7971–7978.
13. Chang, Z.Y.; Zhang, C.C.; Li, C. J. Motor imagery EEG classification based on transfer learning and multi-scale convolution network. *Micromachines*, **2022**, *13*: 927.
14. Yang, L.; Song, Y.H.; Ma, K.; *et al.* Motor imagery EEG decoding method based on a discriminative feature learning strategy. *IEEE Trans. Neural Syst. Rehabil. Eng.*, **2021**, *29*: 368–379.
15. Wang, X.Y.; Yang, R.; Huang, M. J. An unsupervised deep-transfer-learning-based motor imagery EEG classification scheme for brain–computer interface. *Sensors*, **2022**, *22*: 2241.
16. Wang, Q.H.; Liu, F.; Wan, G.H.; *et al.* Inference of brain states under anesthesia with meta learning based deep learning models. *IEEE Trans. Neural Syst. Rehabil. Eng.*, **2022**, *30*: 1081–1091.

17. Wan, Z.T.; Yang, R.; Huang, M.J.; *et al.* A review on transfer learning in EEG signal analysis. *Neurocomputing*, **2021**, *421*: 1–14.
18. Ji, D.A.; Wang, C.; Li, J.H.; *et al.* A review: Data driven-based fault diagnosis and RUL prediction of petroleum machinery and equipment. *Syst. Sci. Control Eng.*, **2021**, *9*: 724–747.
19. Wang, Y.Z.; Zou, L.; Ma, L.F.; *et al.* A survey on control for Takagi-Sugeno fuzzy systems subject to engineering-oriented complexities. *Syst. Sci. Control Eng.*, **2021**, *9*: 334–349.
20. Hu, J.; Jia, C.Q.; Liu, H.J.; *et al.* A survey on state estimation of complex dynamical networks. *Int. J. Syst. Sci.*, **2021**, *52*: 3351–3367.
21. Wan, Z.T.; Yang, R.; Huang, M.J.; *et al.* Segment alignment based cross-subject motor imagery classification under fading data. *Comput. Biol. Med.*, **2022**, *151*: 106267.
22. Zhang, X.X.; She, Q.S.; Chen, Y.; *et al.* Sub-band target alignment common spatial pattern in brain-computer interface. *Comput. Methods Programs Biomed.*, **2021**, *207*: 106150.
23. Koles, Z.J.; Lazar, M.S.; Zhou, S. Z. Spatial patterns underlying population differences in the background EEG. *Brain Topogr.*, **1990**, *2*: 275–284.
24. Müller-Gerking, J.; Pfurtscheller, G.; Flyvbjerg, H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, **1999**, *110*: 787–798.
25. Decety, J. The neurophysiological basis of motor imagery. *Behav. Brain Res.*, **1996**, *77*: 45–52.
26. Pfurtscheller, G.; Aranibar, A. Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalogr. Clin. Neurophysiol.*, **1977**, *42*: 817–826.
27. Pfurtscheller, G.; da Silva, F. L. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.*, **1999**, *110*: 1842–1857.
28. Krause, C.M.; Pörn, B.; Lang, A.H.; *et al.* Relative alpha desynchronization and synchronization during speech perception. *Cognit. Brain Res.*, **1997**, *5*: 295–299.
29. Ju, Y.M.; Tian, X.; Liu, H.J.; *et al.* Fault detection of networked dynamical systems: A survey of trends and techniques. *Int. J. Syst. Sci.*, **2021**, *52*: 3390–3409.
30. Lemm, S.; Blankertz, B.; Curio, G.; *et al.* Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans. Biomed. Eng.*, **2005**, *52*: 1541–1548.
31. Onaran, I.; Ince, N.F. Extraction of spatially sparse common spatio-spectral filters with recursive weight elimination. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering, San Diego, 06–08 November 2013; IEEE: San Diego, 2013; pp. 1291–1294. doi: [10.1109/NER.2013.6696177](https://doi.org/10.1109/NER.2013.6696177)
32. Jin, J.; Miao, Y.Y.; Daly, I.; *et al.* Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw.*, **2019**, *118*: 262–270.
33. Geng, H.; Liu, H.J.; Ma, L.F.; *et al.* Multi-sensor filtering fusion meets censored measurements under a constrained network environment: Advances, challenges and prospects. *Int. J. Syst. Sci.*, **2021**, *52*: 3410–3436.
34. Novi, Q.; Guan, C.T.; Dat, T.H.; *et al.* Sub-band common spatial pattern (SBCSP) for brain-computer interface. In Proceedings of the 2007 3rd International IEEE/EMBS Conference on Neural Engineering, Kohala Coast, 02–05 May 2007; IEEE: Kohala Coast, 2007; pp. 204–207. doi: [10.1109/CNE.2007.369647](https://doi.org/10.1109/CNE.2007.369647)
35. Ang, K.K.; Chin, Z.Y.; Zhang, H.H.; *et al.* Filter bank common spatial pattern (FBCSP) in brain-computer interface. In Proceedings of 2008 IEEE International Joint Conference on Neural Networks, Hong Kong, China, 01–08 June 2008; IEEE: Hong Kong, China, 2008; pp. 2390–2397. doi: [10.1109/IJCNN.2008.4634130](https://doi.org/10.1109/IJCNN.2008.4634130)
36. Higashi, H.; Tanaka, T. Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification. *IEEE Trans. Biomed. Eng.*, **2013**, *60*: 1100–1110.
37. Mousavi, E.A.; Maller, J.J.; Fitzgerald, P.B.; *et al.* Wavelet common spatial pattern in asynchronous offline brain computer interfaces. *Biomed. Signal Process. Control*, **2011**, *6*: 121–128.
38. Zhang, Y.; Nam, C.S.; Zhou, G.X.; *et al.* Temporally constrained sparse group spatial patterns for motor imagery BCI. *IEEE Trans. Cybern.*, **2019**, *49*: 3322–3332.
39. Cheng, H.J.; Wang, Z.D.; Wei, Z.H.; *et al.* On adaptive learning framework for deep weighted sparse autoencoder: A multiobjective evolutionary algorithm. *IEEE Trans. Cybern.*, **2022**, *52*: 3221–3231.
40. Sun, J.Y.; Wang, Z.D.; Yu, H.; *et al.* Two-stage deep regression enhanced depth estimation from a single RGB image. *IEEE Trans. Emerging Topics Comput.*, **2022**, *10*: 719–727.
41. Lotte, F.; Guan, C. T. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Trans. Biomed. Eng.*, **2011**, *58*: 355–362.
42. He, H.; Wu, D.R. Spatial filtering for brain computer interfaces: A comparison between the common spatial pattern and its variant. In Proceedings of 2018 IEEE International Conference on Signal Processing, Communications and Computing, Qingdao, 14–16 September 2018; IEEE: Qingdao, 2018; pp. 1–6. doi: [10.1109/ICSPCC.2018.8567789](https://doi.org/10.1109/ICSPCC.2018.8567789)
43. Kachenoura, A.; Albera, L.; Senhadji, L.; *et al.* ICA: A potential tool for BCI systems. *IEEE Signal Process. Mag.*, **2008**, *25*: 57–68.
44. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, **2004**, *134*: 9–21.
45. Blankertz, B.; Müller, K.R.; Krusienski, D.J.; *et al.* The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.*, **2006**, *14*: 153–159.
46. Song, B.Y.; Miao, H.M.; Xu, L. Path planning for coal mine robot via improved ant colony optimization algorithm. *Syst. Sci. Control Eng.*, **2021**, *9*: 283–289.
47. Xu, L.; Song, B.Y.; Cao, M. Y. An improved particle swarm optimization algorithm with adaptive weighted delay velocity. *Syst. Sci. Control Eng.*, **2021**, *9*: 188–197.
48. Jia, X. C. Resource-efficient and secure distributed state estimation over wireless sensor networks: A survey. *Int. J. Syst. Sci.*, **2021**, *52*: 3368–3389.

Citation: Yu, N.; Yang, R.; Huang, M. Deep Common Spatial Pattern Based Motor Imagery Classification with Improved Objective Function. *International Journal of Network Dynamics and Intelligence*. <https://doi.org/10.53941/ijndi0101007>

Publisher’s Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license <https://creativecommons.org/licenses/by/4.0/>.