




Capped L21-norm-based common spatial patterns for EEG signals classification applicable to BCI systems

Jingyu Gu¹ · Jiuchuan Jiang² · Sheng Ge¹ · Haixian Wang¹ 

Received: 14 June 2022 / Accepted: 6 January 2023 / Published online: 20 January 2023
© International Federation for Medical and Biological Engineering 2023

Abstract

The common spatial patterns (CSP) technique is an effective strategy for the classification of multichannel electroencephalogram (EEG) signals. However, the objective function expression of the conventional CSP algorithm is based on the L2-norm, which makes the performance of the method easily affected by outliers and noise. In this paper, we consider a new extension to CSP, which is termed capped L21-norm-based common spatial patterns (CCSP-L21), by using the capped L21-norm rather than the L2-norm for robust modeling. L21-norm considers the L1-norm sum which largely alleviates the influence of outliers and noise for the sake of robustness. The capped norm is further used to mitigate the effects of extreme outliers whose signal amplitude is much higher than that of the normal signal. Moreover, a non-greedy iterative procedure is derived to solve the proposed objective function. The experimental results show that the proposed method achieves the highest average recognition rates on the three real data sets of BCI competitions, which are 91.67%, 85.07%, and 82.04%, respectively.

Jingyu Gu and Jiuchuan Jiang contributed equally.

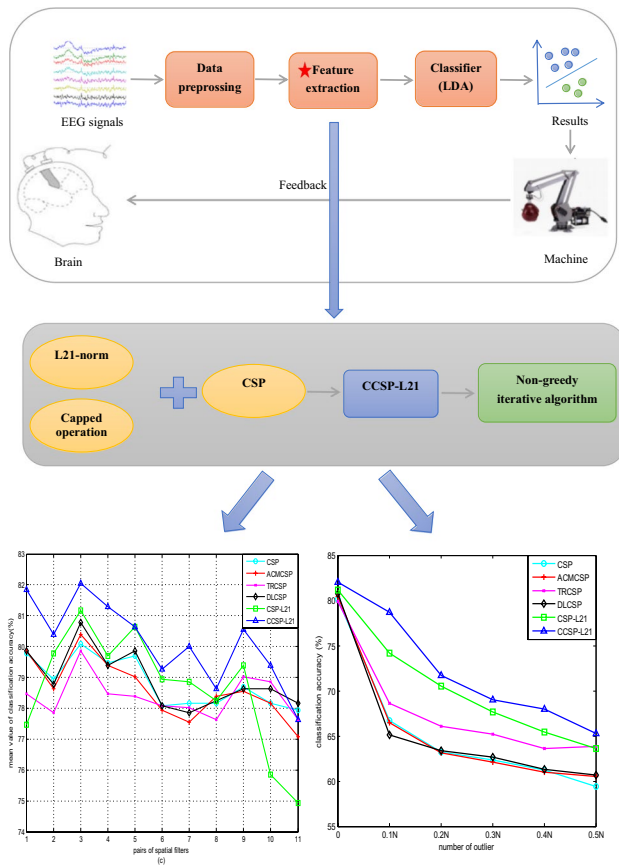
✉ Haixian Wang
hxwang@seu.edu.cn

¹ Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, Jiangsu, People's Republic of China

² School of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210003, Jiangsu, People's Republic of China

Graphical abstract

Capped L21-norm-based common spatial patterns—a robust model for EEG signals classification



Capped L21-Norm-Based Common Spatial Patterns—a Robust Model for EEG Signals Classification

Keywords Brain-computer interfaces (BCI) · Common spatial patterns (CSP) · Capped L21-norm · Robust modeling

1 Introduction

Brain-computer interface (BCI) is an advanced communication system designed to bypass muscle tissue and establish communication pathways between the brain and the external environment, thus effectively decoding the brain's state of mind [1]. In recent years, with the interdisciplinary integration of computer science, brain science, neurobiology, and intelligent control, the application of BCI technology has expanded from the medical and health fields to the entertainment, education, intelligent home, and even military fields. A complete BCI system is generally composed of the following four parts: signal acquisition

module, signal processing module, control signal output module, and feedback module [2].

Electroencephalography (EEG) is widely used to measure neurophysiological activity in the signal acquisition module because of its noninvasiveness, high temporal resolution, and inexpensive recording device [3]. However, EEG-based BCI systems also have some drawbacks that cannot be ignored, such as non-stationary nature, the characteristics of non-linearity, and low signal-to-noise ratio (SNR). There is no doubt that the shortcomings mentioned above lead to higher requirements for subsequent signal processing technology. Thus, seeking robust feature extraction methods has become a key issue in recent research [4].

Common spatial patterns (CSP), as one of the popular and representative spatial filtering algorithm, is good at extracting features associated with motor imagery tasks. In recent years, with the wide application of CSP algorithm, researchers have put forward many new ideas to further optimize the performance of BCI system. For example, Afrakhteh et al. indicated that higher recognition accuracy can be achieved by using CSP algorithm for feature extraction and dimension reduction, combined with some evolutionary approach and improved neural network algorithm [5–8]. Many scholars combined CSP algorithm with many classical methods, such as Mel-frequency-based CSP (MF-CSP) [9], CSP combined with wavelet packet decomposition (wavelet-CSP) [10], and improved common spatial patterns (B-CSP) [11]. Besides, Rahman et al. proposed multiclass CSP (M-CSP) [12] by extending CSP from two to multiple classes.

In fact, the classical CSP itself also has some drawbacks. Although the traditional CSP algorithm is simple and efficient, its covariance matrix estimation is based on the square of Euclidean distance, which makes the performance of this method vulnerable to outliers and noise [13]. In order to improve the robustness and sparsity of the CSP algorithm, some extensions have been put forward by modifying its objective function, such as L1-norm-based CSP (CSP-L1) [14, 15], sparse CSP-L1 (sp-CSPL1) [16], regularized CSP-L1 with a waveform length (wlCSPL1) [17], local temporal CSP (LTCSP) [18], local temporally correlated CSP (LTC-CSP) [19], local temporal joint-recurrence CSP (LTRCSP) [20], and Lp-norm-based local temporally correlated CSP (LTCCSP-Lp) [21]. Among them, whereas the extensions based on L1-norm are popular and have been able to seek robust spatial filters effectively, the L1-norm is unable to characterize the geometric structure of the data well, and the absolute value operator makes the calculation difficult. Therefore, we were inspired to use the L21-norm which has the advantages of rotational invariance and geometric structure characterization. L21-norm-based CSP (CSP-L21) [22] and regularized CSP with the L21-norm (RCSP-L21) [23] are proposed accordingly.

In this paper, we consider a new algorithmic form with better stability and robustness by replacing the L2-norm with the capped L21-norm. This method, called the capped L21-norm-based CSP (CCSP-L21), is motivated by ideas underpinning some classical pattern recognition algorithms in the machine learning field. Compared with other extensions to CSP, CCSP-L21 has two main highlights as follows. On the one hand, by employing the L21-norm as the basic metric, we enhance the robustness of our new approach to achieve better classification performance. In fact, this enhancement is achieved by removing the square operator and has been applied in many feature selection algorithms, such as the rotational invariant L1-norm principal component analysis (R1-PCA) [24, 25], discriminant analysis via joint Euler transform and L21-norm (e-LDA-L21) [26], and L21-norm-based discriminant locality preserving projections (L21-DLPP) [27]. On the other hand, to further reduce the negative impact of some outliers with large amplitudes that appear during the signal acquisition process, we apply the capped norm to our new approach. Recently, some studies have also shown that methods integrating capped norms can obtain more discriminative features. For example, Lai et al. [28] presented a robust locally discriminant analysis via capped norm (RLDA) by mixing the L21-norm, capped norm, regularized term, and local structure information. Moreover, Wang et al. [29] proposed capped Lp-norm linear discriminant analysis (CappedLDA) to enhance the robustness of the algorithm.

The remainder of this paper is organized as follows. We define some notations and briefly introduce the conventional CSP in Sect. 2. In Sect. 3, the CCSP-L21 is presented, and the corresponding non-greedy iterative algorithm is introduced. We carry out a set of experiments on the three real EEG data sets and discuss the results in Sect. 4. Finally, Sect. 5 is the summary of the paper.

2 Brief review of conventional CSP

As one of the most commonly used feature extraction methods, common spatial patterns (CSP) have good performance in the classification of multichannel EEG signals [30]. As we all know, it is generally applied to a two-class paradigm. Let $X^1, X^2, \dots, X^{t_x} \in R^{C \times N}$ be the EEG signals of one mental task, while $Y^1, Y^2, \dots, Y^{t_y} \in R^{C \times N}$ be the other condition. Among the notations above, C represents the number of electrodes (channels), N is the number of recording time points in a trial, and t_x and t_y denote the numbers of trials that belong to the two classes, respectively. For the sake of expression, the columns of X and Y are relabeled as $X = (x_1, x_2, \dots, x_m) \in R^{C \times m}$ ($m = N \times t_x$) and $Y = (y_1, y_2, \dots, y_n) \in R^{C \times n}$ ($n = N \times t_y$). Here, m and n represent the numbers of sampled points from the two brain

states. In addition, the trial segments are assumed to have already gone through the filtering of a specific frequency band, the decentralization of the mean value, and the preprocessing of normalization [31].

The CSP algorithm aims to find an optimal spatial filter $w \in R^C$ that projects multichannel EEG signals into a new space such that the variance of one class is maximized while that of the other class is minimized. Mathematically, the objective function can be given as follows:

$$J_{\text{CSP}}(w) = \frac{w^T C^x w}{w^T C^y w} \tag{1}$$

Here, the covariance matrices of the two classes $C^x \in R^{C \times C}$ and $C^y \in R^{C \times C}$ can be calculated as Eqs. (2) and (3), respectively:

$$C^x = \frac{1}{t_x} X X^T \tag{2}$$

$$C^y = \frac{1}{t_y} Y Y^T \tag{3}$$

where T is a transpose operator. The solution of the objective function is essentially a generalized eigenvalue problem, which can be solved by the following equation:

$$C^x w = \lambda C^y w \tag{4}$$

where the eigenvalue λ represents the ratio of the variances of the two classes.

Finally, we select the few leading eigenvectors associated with the largest and smallest eigenvalues as spatial filters. Then, the normalized log-variances of these components are used as features, which are fed into the linear discriminant analysis (LDA) classifier.

3 Capped L21-norm-based common spatial patterns (CCSP-L21)

It is clear that the objective function expression of the traditional CSP algorithm is based on the square of Euclidean distance, which makes the performance of the method easily affected by outliers and noise. To address this problem, a new robust extension is considered in the paper. We term it capped L21-norm-based common spatial patterns (CCSP-L21).

In this section, we present the new objective function of our proposed method first. Then, a non-greedy iterative algorithm [32] is designed to solve the optimization problem. At last, the suitable features are extracted for classification.

3.1 Objective function

For the convenience of calculation, we rewrite the objective function of the classical CSP by substituting Eqs. (2) and (3) into Eq. (1), which is shown as Eq. (5):

$$J_{\text{CSP}}(w) = \frac{w^T C^x w}{w^T C^y w} = \frac{\frac{1}{t_x} \|w^T X\|_2^2}{\frac{1}{t_y} \|w^T Y\|_2^2} = \frac{\frac{1}{t_x} \sum_{i=1}^m (w^T x_i)^2}{\frac{1}{t_y} \sum_{j=1}^n (w^T y_j)^2} \quad (5)$$

where $\|\cdot\|$ denotes the L2-norm.

To obtain more discriminative features, the objective function can be further reformulated by the capped L21-norm as follows:

$$J_{\text{CCSP-L21}}(W) = \frac{\|W^T X\|_{\text{cap21}}}{\|W^T Y\|_{\text{cap21}}} = \frac{\sum_{i=1}^m \min(\|W^T x_i\|_2, \varepsilon)}{\sum_{j=1}^n \min(\|W^T y_j\|_2, \varepsilon)} \quad (6)$$

where $\|\cdot\|_{\text{cap21}}$ denotes the capped L21-norm, ε ($\varepsilon > 0$) is a thresholding parameter that is used to pick out the extreme data outliers, $W \in R^{C \times d}$ ($d < C$) is an optimal projection matrix for dimension reduction, and d represents the number of extracted features for classification.

According to simple algebraic theory, objective function (6) is equal to the following formulation:

$$J_{\text{CCSP-L21}}(W) = \frac{\text{tr}(W^T X D_x X^T W)}{\text{tr}(W^T Y D_y Y^T W)} \quad (7)$$

$$D_x = \text{diag}\left(\frac{\text{Ind1}}{\|x_1\|_2}, \frac{\text{Ind1}}{\|x_2\|_2}, \dots, \frac{\text{Ind1}}{\|x_d\|_2}\right) \quad (8)$$

$$D_y = \text{diag}\left(\frac{\text{Ind2}}{\|y_1\|_2}, \frac{\text{Ind2}}{\|y_2\|_2}, \dots, \frac{\text{Ind2}}{\|y_d\|_2}\right) \quad (9)$$

where $\text{tr}(\cdot)$ is the trace operator and Ind1 and Ind2 represent the indicative functions defined as Eqs. (10) and (11):

$$\text{Ind1} = \begin{cases} 1 & \text{if } \|x_d\| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$\text{Ind2} = \begin{cases} 1 & \text{if } \|y_d\| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

3.2 Iterative algorithm

Obviously, it is difficult to find the solution to the objective function of our proposed approach. To obtain the optimal projection matrix W , we consider a non-greedy iterative algorithm by constructing an auxiliary function, with

the assistance of the alternating renewal process, subgradient algorithm, and Armijo line search method.

The following theorem is introduced to provide an auxiliary function for objective optimization:

Theorem 1: Suppose that the matrix functions $M(U)$ and $N(U)$ are positive definite, we have:

$$\lambda_{\max} = \frac{M(U^*)}{N(U^*)} = \max_{U^T U = I_p} \frac{M(U)}{N(U)} \quad (12)$$

if and only if:

$$M(U^*) - \lambda_{\max} N(U^*) = \max(M(U) - \lambda_{\max} N(U)) = 0 \quad (13)$$

Thus, objective function (7) can be inverted into the following equation with the form of the corresponding trace difference:

$$W_{\text{opt}} = \arg \max_{W^T W = I_d} \frac{\|W^T X\|_{\text{cap21}}}{\|W^T Y\|_{\text{cap21}}} = \arg \max_{W^T W = I_d, \lambda} (\|W^T X\|_{\text{cap21}} - \lambda \|W^T Y\|_{\text{cap21}}) \quad (14)$$

Because the specific derivation and the convergence proof of the iterative process have been mentioned in the article [22], the entire iteration steps are directly given here to avoid repetition, as shown in Table 1.

3.3 Feature extraction

The optimal projection matrix W obtained by the above iterative algorithm can be regarded as a set of multiple orthogonal spatial filters. Therefore, we relabel the columns of W as w_1, w_2, \dots, w_d . Suppose that Z denotes the EEG trial, and feature f is extracted as:

$$f = (\|w_1 Z\|_2, \|w_2 Z\|_2, \dots, \|w_d Z\|_2)^T \quad (15)$$

where d represents the number of spatial filters.

4 Experiment

In the experiments, we use three public BCI competition-based EEG, data sets IIIa and IVa of BCI competition III, and data set IIa of BCI competition IV [33], to prove the effectiveness of the proposed CCSP-L21 approach. In addition, other extensions to the original CSP methods are also introduced for comparison purposes. Afterwards, we compare the performances of all methods when outliers of different frequencies occur. It should be mentioned that linear discriminant analysis (LDA) is used as a classifier to evaluate the algorithm performance here.

Table 1 Iterative algorithm procedure of CCSP-L21

Input:
 Two classes of EEG data matrices X and Y , Armijo line search parameter β ($0 < \beta < 1$), thresholding parameter ε ($\varepsilon > 0$) and iteration number t .
 Initialize $W_0 \in R^{C \times d}$ randomly, satisfying $W_0^T W_0 = I_d$

Output:
 Optimal spatial filter matrix W

Procedure:
 1) Set $t = 1$
 2) Suppose the solution of the $(t-1)$ th iteration has been obtained, which is W_{t-1} . Calculate λ_t by using the equation: $\lambda_t = \frac{\|W_{t-1}\|^T X\|_{\text{cap}2,1}}{\|W_{t-1}\|^T Y\|_{\text{cap}2,1}}$
 3) According to Eq. (14), transform the fixing problem of λ_t into optimizing the formula: $\arg \max_{W^T W = I_d} (\|W^T X\|_{\text{cap}2,1} - \lambda_t \|W^T Y\|_{\text{cap}2,1})$. Mark the formula in the bracket as $F(W)$
 4) Calculate the subgradient of $F(W)$: $\nabla F(W_{t-1}) = X D^X X^T W_{t-1} - \lambda_t Y D^Y Y^T W_{t-1}$. Here, D^X and D^Y are all diagonal matrices, satisfying: $D^X = \text{diag}(\frac{\text{Ind}1}{\|W^T X(:,1)\|_2}, \frac{\text{Ind}1}{\|W^T X(:,2)\|_2}, \dots, \frac{\text{Ind}1}{\|W^T X(:,d)\|_2})$, $D^Y = \text{diag}(\frac{\text{Ind}2}{\|W^T Y(:,1)\|_2}, \frac{\text{Ind}2}{\|W^T Y(:,2)\|_2}, \dots, \frac{\text{Ind}2}{\|W^T Y(:,d)\|_2})$, where $\text{Ind}1$ and $\text{Ind}2$ are represented as Eqs. (10) and (11), respectively
 5) Update $W_{\text{new}} = W_{t-1} + \beta \cdot \nabla F(W_{t-1})$, which satisfies the orthogonalization operation
 6) Calculate $F(W_{\text{new}})$. If $F(W_{\text{new}}) \geq F(W_{t-1}) = 0$, let $W_t = W_{\text{new}}$ and go to step 7; otherwise, update $\beta = \frac{\beta}{1+\alpha}$ ($\alpha \in [0, 1]$) and go to step 5
 7) Set $t \leftarrow t + 1$
 8) Stop the iteration until convergence

4.1 Real EEG data sets

The three real data sets record EEG signals, while the subjects imagine limb movements (e.g., hand or foot movements) [34]. What needs illustration is that only the classifications of the data for two classes are considered in the experiment. The detailed statistical information of the three publicly available data sets is summarized in Table 2.

4.2 Preprocessing of the EEG signals

For the three data sets introduced above, the raw EEG signals need a series of preprocessing operations before the experiment. The original signals are first filtered by a fifth-order Butterworth filter with cutoff frequencies of 8 and 35 Hz composing both the α -band and β -band, respectively. There is an optimal time period for an EEG to detect an event-related synchronization (ERS) or an event-related desynchronization (ERD). Thus, the EEG segments recorded from 0.5 s to 2.5 s after the visual cue are chosen for the first and third data sets. Specifically, inspired by the winner of BCI competition IV and the relevant pretreatment method

mentioned in the corresponding article [35], we use a time interval from 0.5 s to 3.75 s on the second data set.

4.3 Experimental settings

The CCSP-L21 algorithm involves three parameters: the line search parameters β and α and the thresholding parameter ε . The set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ is designed for β empirically, while α randomly takes a value between 0 and 1. Note that we need to run the program ten times to ensure the stability of the algorithm. Based on the experience in article [29], the set $\{1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2\}$ for thresholding parameter ε is selected on an approximate scale. In addition, as one of the comparison methods, the TRCSP algorithm has regularization parameters, which are searched in set $\{1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e1, 1e2\}$ by a tenfold cross-validation.

In particular, the pairs of filters in the experiment vary from 1 to $0.5 \times C$ here rather than using one fixed value. Then, d -D feature vectors, which are set as an input to the

Table 2 Detailed statistical information of the three real EEG data sets used for the experiment

Data set	BCI competition III								BCI competition IV	
	Data set IIIa			Data set IVa					Data set IIa	
Subject	s1	s2	s3	aa	al	av	aw	ay	A01E-A09E	
No. of training trials	90	60	60	168	224	84	56	28	144	
No. of testing trials	90	60	60	112	56	196	224	252	144	
No. of total trials	180	120	120	280	280	280	280	280	288	
No. of channels	60			118						22
Sample frequency	250 Hz			100 Hz						250 Hz
MI classification	Left vs. right hand			Right hand vs. foot						Left vs. right hand

LDA classifier, can be obtained, where d denotes the number of filters.

4.4 Outlier simulation

In order to further verify the robustness of the algorithm, a C -dimensional Gaussian distribution $N(m + 3\sigma, \Sigma)$ is used to generate outliers with the numbers varying from 0 to $0.5N$ with step $0.1N$. Here, m represents the mean vector, σ is the standard deviation vector of the EEG training data, Σ denotes the covariance matrix of the EEG training samples, and N is the number of recording time points.

4.5 Results and discussion

In this section, the performance of the proposed CCSP-L21 algorithm is verified by comparing with five relevant methods on the three public BCI competition EEG data sets mentioned above. Other than the classical CSP algorithm, we also use some other extensions for comparison, including the CSP with Tikhonov regularization (TRCSP) [35], the CSP with weighted average covariance matrix (ACMCSP) [36], the regularized CSP based on diagonal loading (DLCSP) [37], and the L21-norm-based CSP (CSP-L21) [22].

Figure 1 displays the average recognition rates of the above six methods as the pair of filters change. It can be seen that the blue curve representing the classification accuracies of the CCSP-L21 algorithm is above the curves of the other algorithms in most cases. In addition, our proposed approach also achieves the highest recognition rates, which is enough to show its superior performance in the task of recognizing motor imagery (MI)-based EEG signals.

Thus, we can obtain the corresponding filter pairs of the six methods when the optimal recognition rates are reached on the three real data sets. For the three individuals in data set IIIa of BCI competition III, the optimal spatial filter pairs for the six methods CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21 are 2, 3, 8, 2, 5, and 4, respectively; 3, 1, 1, 2, 2, and 3 filter pairs are selected for data set IVa of BCI competition III, which has five subjects. On data set IIa of BCI competition IV, the best accuracy can be achieved by applying three filter pairs for all the above methods.

Next, as shown in Tables 3 and 4, the optimal classification accuracies of each algorithm on the three data sets are calculated, and the data with the best performance in each subject are represented in bold to make the results easy to observe. It should be noted that the last line lists the results of the BCI winners just for the integrity of the results rather than as a comparison.

Clearly, the classical CSP and the other four extensions have their advantages in some subjects. However, the proposed CCSP-L21 algorithm performs better than the other methods under most circumstances. For some subjects, such as s1, s3, a1, and A08E, the recognition rates are above 98%. Among them, the rates of individuals s1 and a1 even reach 100%. Compared with the traditional CSP, the mean classification accuracies of CCSP-L21 increase by approximately 3.15%, 4.41%, and 1.95%, which fully demonstrates the effectiveness of the capped L21-norm. Besides, the performance of CCSP-L21 algorithm has also been improved to some extent by comparing with CSP-L21, which demonstrates the effectiveness of introducing the capped norm.

Afterwards, the robustness of the CCSP-L21 algorithm should be further evaluated with the addition of artificial outliers. By observing Fig. 2, we can see the curve of the

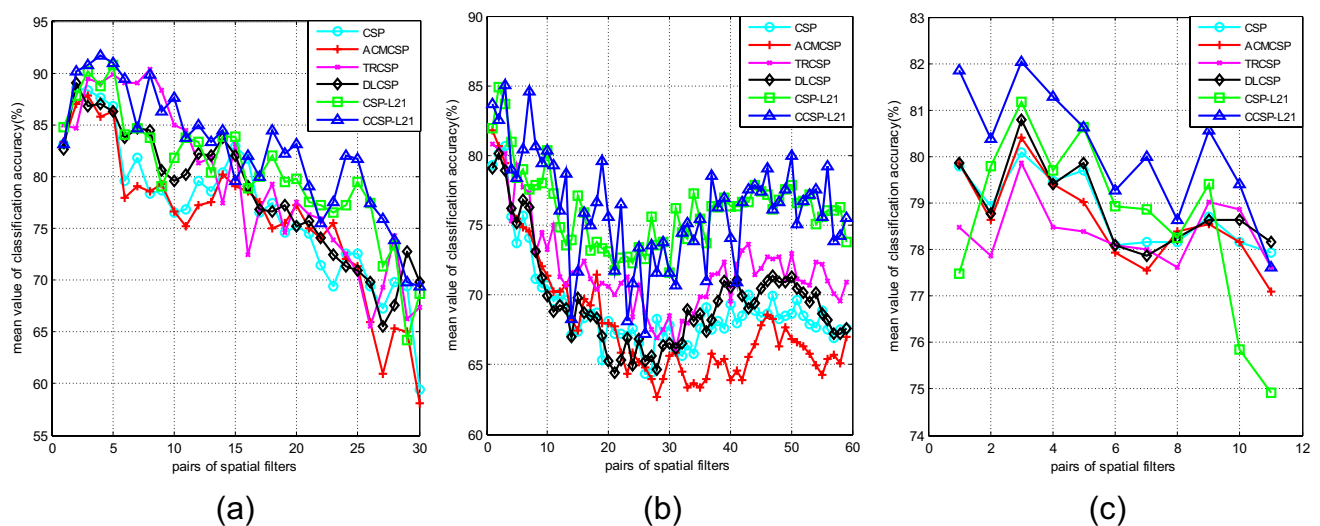


Fig. 1 The average classification accuracies as the pairs of spatial filters change by six different extensions to CSP, i.e., classical CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21. **a** Data set IIIa, **b** IVa of BCI competition III, and **c** data set IIa of BCI competition IV

Table 3 Classification accuracies of CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21 on the subjects of data sets IIIa and IVa of BCI competition III without outliers added. The BCI winner values with underline on data set IIIa of BCI competition III are the kappa scores. Values in bold indicate the best recognition rate for each subject

Data set	BCI competition III									
	Data set IIIa				Data set IVa					
	Subject	s1	s2	s3	Mean	aa	al	av	aw	ay
CSP	98.89	68.33	98.33	88.52	69.64	100.00	65.31	93.75	74.60	80.66
ACMCSP	96.67	70.00	96.67	87.78	65.18	100.00	71.94	92.86	78.97	81.79
TRCSP	94.44	80.00	96.67	90.37	65.18	98.21	72.45	89.73	78.57	80.83
DLCSP	98.89	70.00	98.33	89.07	71.43	91.07	65.31	92.86	80.16	80.16
CSP-L21	98.89	75.00	98.33	90.74	78.57	100.00	66.33	91.96	87.70	84.91
CCSP-L21	100.00	76.67	98.33	91.67	73.46	100.00	70.90	95.41	85.56	85.07
BCI winner	<u>0.82</u>	<u>0.76</u>	<u>0.80</u>	<u>0.79</u>	95.50	100.00	80.60	100.00	97.60	94.74

Table 4 Classification accuracies of CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21 on the subjects of data set IIa of BCI competition IV without outliers added. The BCI winner values with underline are the kappa scores. Values in bold indicate the best recognition rate for each subject

Data set	BCI competition IV									
	Data set IIa									
Subject	A01E	A02E	A03E	A04E	A05E	A06E	A07E	A08E	A09E	Mean
CSP	89.58	61.81	96.53	73.61	67.36	65.28	80.56	94.44	91.67	80.09
ACMCSP	89.58	59.03	96.53	72.92	72.92	68.06	78.47	94.44	91.67	80.40
TRCSP	90.97	58.33	96.53	64.58	70.14	68.06	84.72	94.44	90.97	79.86
DLCSP	89.58	61.11	96.53	73.61	68.75	68.06	82.64	95.14	91.67	80.79
CSP-L21	90.97	56.25	96.53	76.39	68.75	68.75	82.64	97.22	93.06	81.17
CCSP-L21	92.36	61.81	96.53	75.00	71.58	68.75	81.94	98.06	92.36	82.04
BCI winner	<u>0.68</u>	<u>0.42</u>	<u>0.75</u>	<u>0.48</u>	<u>0.40</u>	<u>0.27</u>	<u>0.77</u>	<u>0.75</u>	<u>0.61</u>	<u>0.57</u>

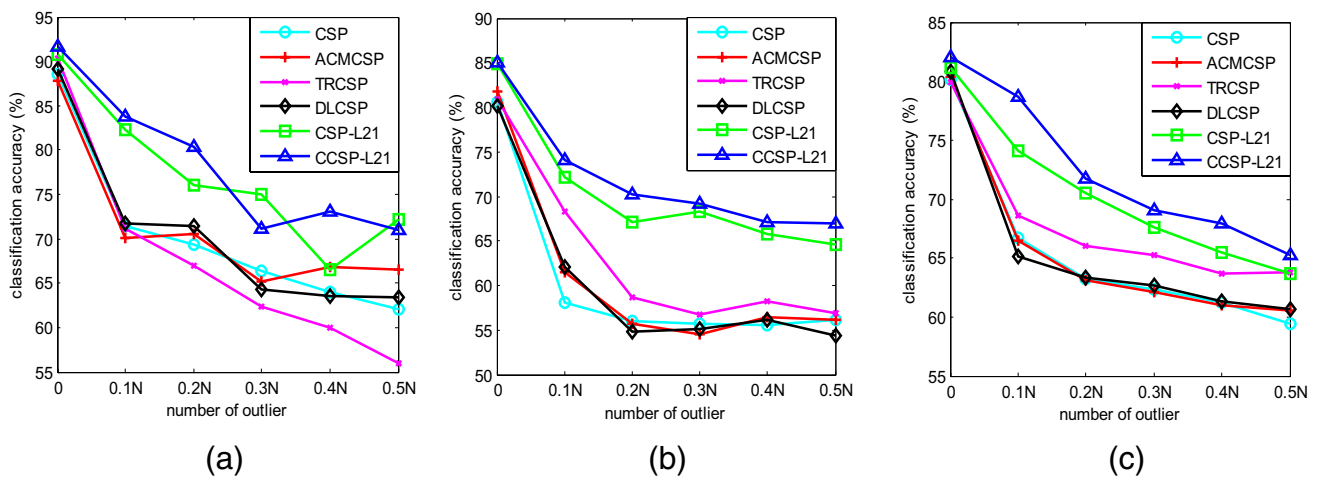


Fig. 2 Average classification accuracies of CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21 for the subjects of the three real EEG data sets with outliers added. The numbers of the outliers are

0.1 N, 0.2 N, 0.3 N, 0.4 N, and 0.5 N. **a** Data set IIIa of BCI competition III. **b** Data set IVa of BCI competition III. **c** Data set IIa of BCI competition IV

average recognition rates of the subjects varying with the outliers’ frequencies for each data set. As the frequency of the outliers increases, CCSP-L21 still has excellent discrimination accuracies, which always exceeds 65%, while the performance of the other methods deteriorates. Furthermore, taking the data set IVa of BCI competition III as an

example, the average classification accuracy of each subject is calculated by using the contaminated data set in Table 5. As you can see from the table, for the subject ay, the TRCSP algorithm performs best for its regularization term, which can effectively alleviate the overfitting problem in small data samples [38]. In addition, for several other subjects, there

Table 5 Average classification accuracies of the CSP, ACMCSP, TRCSP, DLCSP, CSP-L21, and CCSP-L21 methods for the five subjects with increasing outlier occurrence frequencies on data set IVa of

BCI competition III. Values in bold indicate the best average recognition rate for each subject. The numbers of the outliers vary from 0 to $0.5 N$ with step $0.1 N$

	CSP	ACMCSP	TRCSP	DLCSP	CSP-L21	CCSP-L21
aa	51.38	51.41	50.89	51.95	60.18	60.64
al	69.29	71.28	70.24	69.38	94.20	95.06
av	55.12	56.05	56.46	54.57	63.28	64.41
aw	58.12	57.91	58.77	57.99	60.64	64.98
ay	67.98	68.45	80.01	68.5	74.16	75.51
Mean	60.38	61.02	63.27	60.48	70.49	72.12

is no doubt that the CCSP-L21 algorithm obtains the best classification accuracy among all the methods. It can be seen that the performance of CCSP-L21 is superior to CSP-L21 in the case of outliers, which further proves the robustness of CCSP-L21. What is more, compared with the other extensions, the average recognition rates of CCSP-L21 are always improved by approximately 10%. According to the analysis above, we conclude that the CCSP-L21 approach is able to effectively reduce the impact of outliers.

Moreover, the determination of the line search parameter β and the thresholding parameter ϵ deserve to be discussed. We take the three subjects in data set IIIa of BCI competition III as research objects and draw the 3-D histogram of classification accuracies for each subject while varying the values of the parameters β and ϵ in Fig. 3. Empirically, the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ is designed for β , while the thresholding parameter ϵ is searched in set $\{1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2\}$ according to the relevant article [29]. It can be observed that the optimal value selection of β for each subject has a distinct difference because the state of the brain wave varies greatly between individuals. Moreover, the accuracy is generally able to reach a high level when ϵ takes the value from the list

$\{1e-2, 1e-1, 1\}$. If the thresholding parameter is set too large, the effect of filtering outliers will become worse. In contrast, we will lose much useful information by choosing a threshold value that is too small. This conclusively proves that the parameter plays an important role in the algorithm and needs to be adjusted carefully.

Last but not least, we analyze the computational complexity of the proposed method and other versions of the CSP-based method. For classical CSP, TRCSP, ACMCSP, and DLCSP, the main computational complexity comes from solving the eigen-equation, which needs computational complexity $O(C^3)$. For CSP-L21 and CCSP-L21, the non-greedy iterative procedure is derived to solve the proposed objective function. In reality, the dimensions of original EEG data are always larger than other constants, so we only consider the most important matrix calculation and the number of iterations. Therefore, if the iteration steps is T , the total computational complexity is $O((m+n)CT)$, where C denotes the number of electrodes (channels), and m and n represent the numbers of sampled points from the two brain states. It can be seen that the two algorithms based on the L21-norm are affected by the numbers of iteration, which are related to the setting of initial value and step size parameters.

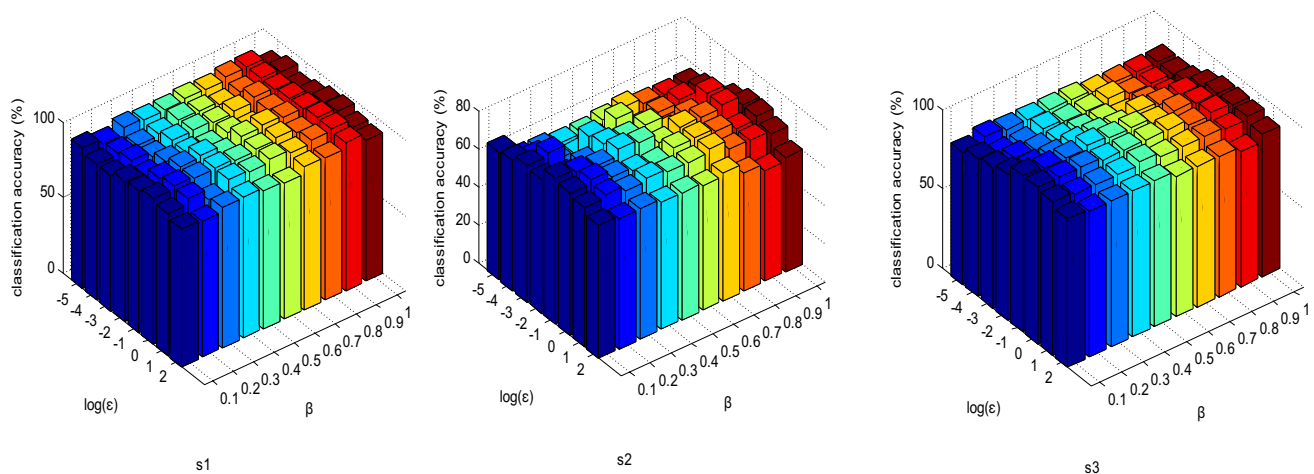


Fig. 3 Classification accuracies of CCSP-L21 for each subject from the data set IIIa of BCI competition III change with the value of the line search parameter β and the thresholding parameter ϵ

To sum up, the experiments on both noisy and unnoisy data sets demonstrate the robustness and superiority of the CCSP-L21 algorithm. However, the proposed method involves some parameters, which introduce uncertainty to the system. How to adjust the parameters optimally or design a more stable solution approach with fewer parameters will continue to be considered. In addition, improving processing speed and using more kind of noise can also be investigated in the future work.

5 Conclusion

In this paper, we propose the capped L21-norm-based common spatial patterns, named CCSP-L21. The algorithm aims to construct more robust models by introducing the capped L21-norm to redefine a covariance matrix of EEG data. Among them, the L21-norm removes the influence of the square operator, while the “capped” operation further achieves the goal of filtering extreme outliers. A non-greedy iterative algorithm is designed to compute the optimal solutions of the proposed CCSP-L21. Experimental results show that the CCSP-L21 method outperforms the classical CSP and other extensions. In future works, finding more appropriate parameters is a significant problem that deserves further consideration.

Acknowledgements The authors would like to thank the anonymous reviewers for their useful suggestions.

Funding This work was supported by the National Natural Science Foundation of China under Grant 62176054 and the Key Research and Development Plan (Industry Foresight and Common Key Technology) of Jiangsu Province, China, under Grant BE2022157.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Chaudhary U, Birbaumer N, Ramos-Murguialday A (2016) Brain-computer interfaces for communication and rehabilitation. *Nat Rev Neurol* 12(9):513–525
- Bashashati A, Fatourehchi M, Ward RK et al (2007) A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *J Neural Eng* 4(2):R32
- Blankertz B, Tomioka R, Lemm S et al (2008) Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag* 25(1):41–56
- Lotte F, Congedo M, Lécuyer A et al (2007) A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng* 4(2):R1–R13
- Afrakhteh S, Mosavi MR, Khishe M et al (2020) Accurate classification of EEG signals using neural networks trained by hybrid population-physic-based algorithm. *Int J Autom Comput* 17:108–122
- Mosavi MR, Ayatollahi A, Afrakhteh S (2021) An efficient method for classifying motor imagery using CPSO-trained ANFIS prediction. *Evol Syst* 12:319–336
- Afrakhteh S, Mosavi MR (2020) An efficient method for selecting the optimal features using evolutionary algorithms for epilepsy diagnosis. *J Circ Syst Comput* 29(12):2050195
- Afrakhteh S, Mosavi MR (2020) Applying an efficient evolutionary algorithm for EEG signal feature selection and classification in decision-based systems. In: *Energy efficiency of medical devices and healthcare applications*. Elsevier, pp 25–52. <https://doi.org/10.1016/B978-0-12-819045-6.00002-9>
- Hooda N, Kumar N (2019) Cognitive imagery classification of EEG signals using CSP-based feature selection method. *IETE Tech Rev* 37:315–326
- Dai Y, Zhang X, Chen Z et al (2018) Classification of electroencephalogram signals using wavelet-CSP and projection extreme learning machine. *Rev Sci Instrum* 89(7):074302
- Tang Z, Li C, Wu J et al (2019) Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI. *Front Inf Technol Electron Eng* 20(8):1087–1098
- Rahman AU, Tubaishat A, Al-Obeidat F et al (2022) Extended ICA and M-CSP with BiLSTM towards improved classification of EEG signals. *Soft Comput* 26:10687–10698
- Müller-Gerking J, Pfurtscheller G, Flyvbjerg H (1999) Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin Neurophysiol* 110(5):787–798
- Tang Q, Wang J, Wang H (2014) L1-norm based discriminative spatial pattern for single-trial EEG classification. *Biomed Signal Process Control* 10(3):313–321
- Wang H, Tang Q, Zheng W (2012) L1-norm-based common spatial patterns. *IEEE Trans Biomed Eng* 59(3):653–662
- Li X, Lu X, Wang H (2016) Robust common spatial patterns with sparsity. *Biomed Signal Process Control* 26:52–57
- Wang H, Li X (2016) Regularized filters for L1-norm-based common spatial patterns. *IEEE Trans Neural Syst Rehabil Eng* 24(2):201–211
- Wang H, Zheng W (2008) Local temporal common spatial patterns for robust single-trial EEG classification. *IEEE Trans Neural Syst Rehabil Eng* 16(2):131–139
- Zhang R, Xu P, Liu T et al (2013) Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery. *Comput Math Methods Med* 2013:591216
- Deng Y, Li Z, Wang H, Lu X, Fan H (2020) Local temporal joint recurrence common spatial patterns for MI-based BCI. In: *2020 IEEE 4th Information technology, networking, electronic and automation control conference*, vol 1. IEEE, Chongqing, China, pp 813–816. <https://doi.org/10.1109/ITNEC48623.2020.9084657>
- Fang N, Wang H (2017) Generalization of local temporal correlation common spatial patterns using Lp-norm ($0 < p < 2$). *Int Conf Neural Inf Process* 10637:769–777
- Gu J, Wei M, Guo Y, Wang H (2021) Common spatial pattern with L21-norm. *Neural Process Lett* 53:3619–3638
- Gu J, Cai Q, Gong W, Wang H (2021) L21-norm-based common spatial pattern with regularized filters. In: *2021 IEEE 4th Advanced information management, communicates, electronic and automation control conference*, vol 4. IEEE, Chongqing, China, pp 1746–1751. <https://doi.org/10.1109/IMCEC51613.2021.9482128>
- Ding C, Zhou D, He X, Zha H (2006) R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In: *Proceedings of the 23rd international conference on Machine learning (ICML '06)*. Association for Computing Machinery, pp 281–288. <https://doi.org/10.1145/1143844.1143880>
- Yang Y, Shen H, Ma Z et al (2011) L2,1-norm regularized discriminative feature selection for unsupervised learning. In: *Proceedings*

- of the Twenty-Second international joint conference on Artificial Intelligence (IJCAI'11). AAAI Press, pp 1589–1594. <https://dl.acm.org/doi/10.5555/2283516.2283660>
26. Liao S, Gao Q, Yang Z, Chen F (2018) Discriminant analysis via joint Euler transform and L2,1-norm. *IEEE Trans Image Process* 27(11):5668–5682
 27. Liu Y, Gao Q, Gao X, Shao L (2018) L2,1-norm discriminant manifold learning. *IEEE Access* 6:40723–40734
 28. Lai Z, Liu N, Shen L et al (2019) Robust locally discriminant analysis via capped norm. *IEEE Access* 7:4641–4652
 29. Wang Z, Nie F, Zhang C et al (2020) Capped Lp-norm LDA for outliers robust dimension reduction. *IEEE Signal Process Lett* 27:1315–1319
 30. Wolpaw J, Birbaumer N, McFarland D, Pfurtscheller G, Vaughan T (2002) Brain-computer interfaces for communication and control. *Clin Neurophysiol* 113(6):767–791
 31. Parra L, Spence C, Gerson A, Sajda P (2005) Recipes for linear analysis of EEG. *Neuroimage* 28(2):326–341
 32. Luo M, Nie F, Chang X et al (2016) Avoiding optimal mean robust PCA/2DPCA with non-greedy L1-norm maximization. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, pp 1802–1808. <https://dl.acm.org/doi/abs/10.5555/3060832.3060873>
 33. Tangermann M, Müller KR, Aertsen A et al (2012) Review of the BCI competition IV. *Front Neurosci* 6(55):2
 34. Pfurtscheller G, Neuper C, Flotzinger D et al (1997) EEG-based discrimination between imagination of right and left hand movement. *Electroencephalogr Clin Neurophysiol* 103(6):642
 35. Lotte F, Guan C (2011) Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans Biomed Eng* 58(2):355–362
 36. Kawanabe M, Vidaurre C (2009) Improving BCI performance by modified common spatial patterns with robustly averaged covariance matrices. *World Cong Med Phys Biomed Eng* 25:279–282. https://doi.org/10.1007/978-3-642-03889-1_75
 37. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88(2):365–411
 38. Lu H, Eng HL, Guan C et al (2010) Regularized common spatial pattern with aggregation for EEG classification in small-sample setting. *IEEE Trans Biomed Eng* 57(12):2936–2946

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Jingyu Gu was born in Jingjiang, China, in 1996. She received the B.E. degree in computer science and technology from the Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China, in 2018, and M.S. degree in biological science and medical engineering from Southeast University, Nanjing, Jiangsu, China, in 2021. She is currently working toward her Ph.D. degree at the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological

Science & Medical Engineering, Southeast University, Nanjing, Jiangsu, China. Her research interest focuses on EEG signal processing.



Jiuchuan Jiang received the Ph.D. degree in Computer Science from Nanyang Technological University, Singapore. He is currently an associated professor with the School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China. His main research interests include crowdsourcing, multiagent systems, and social networks. He has published several scientific articles in refereed journals and conference proceedings, such as the *IEEE Transactions on Parallel and Distributed Systems*, the *IEEE Transactions on*

Systems, Man, and Cybernetics: Systems, the *ACM Transactions on Autonomous and Adaptive Systems*, and the *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.



Sheng Ge received the B.E. degree from the Nanjing University of Science and Technology, China, in 1997, and the M.E. and Ph.D. degrees from Yamaguchi University, Japan, in 2001 and 2004, respectively. From 2004 to 2005, he was a Research Assistant with the Faculty of Engineering, Yamaguchi University. From 2005 to 2007, he was a COE Research Fellow with the Graduate School of Information Science and Electrical Engineering, Kyushu University. In 2007, he was an Assistant Professor with Kyushu University.

From 2007 to 2009, he was a Post-Doctoral Fellow with the Department of Radiology, University of California at San Diego, San Diego, USA. From 2010 to 2012, he was an Associate Professor with the Nanjing University of Science and Technology, China. Since 2010, he has been an Associate Professor with the Key Laboratory of Child Development and Learning Science, Ministry of Education, Research Center for Learning Science, School of Biological Science and Medical Engineering, Southeast University, China. His current research interests include brain science, brain-computer interface, and biomedical signal processing.



Haixian Wang received the B.S. and M.S. degrees in statistics and the Ph.D. degree in computer science from Anhui University, Hefei, Anhui, China, in 1999, 2002, and 2005, respectively. Currently, he is a full professor with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing, Jiangsu, China. His research interests focus on biomedical signal processing, brain-computer interfaces, and machine learning. He

has published over one hundred papers in these fields.