

Received July 31, 2020, accepted August 6, 2020, date of publication August 14, 2020, date of current version September 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016700

An Adaptive CSP and Clustering Classification for Online Motor Imagery EEG

QIN JIANG¹, YI ZHANG², GENGYU GE¹, AND ZHIRONG XIE³

¹School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

²Advanced Manufacturing and Automation Engineering Laboratory, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

³Key Laboratory of Optoelectronic Information Sensing and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding authors: Qin Jiang (namy_jiang@hotmail.com) and Yi Zhang (zhangyi@cqupt.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant 51775076, and in part by the Ph.D. Program of Chongqing University of Posts and Telecommunications under Grant BYJS201910.

ABSTRACT A potential limitation of motor imagery (MI) based brain-computer interface (BCI) (MI-BCI) is that it usually requires a relatively long time to record sufficient electroencephalogram (EEG) data for robust feature extraction and classification. Moreover, due to the non-stationarities in EEG signals, the offline training model has poor adaptability and classification ability in cross-session or sample-wise online testing. **Methods:** To address the problems, we propose a model updating scheme with adaptive and fast operation. Based on the Common Spatial Pattern (CSP), we propose an online and fast generalized eigendecomposition method by Recursive Least Squares updates of the CSP filter coefficients (RLS-CSP), which allows incremental training for CSP spatial filters. Additionally, we present an Incremental Self-training Classification algorithm based on Density Clustering (ISCDC) to select high-confidence samples to update spatial filters and classifier, and classify at the same time. **Results:** We conducted extensive experiments to validate the efficiency of the proposed adaptive CSP and classifier on the BCI III_IVa and BCI III_V data sets. Experimental results demonstrate that RLS-CSP outperforms significantly in a small sample setting (SSS), and ISCDC has great adaptability in cross-session and non-stationary EEG signals. The results indicate that our proposed methods are feasible to improve the real-time performance of online BCI system.

INDEX TERMS Brain-computer interface, common spatial pattern, recursive least squares, self-training classification, density clustering.

I. INTRODUCTION

Brain-Computer Interface (BCI) is a human-computer interaction technology. It does not rely on the peripheral nerve and muscle system and aims to provide a bridge between human brain and external devices [1], [2]. It has demonstrated broad application prospects in the rehabilitation of disabled people and auxiliary control of healthy people [3].

MI-BCI generates correlation signals by thinking, which is accompanied by event-related desynchronization and event-related synchronization (ERD/ERS) in functional motor areas [4], [5]. Effective characterization of ERD/ERS phenomenon is of vital importance to a MI-BCI system. The Common Spatial Pattern (CSP), is a widely used a time-spatial feature extraction method in MI-BCI system, which is effective to extract the frequency band variances as

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang¹.

features [6], [7] by conducting spatial filtering on multichannel simultaneously [8].

A. SMALL SAMPLE SETTING

CSP is highly dependent on sample-based covariance [2], and is very sensitive to noise and prone to overfitting in a small sample setting (SSS) [9]. The regularization CSP is developed to estimate a covariance matrix by adding a-priori information into the CSP learning process, under the form of regularization terms [10], [11]. The regularization CSP can be implemented in two levels. One method is performed at the covariance matrix estimation level, since spatial covariance matrix estimates can suffer from noise or small training sets. The other approach is regularizing CSP at the level of the objective function itself. Tikhonov Regularization CSP (TRCSP) reduces deviation by adding identity matrix to the denominator of the objective function to constrain the norm of spatial filters [12], [13]. TRCSP performs equal

weighted penalty on all signal channels, whereas it ignores the particularity of the channels in reflecting region. Therefore, the weighted Tikhonov Regularization CSP (wTRCSP) is improved based on TRCSP by introducing priori information of channels weights, which can be obtained from training samples of auxiliary objects. However, wTRCSP ignored the personality and individual differences between auxiliary subjects and the target subject. Invariant CSP (iCSP) maintains the robust of spatial filters under given noise by adding general invariances to the objective function [15]. However, in order to compute the penalty covariance matrix, an extra artifact recording and additional preprocessing methods are needed. Complex CSP (CCSP) uses the samples of other subjects to compensate for the estimation bias of covariance matrix [16], which may aggravate the estimation bias due to the differences between subjects. In Regularized CSP with Selected Subjects (SSRCSP), samples of other subjects are selectively added into covariance matrix, however, when the data set is large, the selection process takes a long time [9]. Besides, a filter band regularization with CSP (FBRCSP) [2] has been proposed to overcome dependence on the frequency and covariance matrix estimation, which a regularization CSP [9] is worked as spatial filters on each frequency filter.

Moreover, subject-to-subject feature transfer provides a promising approach to learn reliable features from the limited data of the target subject with help of sufficient data from other subjects [17]–[19]. A sparse representation-based classification (SRC) scheme has demonstrates its advantage in exploring potential relationship of CSP features among subjects [20]. Based on SRC, the Sparse Group Representation Model (SGRM) aims at finding out the most significant training feature from both the target and other subjects by exploiting two norm regularizations [8]. A sparse filter band CSP (SFBCSP) [21] is proposed to improve the selection of filter band in a supervised way by exploiting sparse representation learning [20], [22], and a temporally constrained sparse group spatial pattern (TSGSP) [22], for the simultaneous optimization of filter bands and time window within CSP to further improve accuracy of MI-related EEG. However, the sparse-based method represents a linear relationship of all the training samples, which puts forward higher requirements on the quality of the training samples [20].

It should be noted that all the regularized-based or sparse-based CSP algorithms are affected by their hyper-parameters. However, the parameter optimization based on cross-validation is relatively time-consuming and requires additional dataset for validation, which inevitably produces parameter deviations in SSS [9], [20]. Besides, the parameters need to be re-determined when adding new samples, which limits the practicality of BCI system to some extent.

Aiming at time-consuming and poor adaptability of the offline regularization CSP in MI-BCI system, this paper proposes RLS-CSP, an online and fast generalized eigen-decomposition method, which updates the filter coefficients by Recursive Least Squares (RLS). Compared with

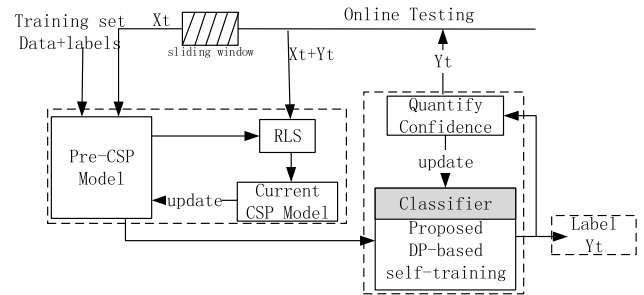


FIGURE 1. The proposed online training and adaptation scheme.

regularization CSP that relies on batch computing, RLS-CSP allows adding testing samples to alleviate the overfitting and improve the adaptability of the RLS-CSP filters. How to select high-confidence samples is very important to the performance of RLS-CSP.

B. NON-STATIONARITY IN THE TESTING PHASE

Because of the non-stationarities in EEG, the feature distribution change obviously across sessions. Therefore, spatial filters adaptive improvements are necessary to maintain their high performance in a long duration. One of the ways to keep CSP adaptive is to update the initial spatial filters with high-confidence testing samples [23]–[25]. To evaluate the confidence, this paper proposes an Incremental Self-training Classification method based on Density Clustering (ISDC) by combining Density Peaks Clustering (DPC) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). In ISDC, the self-training method based on density peaks is redesigned to classify in SSS, and quantify the reliability of classification. Moreover, a parameter-free local noise filter based on DBSCAN is proposed to filter out mislabeled instances. Compared with existing classifications used in MI-BCI, the ISDC can remove mislabeled samples by exploiting the information of both labeled data and unlabeled data, and assess the reliability of classification.

Fig.1 demonstrates the framework of our proposed approach for a simple-wise online learning. The main contributions of this work are as follows.

1) *Aiming at feature extraction in SSS*, this paper proposes an adaptive CSP method (RLS-CSP), which allows incremental updating covariance matrix and fast decomposing eigenvectors by RLS. The RLS-CSP has two advantages: a) allows incremental computation of the filter coefficients which is efficient regarding the required memory and computational effort, and b) allows the incorporation of new samples to adapt the current filter, improving overfitting gradually.

2) *Adaptation classification of time-varying EEG*, a cluster classification method based on density clustering is proposed to execute reliability evaluation on testing samples, remove outliers, and adaptively update cluster centers and spatial filters. Moreover, that distinguishes from recent investigated batch-base clustering [26–29], our algorithm performs efficient classification in SSS, and is applicable in sample-wise online BCI.

3) **In the case of SSS and cross-session**, A series of detailed experiments are designed, with RLS-CSP for feature extraction and ISDC for classification.

The rest of this paper is organized as follows. Section II details the basic principle of proposed RLS-CSP algorithm. Section III describes the proposed adaptation classification and the implementation framework of our algorithms. Section IV provides a detailed description of the experiment design. Section V evaluates the performance of our method on two datasets. Finally, the conclusions and future work are presented in Section VI.

II. ADAPTIVE CSP SPATIAL FILTERS IN SSS

Since the CSP algorithm is the basis of our method, we shortly describe it in the beginning. Afterwards, we introduce the RLS feature decomposition method [30] to incorporate new arrival samples to fit the current spatial filters in an application.

A. BASIC CSP

Suppose $X_c \in \mathbb{R}^{M \times T}$ represent a single-trial time-space matrix, with M channels and T samples in each channel. We suppose except for different cognitive tasks, the matrix and recording conditions are the same in two classes indexed by $c \in \{1, 2\}$. CSP aims at learning spatial filters which maximize the variance of band-pass filtered EEG signals from one class while minimizing their variance from the other class, which its objective function can be defined as:

$$\begin{aligned} g(\omega) = \arg \max_{\omega} \frac{\omega \sum_1 \omega^T}{\omega (\sum_2) \omega^T} \\ \text{subject to } \omega (\sum_2) \omega^T = K \end{aligned} \quad (1)$$

where \sum_c is mean covariance matrix for class c , ω denotes spatial filters with K a real constant. Using the Lagrange multiplier method, this constrained optimization problem amounts to extremizing the following function:

$$L(\omega, \lambda) = \omega \sum_1 \omega^T - \lambda (\omega (\sum_2) \omega^T - K) \quad (2)$$

The spatial filters ω extremizing L are such that the derivative of L with respect to ω equals zero:

$$\sum_1 \omega = \lambda (\sum_2) \omega \quad (3)$$

Equation (3) is a standard eigenvalue decomposition problem. The spatial filters ω are the eigenvectors of $V = (\sum_2)^{-1} \sum_1$, which correspond to its largest and lowest eigenvalues. The extracted features are the logarithm of the EEG signal variance after projection onto ω .

Obviously, the optimization of the spatial filter is based on the variance estimation. However, in SSS, the covariance matrix estimation is overfitted [9]–[12]. Also, non-stationary factors such as noise, fatigue and emotion lead to the deviation in spatial filter [14], [15]. One way to solve this problem is to adjust spatial filters adaptively through test samples.

B. RLS BASED CSP

The optimal spatial filter of (3) is a standard generalized eigendecomposition. Instead of using the QRD and SVD, we introduce an incremental approach based on the recursive least squares (RLS) method [31].

All the generalized eigenvalues are stationary points of (3). Hence the formula is transformed as:

$$\sum_1 \omega = \frac{\omega \sum_1 \omega^T}{\omega (\sum_2) \omega^T} \sum_2 \omega \quad (4)$$

Left multiplying (4) by \sum_2^{-1} becomes:

$$\omega = \frac{\omega \sum_2 \omega^T}{\omega (\sum_1) \omega^T} \sum_2^{-1} \sum_1 \omega \quad (5)$$

Here Recursive Stochastic algorithm [27] is used to approximate the maximum eigenvector of the n th iteration as:

$$\begin{aligned} \omega_1(n) \frac{\omega_1(n-1) \sum_2^1(n) \omega_1(n-1)^T}{\omega_1(n-1) (\sum_1^1(n) \omega_1(n-1)^T} \\ * [\sum_2^1(n)^{-1} \sum_1^1(n) \omega_1(n-1)] \end{aligned} \quad (6)$$

where $\omega_i(n)$ represents the i th eigenvector of n th iteration, $\sum_c^i(n)$ denotes the c -class covariance matrix corresponding to the i th spatial filter.

Equation (6) explains how to approximate the largest eigenvector in the n th iteration with the largest generalized eigenvector in the $n-1$ th iteration. Especially, the eigenvectors of ω are arranged in descending order.

Computation $\sum_2^1(n)^{-1}$: According to Sherman-Morrison-Woodbury theorem, the inverse of $\sum_2^1(n)$ can be implemented using recursive estimators, then,

$$\begin{cases} \sum_2^1(n)^{-1} = \sum_2^1(n-1)^{-1} - \sum_2^1(n-1)^{-1} \Theta \sum_2^1(n-1)^{-1} \\ \Theta = X_2(n)(I + X_2^T(n) \sum_2^1(n-1)^{-1} X_2(n)) X_2^T(n) \end{cases} \quad (7)$$

where $X_c(n)$ denotes new sample of c -class in n th iteration, the superscript T denotes the transpose of a matrix.

Covariance matrix $\sum_c^1(n)$: The covariance matrix can be updated by:

$$\sum_c^1(n) = \frac{\sum_c^1(n-1) + X_c(n) X_c^T(n)}{2} \quad (8)$$

Lower-Order Filters: By using (6) the first spatial filter is obtained. For the minor components, a standard deflation procedure [26] is considered:

$$\begin{cases} \sum_1^i(n) = [I + \frac{\sum_1^{i-1}(n) \omega_{i-1}(n) \omega_{i-1}^T(n)}{\omega_{i-1}^T(n) \sum_1^{i-1}(n) \omega_{i-1}(n)}] \sum_1^{i-1}(n) \\ \sum_2^i(n) = \sum_2^{i-1}(n) \end{cases} \quad (9)$$

Hence the low-order filter $\omega_i(n)$ is given by (6) with $\sum_1^1(n)$ and $\sum_2^1(n)$ replaced by $\sum_1^i(n)$ and $\sum_2^i(n)$.

Note that the deflation given by (9) does not increase the complexity of the algorithm because all the terms in (9) are pre-computed in pre-iteration, and each filters $\omega_i(n)$ in n th iteration depends on the preceding iteration filter $\omega_i(n - 1)$.

The complete algorithm is detailed in Alg.1.

Algorithm 1 RLS Based CSP(RLS_CSP)

Input: Training data $X_i \in R^{M \times T}$, ($i = 1, 2, \dots, N$)
with labels $Y_i \in R^{N \times T}$, ($i = 1, 2, \dots, N$)
Testing data $X_t \in R^{M \times T}$, ($t = 1, 2, \dots, N$)
Output: The new spatial filter $\hat{\omega}$

1. Initialization:
 $\Sigma_2^1(0) \setminus \Sigma_2^1(0) \setminus \Sigma_2^1(0)^{-1}$; $\omega(0)$
 $n = 0$ and $\omega(0) = \omega(0)$

2. Begin:
While the new sample available **do**
counter $n + +$
Update $\Sigma_1^1(n)$ or $\Sigma_2^1(n)$ by (8), and $\Sigma_2^1(n)^{-1}$ by (7)
for $i \in [1, 2, \dots, m]$ **do**
if $i > 1$ **then**
using (9) to get $\Sigma_1^1(n)$
 $\omega_i(n - 1) = \hat{\omega}(n - 1)$
Calculate $\omega_i(n)$ using (6)
Normalize $\omega_i(n)$
Save $\omega_i(n)$ in $\hat{\omega}(n)$
Get $\hat{\omega}(n)$

III. ADAPTIVE CLASSIFICATION OF TIME-VARYING EEG

Notations: $CoNb(x)$ is the set which lies in the region with x as the center and cutoff distance $dist_{cutoff}$ as the radius, defining as $CoNb(x) = \{x_i \in D | dist(x_i, x) \leq dist_{cutoff}\}$.

$ECoNb(x)$ is defined as $ECoNb(x) = \{x_i \in D | dist(x_i, x) \leq \lambda * dist_{cutoff}\}$, where $\lambda = 1, \dots, N$, N is a non-zero integer. $ECoNb(x)$ is the set which lies in the field of extended $CoNb(x)$.

$X_i \in R^{M \times T}$ is a single-trial EEG signal, with M channels and T sample points.

$x_i \in R^m$ is a m -dimension feature vector.

$L = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ is the set of all labeled samples.

$U = \{x_1^u, x_2^u, \dots, x_t^u\}$ is a training set without labels.

A. DENSITY CLUSTERING-BASED CLASSIFICATION

Peak density clustering is a widely used unsupervised clustering method, which is based on the idea that clusters centers have a higher density than their neighbors [26]. For each sample x_i two quantities have to be present: its local density ρ_i and its distance δ_i from points of higher density.

The local density ρ_i of data point x_i is defined as:

$$\rho_i = \sum_{x_j} \phi(dist(x_i, x_j), dist_{cutoff}) \tag{10}$$

where $dist_{cutoff}$ is a cutoff distance, $\phi(x, y)$ is usually defined as a sign function, i.e. $\phi(x, y) = 1$ if $x - y \leq 0$ and $\phi(x, y) = 0$

otherwise. However, for computing the density for cases in SSS, the exponential kernel is suggested $\phi(x, y) = \exp^{-x^2/y^2}$.

δ_i measures minimum distance between the sample x_i and any other point with higher density:

$$\delta_i = \begin{cases} \min_{j, \rho_j > \rho_i} dist(x_i, x_j) \\ \max_{j, \rho_i \geq \rho_j} dist(x_i, x_j) \end{cases} \tag{11}$$

The cluster center is quantified using relatively large δ_i and ρ_i . Then, the samples are clustered in the direction of decreasing density with the cluster center as the starting point. Similar to the DPC principle, DBSCAN calculates the local density by ϵ_i - neighborhood, and the number of samples in neighborhood is the wanted density. Points with a density higher than $MinPts$ are considered cluster centers. Eventually the samples are classified according to the neighborhood accessibility starting from the central clustering. Readers can refer to [27] for more details about DBSCAN.

Once the cluster centers are selected, the testing samples can be classified by calculating the distance to the cluster centers. However, for time-varying samples, especially in online BCI, the clustering-based classification has the following disadvantages:

1) Because of the limitation in SSS, the obtained cluster center is a local optimum under current samples, not a global one for the whole test set.

2) Batch-based updating method for cluster center burdens time-consuming, which makes it unsuitable for sample-wise online BCI system.

3) Due to the non-stationarity, the influence of test samples on model is different, some are conducive to clustering, and others are not, so filtering outliers is of particular importance for a robust classifier.

B. ADAPTIVE CLASSIFICATION FOR A TIME-VARYING SITUATION: ISCDC

Aimed at above problems, an adaptive density classification method (ISCDC) is proposed. It uses incremental computation to reduce time consumption, automatically updates cluster centers through sample distribution, and performs confidence evaluation and parameter-free noise filtering.

Fig.2 shows a general framework of ISCDC, which can be divided into three steps. At the first step, we use DPC to conduct a supervised classifying on a small amount of labeled data, and construct the neighborhood of each labeled point $CoNb(x)$. Step two is a standard self-training process, where unlabeled samples are initially predicted according to the density with c -class points. Besides, we introduce a concept of $ECoNb(x)$ to assess the reliability of pre-predicted labels and remove noisy instances. Then, continuously add high-confidence samples to the labeled data to expand the neighborhood of the labeled samples and update the spatial filter.

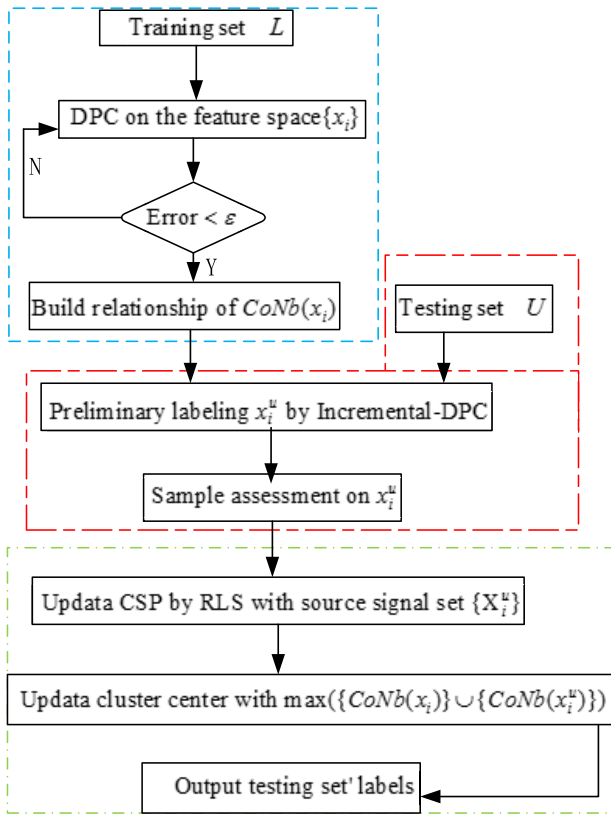


FIGURE 2. The framework of ISCDC.

1) DISCOVER THE UNDERLYING STRUCTURE OF TRAINING DATA

Unsupervised DPC clustering may cause mislabeling [25], [28]. Therefore, we utilize labeled instances to obtain the best $dist_{cutoff}$ by adjusting $dist_{cutoff}$ and repeatedly clustering. The expected result is that the clustering results of the training samples are as consistent as possible with the true labels, and $dist_{cutoff}$ is relatively small. After that, we construct the neighborhood set ($\{CoNb(x_i)\}$) of training samples with $dist_{cutoff}$ being radius in density decreasing order. Algorithm 2 illustrates the procedure of mining the structure of the training samples by our algorithm, and the output is the set $\{CoNb(x_i)\}$.

Algorithm 2 Discover Structure (DS)

Input: $L = \{(x_i, y_i)\}$, $dist_{cutoff}$, ϵ
Output: $CoNb(x_i)$, $i = (1, 2, \dots, l)$, $l < n$
1: Cluster using DPC on $\{x_i\}$ with cutoff $dist_{cutoff}$
2: Validation with the labels $\{y_i\}$, and adjust $dist_{cutoff}$
3: repeat step1&2 until recognition error $< \epsilon$
4: for each x_i in L
 $CoNb(x_i) = \text{find} (dist(x_i, x_j) < dist'_{cutoff})$
 If $CoNb(x_i) == CoNb(x_j)$
 delete $CoNb(x_j)$
end for

2) CLASSIFICATION AND CONFIDENCE EVALUATION OF TESTING SAMPLES

Equation (12) indicates that in Euclidean space, the closer the point is to a group of samples, the higher the local density. Therefore, we get an initial prediction on the tested sample according to the density between the point and different types of labeled ones. After that, it is crucial to assess the reliability of the classification.

Inspired by neighborhood-based clustering (i.e. DBSACN), we introduce an extended neighborhood method to quantify the confidence and remove outliers. We suppose that there is no mutation between adjoining samples of the same type, and the mislabeled sample has a different label from its adjacent neighborhood. We use $CoNb(x_i)$ to represent the neighborhood of a tested instance. The distribution of the testing sample and training samples generally consists of three situations, i.e. the neighborhood of the tested point ($CoNb(x_i)$) (a) contains one category points, (b) includes two categories points, (c) contains no points. For the third case, an extended- $CoNb(x_i)$ (i.e. $ECoNb(x_i)$) is defined.

To quantify the confidence, we modify the definition of harmfulness and usefulness of a sample x_i^u in reference [31], as follows:

$$\begin{cases} Harm(x) = N(\{z | z \in ECoNb(x), l(z) \neq l(x)\}) \\ Help(x) = N(\{z | z \in ECoNb(x), l(z) = l(x)\}) \end{cases} \quad (12)$$

$Harm(x)$ indicates the number of instances z , where $ECoNb(x)$ contains z and z' 's label differs from that of x . By contrast, $Help(x)$ represent the number of a set that $z \in ECoNb(x)$ and $l(z) = l(x)$, where N indicates the number of a set, l denotes the class label of an instance. According to the (14), we can quantify the classification and filter out a noisy point without any additional parameters.

Situation 1: The set $CoNb(x_i^u)$ of point x_i^u contains only one type of points. If $Help(x_i^u) = 0$, x_i^u is regarded as an outlier and the initial classification result is revoked. if $Harm(x_i^u) = 0$, x_i^u is considered reliable, which is beneficial to supply the clustering distribution. So x_i^u is incorporated into training set and its neighborhood relationship $CoNb(x_i^u)$ is preserved, and then its corresponding EEG signal X_i^u is saved for RLS-CSP.

Situation 2: The set $CoNb(x_i^u)$ contains two types of samples. If $Help(x_i^u) > Harm(x_i^u)$, it is considered that the initial classification of x_i^u is correct with a moderate confidence, and then merge x_i^u into training set, without its $CoNb(x_i^u)$. If $Harm(x_i^u) \geq Help(x_i^u)$, x_i^u is treated as a noise point.

Situation 3: The set $CoNb(x_i^u)$ contains no instances. Under the circumstances, we construct a neighborhood $ECoNb(x_i^u)$ of x_i^u with the minimum $\lambda (\lambda \neq 1)$, and then refer to situation 1 and 2 to evaluate the instance x_i^u . Significantly, according to the constrain of "no mutation", if $Harm(x_i^u) = 0$, the neighborhood $ECoNb(x_i^u)$ of x_i^u and its single-trial X_i^u are deleted.

At last, we can update clustering centers with the maximum set of $CoNb(x_i)$, when the training scale accumulates to an

appropriate size. Algorithm 3 describes the pseudo-code of the proposed ISCDC.

Algorithm 3 The ISCDC

Input: testing feature set $U = \{x_1^u, x_2^u, \dots, x_t^u\}$
Output: $U' = \{(x_1^u, y_1), (x_2^u, y_2), \dots, (x_r^u, y_r)\}, (r < t)$
while $x_i^u \in U$ available **do**
 1: Calculate local density ρ_c of x_i^u by (10)
 2: Label x_i^u according to ρ_c , and get $CoNb(x_i^u)$
 3: **if** $CoNb(x_i^u) \neq \emptyset$
 if $Harm(x) == 0$
 save $CoNb(x_i^u)$
 save testing source signal $\{X_t^u\} \cup X_i^u$
 update $CoNb(x_i) \cup x_i^u$
 else if $Harm(x) \geq Help(x)$
 delete noisy sample x_i^u and $CoNb(x_i^u)$
 else if $Harm(x) < Help(x)$
 delete $CoNb(x_i^u)$
 update $CoNb(x_i) \cup x_i^u$
 4: **else**
 Search the minimum of $ECONb(x)$
 if $Harm(x) \geq Help(x)$
 delete noisy sample x_i^u and $CoNb(x_i^u)$
 else
 update $CoNb(x_i) \cup x_i^u$
 5: update spatial filter $\hat{\omega}$ by $\{X_i^u\}$ using Algorithm 1

IV. EXPERIMENT

This section presents a large number of experiments to support the following objectives:

- 1) Investigate how the performance of RLS-CSP is affected by the size of training set and testing duration.
- 2) Investigate how the performance of ISCDC is affected by the size of training set and testing duration.
- 3) Evaluate the efficacy of RLS-CSP combined with ISCDC in continuous BCI system.

A. DATASET DESCRIPTION

1) BCI COMPETITION III DATASET Iva TERMED AS DATASET 1

The dataset contains EEG signals recorded at 118 channels with 1000 Hz sampling rate (downsamples to 100Hz in this paper) from five subjects named ‘‘AA’’, ‘‘AL’’, ‘‘AV’’, ‘‘AW’’ and ‘‘AY’’. For each subject, a total of 280 cue-based trials are available (half for each class of MI). In each trial, a cue was indicated for 3.5 s during which two MI tasks were performed: (R) right hand, (F) right foot. Then the cue was intermitted by periods of random length, 1.75 to 2.25 s, in which the subject could relax. See website http://www.bbc.de/competition/iii/desc_IVa.html for more details about the dataset.

2) BCI COMPETITION III DATASET V TERMED AS DATASET2

In this dataset, EEG was recorded from three subjects at 32-channel with a sampling rate of 512Hz for three tasks

left-hand movement, right-hand movement and word generation (in our experiment word-imagery was eliminated). Each subject performs 4 sessions, each lasting 4 minutes with 5-10 minutes breaks in between them. The subject performed a given task for about 15 seconds and then switched randomly to another task.

B. EXPERIMENT DESIGN

Dataset1 and Dataset2 were filtered by a 5-order Butterworth band-pass filter of 8-30Hz. All trials were normalized before the feature-extraction process, and two pairs of spatial filters were selected for feature extraction.

For Experiment I, II and III were carried out on Dataset 1. In each trial 3-second data was captured after 0.5s of a cue. In Experiment IV, Dataset2, a continuous data set, is used to simulate the an online BCI system. The trial is captured by a 2-second-length sliding window with 1-second overlap.

The specific settings of each experiment are as follows:

1) EXPERIMENT I: RLS-CSP IN SSS

To investigate the performance of RLS-CSP in SSS, a set of experiments were carried out with different size of training set on each subject.

We randomly selected 30trials from each auxiliary subject (15 trials per class) and 0 to 140 trials from the target subject for classifier training, and the rest 140 trials from the target subject for testing. The hyperparameters β and γ (from the set of 10 candidates $\{0.001, 0.01, 0.1, 0.2, \dots, 0.8\}$) and hyperparameter C in SVM were determined by five-fold cross-validation on the calibration data.

RLS-CSP was compared against the conventional CSP as well as other competing CSP:

CSP: the conventional CSP.

CCSP1: reduce the bias of covariance matrix by using other subjects' data with one parameter β [16].

TRCSP: a quadratic penalty β for the spatial filter [9].

SFBCSP: a regularization parameter β in SFBCSP, and 10subbands with 2Hz overlap were used, referred [21] for details.

FBRCSP: 10subbands with 2Hz overlap for EEG filter, and R-CSP [10] with two parameters β and γ for the covariance matrix estimation to extract features, referred [2] for details.

SGRM: two hyperparameters β and γ from 0.001 to 0.01 with an interval of 0.001 were suggested in [8].

2) EXPERIMENT II: RLS-CSP IN NON-STATIONARY EEG

To verify the adaptability of RLS-CSP to the time-varying EEG, we designed the experiments in two scenarios: subject-specific and subject-independent.

Subject-specific means that both training and test data are from the same subject. we used 40trilas to initialize the spatial filters and then analysed the feature distributions of sets of the first-120trials (120st for short) and the second-120trials (120nd for short), which two testing sets were organized in chronological order. Furthermore, we defined a measurement

TABLE 1. The average accuracy (%) for dataset1 in different training size by the RLS-CSP, CSP, SFBCSP, CCSP, TRCSP, FBRCSP and SGRM. the the highest accuracy is marked in boldface.

	10	20	30	40	80	100	120	140	mean±std
RLS-CSP	78.87	80.27	82.05	83.38	87.26	88.13	89.138	90.46	84.94±3.16
CSP	70.3	72.02	74.17	77.18	79.56	82.28	81.71	80.74	76.9±4.52
SFBCSP	73.14	77.68	81.25	82.1	85.42	87.053	88.67	88.53	83.09±5.93
CCSP	70.02	72.92	75.82	76.49	82.855	79.5	81.68	80.01	77.19±4.02
TRCSP	72.16	73.87	77.42	77.183	80.14	79.57	79.715	80.7136	77.71±3.12
FBRCSP	72.53	76.92	77.45	78.3	82.95	83.58	84.653	85.73	81.04±5.19
SGRM	73.48	75.74	76.85	78.68	82.43	81.04	81.97	82.68	79.92±3.27

to quantify the divergence of distributions.

$$DR = \frac{Tr(Sb)}{Tr(Sw)} \quad (13)$$

where $Tr(\cdot)$ denotes the trace of a matrix, Sw is the scatter within classes, Sb is the scatter between classes, and the result DR represents the divergence ratio of feature distribution. The larger the DR , the more separable the two sets of features are.

Subject-independent means that training data comes from auxiliary subjects, and test data from the target subject. The first 40 trials of each subject served as the training set and rest 240 trials for testing. The propose RLS-CSP and ISCDC were used for feature extraction and classification.

3) EXPERIMENT III: ISCDC IN SSS

The training set was randomly selected from 0 to 140 trials for each target subject, and the rest 140 tails for testing. We evaluated the efficacy of ISCDC, by compared with DPC and DBSCAN.

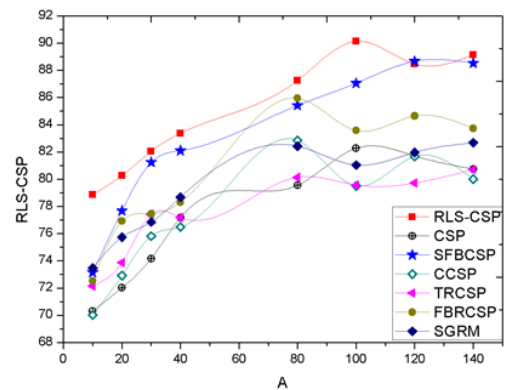
4) EXPERIMENT IV: THE EFFICACY OF RLS-CSP COMBINED WITH ISCDC IN CONTINUOUS BCI SYSTEM

Dataset2 was used to simulate an online BCI system. The first session was used in training phase and reflected a small sample problem. The rest three sessions were used as test data, which possessed the characteristics of time-varying and instability of MI-EEG. We compared the online recognition capabilities among the three algorithms: SGRM, SFBCSP and RLS-CSP, which SFBCSP and RLS-CSP features were classified with ISCDC. Here, we discarded the single-trial with one more label.

V. RESULTS

A. A PERFORMANCE COMPARED WITH DIFFERENT SPATIAL FILTERS

Table 1 and Fig.3 summarize the experimental results for the experiment I, where the average of five repetitions is reported for each subject. Table 1 shows that the RLS-CSP algorithm achieves the overall superior performance compared to other CSP-based algorithms in all scenarios. Specifically, RLS-CSP achieved the highest average accuracy 84.94%, and

**FIGURE 3.** Average accuracy of all subjects with respect to different training sizes.

its performance was improved with the increase of training size, which got its peak 90.46% when the number of training size was 100 trials. All of the methods tend to provide stable classification when using sufficient (more than 80 trials) training samples. The results illustrate that RLS-CSP method can gradually remedy the covariance matrix estimation bias caused by SSS, and update the spatial filter through high confidence test samples to adapt to the time-varying in EEG.

Moreover, as shown in Table 1, the average accuracy of SFBCSP and SGRM were 83.09% and 79.92%, and these of CCSP, TRCSP and FBRCSP were 77.19%, 77.71% and 81.04%, respectively. The sparse-based CSP algorithms were better than the regularized-based CSP algorithms. The possible reason is that in the framework of group sparse learning, compared with the regularization methods, it can eliminate the less relevant features of the non-target objects more rapidly, which contributes more significant test samples, thus improves the classification performance. Besides, since the optimal filter band is generally subject-specific, many researches have proved the filter bands are more effective than a wide frequency band.

Table 2 and Table 3 summarize the classification accuracy of each target subject using 40 and 80 training trials, respectively. Paired-sample t-test was used to investigate the statistical significance of accuracy difference between

TABLE 2. The accuracy (%) on dataset1. The number of training samples from target subject is 40. And p-values are derived by the paired t-tests between results of RLS-CSP and each of other methods.

	RLS-CSP	CSP	SFBCSP	CCSP	TRCSP	FBCRCSP	SGRM
AA	80.42	74.28	79.54	72.17	74.48	76.47	76.13
AL	98.64	93.14	97.37	93.36	93.28	96.57	93.51
AV	72.46	67.73	71.89	68.38	67.97	70	64.45
AW	80.15	73.62	79.61	77.36	76.28	70.28	78.54
AY	85.21	74.14	82.07	71.73	74.15	78.43	80.75
p-value		$\rho < 0.001$	$\rho < 0.05$	$\rho < 0.001$	$\rho < 0.01$	$\rho < 0.05$	$\rho < 0.05$

TABLE 3. The accuracy (%) on of each subject. The number of training samples from target subject is 80. And p-values are derived by the paired t-tests between results of RLS-CSP and each of other methods.

	RLS-CSP	CSP	SFBCSP	CCSP	TRCSP	FBCRCSP	SGRM
AA	89.14	77.86	83.54	76.24	77.91	84.78	80.91
AL	98.64	96.14	98.64	93.36	93.28	96.57	94.56
AV	73.14	71.45	74.89	71.37	71.93	72.07	75.45
AW	84.37	76.62	82.81	79.61	77.85	82.39	85.75
AY	83.69	75.71	84.67	76.21	76.88	78.43	82.49
p-value		$\rho < 0.05$	$\rho = 0.074$	$\rho < 0.01$	$\rho < 0.01$	$\rho < 0.05$	$\rho = 0.193$

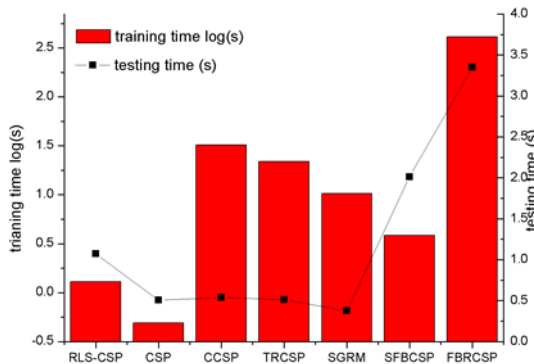


FIGURE 4. The training time and test time for each CSP-based method.

these compared methods (see ρ -value). Table2 indicates that RLS-CSP yields significantly higher accuracy than the others in 40 trials and its performance difference is up to 95%. Nevertheless, Table3 shows sparse-based methods perform better than RLS-CSP. SGRM had a large performance of 75.45% and 85.75% for subject 'AW' and 'AV', and SFBCSP was 84.67% for 'AY'. There was no significant difference between SFBCSP, SGRM and RLS-CSP. In summary, RLS-CSP can improve the robustness and adaptability of a feature model in SSS. However, its advantage is not significantly when the training samples are sufficient.

Fig.4 depicts the training time and testing time for each method, under the environment of MatlabR20116b on a computer with 3.40GHz CPU(i5-7500, 8G RAM). In particular, RLS-CSP training time refers to a process of "offline

spatial filters initialization–ISDCD-based evaluation–spatial filters update", and its testing time is the time to complete "ISDCD-Classification and spatial filters update". The regularization CSP methods(CCSP,TRCSP, FBCRCSP) required much longer training time than the non-regularization algorithms, and the FBCRCSP was up to 410.204s. The sparse-based methods(SGRM, SFBCSP) took a moderate computational efficiency. It is because cross-validation for parameter selection is time-consuming. Since RLS-CSP and CSP did not need cross-validation, the training time was short about 1.293s and 0.492s, respectively. The filter bands-based CSP methods required much more testing time than other algorithms. RLS-CSP method was about 1s for a single-trial testing, which was acceptable in an online test. Therefore, considered both time-consuming and performance in SSS, RLS-CSP presents a comparable advantage than others.

B. EVALUATION ON THE EFFECT OF ADAPTION

To study the effect of non-stationarity on the feature distribution, experiments were carried out as described in Experiment II. Fig.5 demonstrates the testing feature distribution of CSP and RLS-CSP of subject "AY". The x-axis represents the normal vector of the classifier of hyperplane, and the y-axis is the largest PCA-component of testing features. It can be seen from the Fig.5 that the feature distributions of the two testing sets were transferred due to the non-stationarities in EEG signal. For CSP, the features underwent not only a rotation between 120st and 120nd but also their distribution changed, which the 120st had smaller DR

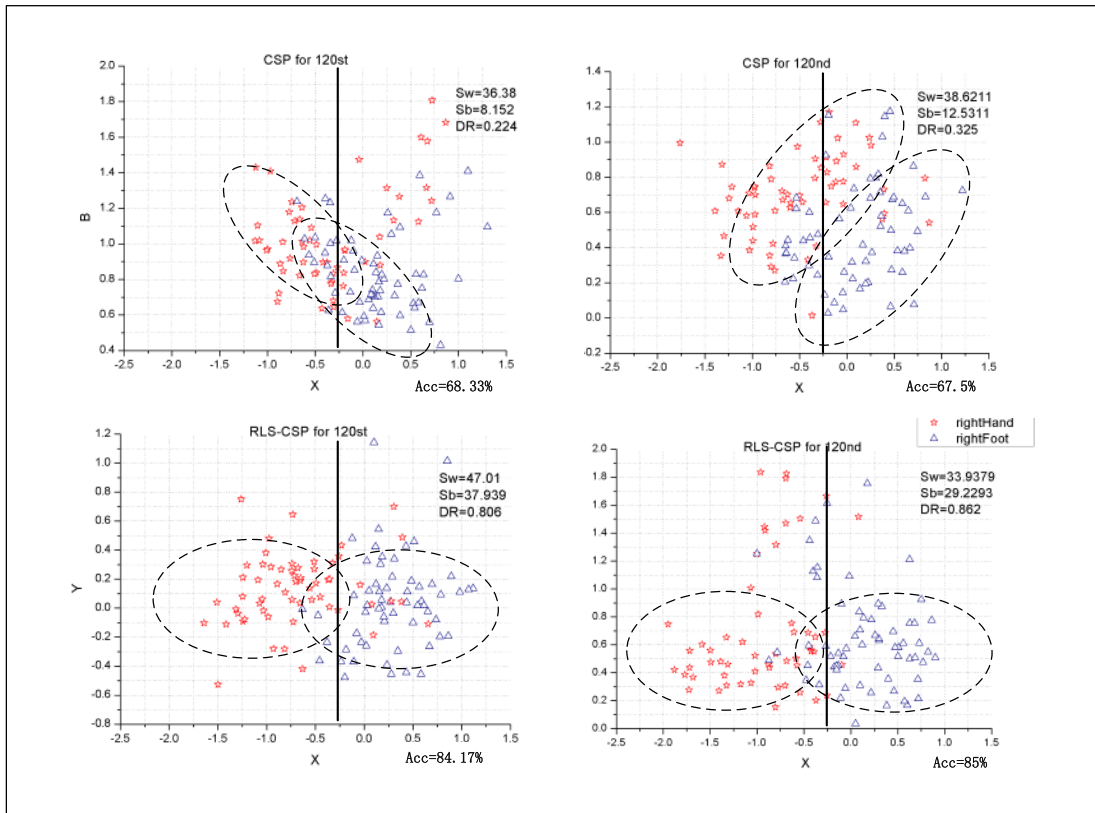


FIGURE 5. The test features for CSP and RLS-CSP of subject "AY".

than the latter. For RLS-CSP, the features were more robust and discriminative than those of CSP, with DR of 0.806 and 0.862, respectively. Under the same testing set, the overlap of RLS-CSP features was smaller than those of CSP, an DR were improved 0.582 and 0.537, respectively. Therefore, it is evident that RLS-CSP is more robust and discriminative than CSP, and more powerful when dealing with time-varying EEG signals.

Table 4 shows the recognition accuracy of RLS-CSP in the subject-independent. It was evident that the performance of RLS-CSP in the subject-independent scenario is poor, which fully illustrated the impact of individual differences on classification. Further observation showed that in general, the accuracy of 120nd training set was higher than that of the 120st. It demonstrates that RLS-CSP is adaptive and can compensate for the initial deviation of the spatial filters. However, once the deviation is too large, the repair degree is limited. Therefore, RLS-CSP is more suitable for a subject-specific scenario.

C. EVALUATION OF CLASSIFICATION PERFORMANCE

The effectiveness of RLS-CSP is based on the merging high-confidence testing samples and filtering outliers. Table 1 reports the results for ISCDC concerning different training sizes and compares them with DPC and DBSCAN. As can be seen in Table 5 that ISCDC was superior to other two methods

with different data sets for all subject, and the performance of all methods were improved with the increase of training size. This may be that in SSS (below 80 trials in our experiments) the cluster center is only a local optimum, not a global one, so the recognition accuracy of DPC and DBSCAN depending on the cluster center are poor, whereas ISCDC is based on the distribution of all test features, the diversity of distributions are conducive to ISCDC classification. Sufficient samples can reduce the deviation of cluster center, and provide a variety of sample distribution information, which expands the neighborhood set of ISCDC, and improves the recognition rate of the three algorithms. In addition, ISCDC has the function of non-parameter singular point filtering, which can gradually overcome the model instability caused by the poor quality of training samples.

D. PERFORMANCE ON THE CONTINUOUS IMAGERY

To verify the application value of proposed adaptive scheme in online BCI system, experiment IV was carried out on Dataset II. Fig. 6 demonstrates that SFBCSP has a high accuracy in the first session of all subjects, and the proposed scheme in our paper performs better in the second and third session. The average accuracies of our method were 78.91%, 70.41% and 59.14% respectively, and compared with the winners (79.6%, 70.31% and 56.02%, respectively), the algorithm used less training samples, which proved the

TABLE 4. The accuracy (%) of subject-independent on two orderly sets. Note: The numbers in solid frames represent results of subject-specific.

source	AA		AL		AV		AW		AY	
	120st	120nd	120st	120nd	120st	120nd	120st	120nd	120st	120nd
AA	89.17	79.67	72.17	75.5	60.5	67.17	60.50	61.33	55.50	58.00
AL	59.67	60.5	98.33	97.5	57.17	57.17	53.83	59.67	79.67	85.50
AV	56.33	58.83	55.5	59.67	76.67	75	52.17	58.83	54.67	57.17
AW	57.5	55.83	63.33	60.83	60.83	60.83	82.45	80.62	55.83	57.50
AY	61.17	63.67	60.33	64.5	62.83	62	60.33	72.0	84.17	85.00
AA+AL	70.5	70.5	97.5	94.17	58	58	53.83	59.67	77.17	78.00
AL+AV	58.83	60.5	98	93	75.5	74.67	48.00	60.50	73.83	77.17
AV+AW	63.83	73	46.33	50.5	66.33	73.83	59.67	66.33	55.50	57.17
AW+AY	58	61.33	44.67	55.5	58.83	58	60.50	56.33	50.50	53.83
AA+AL+AV	58	65.5	93.33	93.33	61.33	68.83	73.83	78.0	54.67	59.67
AL+AV+AW	58.83	57.17	91.33	91.33	57.17	57.17	59.67	65.50	60.50	65.50
AV+AW+AY	60.5	65.5	58	54.67	60.5	68.83	62.17	58.00	54.67	65.50
AA+AL+AV+AW	67.67	69.17	95.5	98.33	57.17	58	66.33	71.33	58.83	60.50
AL+AV+AW+AY	60.5	65.5	60.5	57.17	58	61.33	61.33	68.00	54.67	57.17

TABLE 5. The average accuracy (%) for dataset2 in different training size by the ISCDC, DCP and DBSCAN method.

Method	10	20	30	40	80	100	120	140
ISCDC	78.87±11.33	80.27±10.55	82.05±9.69	83.38±9.68	85.41±9.12	87.05±7.43	88.67±7.55	88.53±7.44
DPC	67.29±13.94	69.77±13.92	74.6±13.62	78.03±11.69	82.63±8.93	84.83±10.04	83.62±8.51	84.78±8.11
DBSCAN	66.97±13.29	68.96±12.15	73.21±11.97	77.24±10.56	83.74±9.19	83.68±9.7	84.72±8.59	83.85±7.6
mean± std	66.97±12.48	69.65±11.75	73.81±10.34	77.07±8.83	83.66±8.78	84.57±8.63	85.32±8.43	85.09±8.3

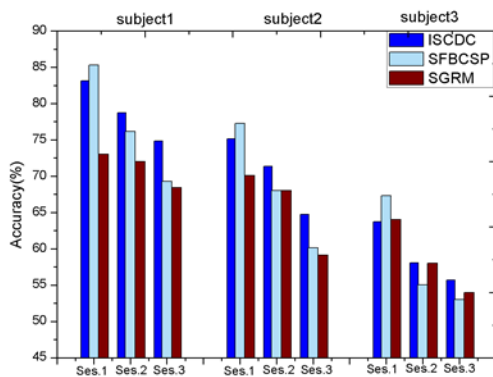


FIGURE 6. The performance of experiment IV on dataset II.

advantages of our algorithm in small sample online BCI. The results indicate that our scheme is capable to update the feature model and classification model, and can better adapt to the online time-varying system.

VI. CONCLUSION

Aiming at the problems of long calibration time and poor stability of an online motor imagery BCI system, in this

paper, a feature model based on recursive least square method (RLS-CSP) is proposed. RLS-CSP uses recursive least squares method to update the spatial filters, which allows incremental computation of the filter coefficients and incorporates new samples to update the filters. Moreover, an adaptive density clustering classification (ISCDC) is proposed to classify the testing sample and quantify its classification confidence.

To verify the efficacy of RLS-CSP spatial filters in SSS, we compared RLS-CSP with CSP, TRCSP, CCSP, SFBCSP, FBRCSP and SGRM in different training sets. As shown in Table 1, RLS-CSP demonstrates better performance. Fig.5 illustrates the feature distributions in RLS-CSP, which proves that RLS-CSP provides more robust and discriminative spatial filters than CSP. Meanwhile, RLS-CSP method is effective to non-stationary EEG signals in SSS. The self-training classification (ISCDC) works for both classification and confidence assessment. Table5 prove the adaptive classification ability of ISCDC. ISCDC achieves excellent performance in SSS and improves its stability by filtering outliers. By classifying in continuous EEG (Fig.6), our propose scheme shows the advantages for small sample setting situation and online

BCI system. Our method provides a new insight to shorten calibration phase and improve the adaptability of an online BCI system.

However, due the individual difference, RSL-CSP has poor recognition effect in subject-independent. In recent years, transfer learning has shown its promise in coping with the subject-transfer in BCI system [17]–[19]. Thus, a transfer learning extension of our RLS-CSP model could be benefit for coping with the inconsistency of data distributions. In this way, less training samples are required to initialize a spatial filter and improve the subject adaptability of RLS-CSP. In addition, with the accumulation of training samples, the computation cost of ISDCD will be increased. An alternative solution is to delete those training samples that do not contribute much to classification, which will be worth our future investigation.

CONFLICT OF INTERESTS

The authors declare no conflict of interests regarding the publication of this article.

ACKNOWLEDGMENT

The authors would like to acknowledge the BCI Competition III Dataset Iva and BCI Competition III dataset V, which was used to test the algorithms proposed in this study.

REFERENCES

- [1] F. Lotte, L. Bougrain, and A. Cichocki et al., "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jul. 2018, Art. no. 031005, doi: [10.1088/1741-2552/aab2f2](https://doi.org/10.1088/1741-2552/aab2f2).
- [2] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018, doi: [10.1109/TNSRE.2017.2757519](https://doi.org/10.1109/TNSRE.2017.2757519).
- [3] F. P. Kalaganis, E. Chatzilari, S. Nikolopoulos, I. Kompatsiaris, and N. A. Laskaris, "An error-aware gaze-based keyboard by means of a hybrid BCI system," *Sci. Rep.*, vol. 8, no. 1, pp. 257–268, Dec. 2018.
- [4] R. Chaisaen, P. Autthasan, N. Mingchinda, P. Leelaarporn, N. Kunaseth, S. Tammajarung, P. Manoonpong, S. C. Mukhopadhyay, and T. Wilaiprasitporn, "Decoding EEG rhythms during action observation, motor imagery, and execution for standing and sitting," *IEEE Sensors J.*, early access, Jun. 30, 2020, doi: [10.1109/JSEN.2020.3005968](https://doi.org/10.1109/JSEN.2020.3005968).
- [5] S. A. C. Yohanandan, I. Kiral-Kornek, and J. and Tang, "A robust low-cost eeg motor imagery-based brain-computer interface," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* Jul. 2018, pp. 5089–5092.
- [6] K. P. Thomas, C. Guan, C. Tong Lau, A. P. Vinod, and K. Keng 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, doi: [10.1109/TBME.2009.2026181](https://doi.org/10.1109/TBME.2009.2026181).
- [7] S. Kumar and A. Sharma, "A new parameter tuning approach for enhanced motor imagery EEG signal classification," *Med. Biol. Eng. Comput.*, vol. 2018, pp. 1861–1874, Apr. 2018.
- [8] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019, doi: [10.1109/JBHI.2018.2832538](https://doi.org/10.1109/JBHI.2018.2832538).
- [9] 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, doi: [10.1109/TBME.2010.2082539](https://doi.org/10.1109/TBME.2010.2082539).
- [10] 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, doi: [10.1109/TBME.2010.2082540](https://doi.org/10.1109/TBME.2010.2082540).
- [11] S.-H. Park and S.-G. Lee, "Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern," *IEEE Sensors J.*, vol. 17, no. 10, pp. 2977–2983, May 2017, doi: [10.1109/JSEN.2017.2671842](https://doi.org/10.1109/JSEN.2017.2671842).
- [12] T. N. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, F. M. Alotaiby, K. Aljibrin, S. R. Alrshod, I. M. Alkhanin, and N. Alrajhi, "A common spatial pattern approach for scalp EEG seizure detection," in *Proc. 5th Int. Conf. Electron. Devices, Syst. Appl. (ICEDSA)*, Dec. 2016, pp. 1–5, doi: [10.1109/ICEDSA.2016.7818564](https://doi.org/10.1109/ICEDSA.2016.7818564).
- [13] R. Zhang, X. Xiao, Z. Liu, W. Jiang, J. Li, Y. Cao, J. Ren, D. Jiang, and L. Cui, "A new motor imagery EEG classification method FB-TRCSP+RF based on CSP and random forest," *IEEE Access*, vol. 6, pp. 44944–44950, 2018, doi: [10.1109/ACCESS.2018.2860633](https://doi.org/10.1109/ACCESS.2018.2860633).
- [14] W. Samek, C. Vidaurre, and K. R. Mueller, "Stationary common spatial patterns for brain-computer interfacing," *J. Neural Eng.*, vol. 9, no. 12, pp. 13–26, 2012.
- [15] T. Ha Nguyen, S.-M. Park, K.-E. Ko, and K.-B. Sim, "Invariant common spatial pattern advanced feature extraction in μ rhythms of EEG signals," in *Proc. 38th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Montreal, QC, Canada, Oct. 2012, pp. 1535–1539, doi: [10.1109/IECON.2012.6388513](https://doi.org/10.1109/IECON.2012.6388513).
- [16] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, "Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–13, 2016.
- [17] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jul. 7, 2020, doi: [10.1109/TCDS.2020.3007453](https://doi.org/10.1109/TCDS.2020.3007453).
- [18] H. He and D. Wu, "Transfer learning for Brain-Computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.
- [19] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 465–479, Mar. 2012.
- [20] Y. Shin, S. Lee, J. Lee, and H.-N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems," *J. Neural Eng.*, vol. 9, no. 5, Oct. 2012, Art. no. 056002.
- [21] Y. Zhang, G. Zhou, J. Jin, X. 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, doi: [10.1016/j.jneumeth.2015.08.004](https://doi.org/10.1016/j.jneumeth.2015.08.004).
- [22] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. 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.
- [23] M. Chen, X. Tan, and L. Zhang, "An iterative self-training support vector machine algorithm in brain-computer interfaces," *Intell. Data Anal.*, vol. 20, no. 1, pp. 67–82, Jan. 2016.
- [24] H. Woehle, M. M. Krell, S. Straube, S. K. Kim, E. A. Kirchner, and F. Kirchner, "An adaptive spatial filter for user-independent single trial detection of event-related potentials," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1696–1705, Jul. 2015, doi: [10.1109/TBME.2015.2402252](https://doi.org/10.1109/TBME.2015.2402252).
- [25] J. Li, Q. Zhu, and Q. Wu, "A self-training method based on density peaks and an extended parameter-free local noise filter for k nearest neighbor," *Knowl.-Based Syst.*, vol. 184, Nov. 2019, Art. no. 104895.
- [26] A. Rodriguez and L. Alessandro, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [27] E. Martin, H. P. Kriegel, and J. Sander, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, pp. 226–231.
- [28] D. Wu, M. Shang, X. Luo, J. Xu, H. Yan, W. Deng, and G. Wang, "Self-training semi-supervised classification based on density peaks of data," *Neurocomputing*, vol. 275, pp. 180–191, Jan. 2018.
- [29] H. Gan, N. Sang, R. Huang, X. Tong, and Z. Dan, "Using clustering analysis to improve semi-supervised classification," *Neurocomputing*, vol. 101, pp. 290–298, Feb. 2013.
- [30] Y. N. Rao and J. C. Principe, "An RLS type algorithm for generalized eigendecomposition," in *Proc. Neural Netw. for Signal Process. XI: IEEE Signal Process. Soc. Workshop*, North Falmouth, MA, USA, Sep. 2001, pp. 263–272, doi: [10.1109/NNSP.2001.943131](https://doi.org/10.1109/NNSP.2001.943131).
- [31] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 551–575, Aug. 1977, doi: [10.1109/TAC.1977.1101561](https://doi.org/10.1109/TAC.1977.1101561).



QIN JIANG received the B.S. and M.S. degrees in biomedical engineering from the Chongqing University of Technology, China. She is currently pursuing the Ph.D. degree in pattern recognition and knowledge discovery with the Chongqing University of Posts and Telecommunications. Her research interests include machine learning, pattern recognition, biomedical signal process, and brain-computer interface.



GENGYU GE received the B.S. degree in information and computing science from Chuzhou University, Chuzhou, China, in 2011, and the M.S. degree in computer system architecture from Southwest University, Chongqing, China, in 2014. He is currently pursuing the Ph.D. degree with the Chongqing University of Posts and Telecommunications. His research interests include mobile robot navigation, semantic slam, and embedded system application.



YI ZHANG received the Ph.D. degree in mechanical manufacturing and automation from the Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. training in intelligent multimode human-computer interaction with the University of Essex, London, U.K. He is currently a Professor and a Ph.D. Supervisor with the Advanced Manufacturing Engineering College, Chongqing University of Posts and Telecommunications. His research interests include robot automatic control and human-computer interaction.



ZHIRONG XIE received the B.S. degree in electronic information engineering from the Sichuan University of Science and Engineering, Zigong, China. He is currently pursuing the M.S. degree with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include deep learning and pattern recognition.

...