IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING FOR COMPUTER-AIDED MEDICAL DIAGNOSIS

Received July 28, 2019, accepted August 6, 2019, date of publication August 12, 2019, date of current version August 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2934519

A General Common Spatial Patterns for EEG Analysis With Applications to Vigilance Detection

HONGBIN YU[1], HONGTAO LU[2,3], SHUIHUA WANG[4], KAIJIAN XIA[5], YIZHANG JIANG[1], AND PENGJIANG QIAN[1]

[1]School of Digital Media, Jiangnan University, Wuxi 214122, China
[2]MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, Shanghai 200240, China
[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[4]School of Architecture Building and Civil engineering, Loughborough University, Loughborough LE11 3TU, U.K.
[5]Soochow University affiliated Changshu Hospital—Changshu No.1 People's Hospital, Changshu 215500, China

Corresponding authors: Pengjiang Qian (qianpjiang@jiangnan.edu.cn) and Hongtao Lu (lu-ht@cs.sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241 and Grant 61702225, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160187, in part by the 2016 Qinglan Project of Jiangsu Province, in part by the 2016 Six Talent Peaks Project of Jiangsu Province, and in part by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002.

ABSTRACT Vigilance or sustained attention is an important aspect for people who engaged in long time attention demanding tasks such as monotonous monitoring and driving. Vigilance detection has been an important topic in the field of brain-computer interface (BCI) research. However, the study is limited due to the low SNR (Signal-Noise Ratio) nature of EEG. Common spatial pattern (CSP) is a one of the most effective algorithms for feature extraction method in the BCI study area. The CSP seeks for an optimal projection direction (spatial filter) by maximizing the variance of one class and simultaneously minimizing the variance of the other class. There is one drawbacks exists in the traditional CSP, that is, the CSP is proposed relies on the assumption that data in each class follow the Gaussian distribution. However, this assumption is not always true for EEG data in practice, especially in the research of vigilance detection based EEG (e.g. during sleep). Thus, traditional CSP suffers performance degradation in case of non-Gaussian distributions. In this paper, we extend the traditional CSP to the general version and proposed nonparametric CSP (NCSP) algorithms which do not explicitly rely on the assumption of the underlying class Gaussian distribution and we then develop a new efficient algorithm based on matrix deflation to solve the proposed NCSP algorithm and its extensions-nonparametric multi-class CSP (NMCSP). Experimental results on EEG-based vigilance estimation and motor imagery recognition task demonstrate the effectiveness and efficiency of our proposed algorithms.

INDEX TERMS Common spatial pattern, nonparametric multiclass CSP, nonparametric CSP, EEG, vigilance detection.

I. INTRODUCTION

Electroencephalography(EEG) is the physiological method of choice to record the electrical activity generated by the brain via electrodes placed on the scalp surface. Studies based on EEG measure the brain's spontaneous electrical activity changes in response to stimulation of sight, sound, or touch. These researches explored some potential applications based on EEG signal to help human to solve some significant troubles, such as disease detection and treatment, stroke patients' rehabilitation as well as paralyzed patients' walking and communications [5], [6], [39]. In recent years, many progress has obtained in noninvasive brain computer interface (BCI) technology [3], [4] and EEG signal analysis [1], [2].

Many occupations require the human keep the mental status with high vigilance, e.g., monitoring power plan, controlling air traffics, driving trucks and high-speed trains [8].

The associate editor coordinating the review of this article and approving it for publication was Yudong Zhang.

Scientist have worked on the research of human vigilance for several decades and much significant progress has made. Many biological and behavioral characteristics such as eye-closure, face expression, head position, heart beating and reaction time [10]–[15], etc. have been used for vigilance detection. EEG signals are also strongly associated with the human vigilance, it can reflect the changes in the state of human's brain, which makes it possible to predict human alertness based on EEG. Neurophysiological studies have shown that human vigilance is a process of dynamic changing, during rest, human mental state toggles among different vigilance levels and the vigilance declines from alertness state (high degree vigilance) to sleepy state (the lowest degree vigilance) and the rhythm of the brain waves varies with the change of the mental states. The β rhythm (14-30Hz) decreases when human's vigilance level decreased, at the same time, the α rhythm (8-13Hz) increases [16], [17].

Vigilance detection based on EEG has always been a hot topic in BCI field. However, the nature of low signal-to-noise ratio (SNR) is one of the most important methodological challenges for EEG data collection and analysis. Researchers have worked hard for several years to explore artifacts removal, dimensionality reduction and feature learning methods to extract discriminant features for EEG analysis, such as principal component analysis (PCA, maximum power feature) [18], [19], independent component analysis (ICA, statistical independence feature) [20] and common spatial pattern (CSP, maximum power ratio feature) [1], [21] [22]–[24], of which CSP is the most widely used one [25].

Given two different classes EEG signals, classical CSP seeks for an optimal projection direction (spatial filter) by maximizing the variance of one class and simultaneously minimizing the variance of the other class. Raw EEG signals are then mapped to this direction, after that, classification become more easier.

CSP projection is actually a spatial filtering process, which was originally designed for two classification problems. It has been extended to multi-class scenarios in several ways [26]–[28]. Dornhege *et al.* proposed a multi-class CSP by jointly approximate diagonalization (JAD) of covariance matrices [26]. Unlike the two-class CSP where two matrices are jointly diagonalized exactly, simultaneous diagonalization of more than two matrices cannot be exactly achieved, only joint approximate diagonalization can be attained. Moritz *et al.* proposed an information theory based multi-class CSP [27], where they show that the two-class CSP maximizes an approximation of mutual information of extracted EEG components and class labels. Based on this relationship, they provided a component selection method in JAD multi-class CSP paradigms by maximizing mutual information of components and class labels, overcoming the drawback of heuristics in multi-class CSP in [26].

In addition to multi-classification CSP algorithms, there are many other versions of CSP algorithms. Falzon *et al.* [29] [30] introduce the complex extension of CSP, which combine the raw EEG signal and its phase information together to implement the classification and recognition task. Park *et al.* [31] improve Falzon's work and proposed an augmented complex common spatial patterns algorithm (ACCSP) and implement the ACCSP algorithm to the motor imagery task get a considerable result. Kawanabe *et al.* [32] applied the maxmin approach to CSP and introduce a robust Common Spatial filter algorithm and improve the performance of the traditional CSP. Zheng and Lin [28] presented an optimal multi-class CSP based on the bayesian error estimation for EEG feature extraction, our proposed general common spatial pattern algorithm is motivated by this method but we have different criteria function.

In this paper, we propose a new CSP algorithm to extract the EEG feature for vigilance estimation. Inspired by the nonparametric linear discriminant analysis (LDA) [33] which was designed to handle the non-Gaussian problem in LDA, in this paper, we first propose a two-class nonparametric common spatial pattern (NCSP) algorithm. In the NCSP, we construct a within class scatter matrix to measure the clustering property of each class, and then simply use the within class scatter matrix to replace the sample variance matrix in the traditional CSP. We also propose a novel multi-class CSP (MCSP) that has an explantation of minimizing Bayesian classification error, but with a different optimization criteria from that in [28], and then extend this multi-class CSP to nonparametric multi-class CSP (NMCSP).
We apply our proposed algorithms to the vigilance estimation as well as the motor imagery task to measure their performance on feature extraction task, comparisons are made with the multi-class CSP in [28], ACCSP in [31], RCSP in [23] as well as the CSP-L1 [34]. Preliminary experiments show better results than the state-of-the-art methods.

The paper is organized as follows: Firstly, we review the traditional CSP and multi-class CSP algorithms in Section II. In Section III, we propose the two-class non-parameter CSP algorithm. In Section IV, A new multi-class CSP algorithm based on a different criterion is proposed, then extend it to our multi-class nonparameteric CSP, the new algorithm is given at last. We validate the performance of our proposed new CSP algorithms on the designed vigilance estimation as well as the motor imagery experiments. The performance comparison result with the baseline as well as the state-of-art algorithms are presented in Section V. Finally, In Section VI concludes our work.

II. CSP AND MULTI-CLASS CSP

In this section, we briefly introduce the classical common spatial pattern (CSP) algorithm and its extensions to multi-class problems. To be more precise, let $\mathbf{X}^c = [\mathbf{x}_1^c, \mathbf{x}_2^c, \ldots, \mathbf{x}_{t_c}^c]$, $c = 1, 2, \ldots C$, are EEG data matrices, where C is the number of classes, $\mathbf{x}_i^c \in R^{D\times S}$ is a $D\times S$ matrix that represents the raw EEG data of the ith trial for class c, $i = 1, 2, \ldots, t_c$, and t_c is the total trial number for class c, D is the number of channels and S is the number of samples in the ith trial of class c. For the classical CSP, $C = 2$. The normalized spatial

covariance matrices for each class are given as

$$\mathbf{R}_c = \frac{\mathbf{X}^c \mathbf{X}^{cT}}{\mathrm{tr}(\mathbf{X}^c \mathbf{X}^{cT})}, \quad c = 1, 2, \ldots, C \tag{1}$$

where tr(·) means the trace of a matrix, and $\mathbf{X}^{cT}$ denotes the transpose of $\mathbf{X}^c$.

A. CSP

CSP is a popular algorithm for calculating spatial filters, and is widely used for extracting features in BCI systems based on event-related (de)synchronization (ERD/ERS) [25], [35]. Given samples corresponding to two classes (e.g., right and left motor imagery) in a high-dimensional space, CSP finds projection directions (i.e.,spatial filters) that maximize the variance for one class and at the same time minimize the variance for the other class.

Finding a projection vector w_{csp} involves solving the following optimization problem,

$$w_{\mathrm{csp}} = \arg\max_w \frac{w^T \mathbf{R}_2\, w}{w^T \mathbf{R}_1 w} \tag{2}$$

or

$$w_{\mathrm{csp}} = \arg\max_w \frac{w^T \mathbf{R}_1\, w}{w^T \mathbf{R}_2 w} \tag{3}$$

The solution to (2) can be obtained by the generalized eigenvector of $\mathbf{R}_1^{-1}\mathbf{R}_2$ ($\mathbf{R}_2^{-1}\mathbf{R}_1$ for (3)) corresponding to the maximal eigenvalue. A limitation of this approach is the sensitivity of the covariance matrices to outliers. For example, it has been found that the largest (or the smallest) eigenvector often captures the myographic activity [1]. Therefore, it has been suggested to consider more than one eigenvectors associated with the largest and lowest eigenvalue ratios [1], [25]. This amounts to calculate a matrix W_{csp} such that

$$W_{\mathrm{csp}} = \arg\max_W \frac{\mathrm{tr}(W^T \mathbf{R}_2 W)}{\mathrm{tr}(W^T \mathbf{R}_1 W)} \tag{4}$$

or

$$W_{\mathrm{csp}} = \arg\max_W \frac{\mathrm{tr}(W^T \mathbf{R}_1\, W)}{\mathrm{tr}(W^T \mathbf{R}_2 W)} \tag{5}$$

Using this projection matrix W_{csp}, the EEG recordings $\mathbf{X}^c$ are linearly projected to $\mathbf{Y}^c$ by

$$\mathbf{Y}^c = W_{\mathrm{csp}}^T \mathbf{X}^c \tag{6}$$

There is an alternative way to solve (4) and (5), where W_{csp} can be obtained by simultaneously diagonalizing the covariance matrices $\mathbf{R}_c$ [25], i.e.,

$$W^T \mathbf{R}_c W = \Lambda_c, \quad c = 1, 2 \tag{7}$$

where Λ_c are diagonal matrices.

B. MULTI-CLASS CSP

How many brain states should be used in a BCI system is an open question, it has been reported that the use of more classes yields better BCI performance than the use of two classes [36]. It has also been pointed out that using more classes has the potential to increase information transfer rate, although the classification performance decreases [26]. On the other hand, in many scenarios, multi-class is essential, for example, in vigilance estimation, different vigilance states involve multiple classes and even continuous states. So, extending the traditional (two-class) CSP to multi-class is of significance. Several extensions of CSP to multi-class paradigms have been proposed till now [26]–[28]. Here we only briefly introduce the multi-class CSP proposed by Zheng and Lin [28] which is closely related to our multi-class CSP developed below. This multi-class CSP is based on the Bayesian error estimation for EEG classification, and the underlying class distribution is assumed to be Gaussian. By assuming all classes have the same prior probability, an upper bound of the Bayesian classification error ϵ has been derived as follows [28]

$$\epsilon \le \sum_{c=1}^{C-1} \sum_{j=c+1}^{C} \frac{1}{C} - \frac{1}{8}\Big(\frac{C^{-2/3} \sum_{c=1}^{C} |w^T(\mathbf{R}_c - \mathbf{R})w|}{2w^T \mathbf{R} w}\Big)^2 \tag{8}$$

where $\mathbf{R} = \frac{1}{C}\sum_{c=1}^{C} \mathbf{R}_c$ is the total covariance matrix. Their multi-class CSP (MCSP) is obtained by minimizing the upper bound (8) which is equivalent to maximize the following criterion

$$w_{\mathrm{mcsp}} = \arg\max_w \frac{\sum_{c=1}^{C} |w^T(\mathbf{R}_c - \mathbf{R})w|}{w^T \mathbf{R} w}. \tag{9}$$

r optimal CSP projection directions can be defined by the following optimization

$$\begin{aligned} w_1 &= \arg\max_w \frac{\sum_{c=1}^{C} |w^T(\mathbf{R}_c - \mathbf{R})w|}{w^T \mathbf{R} w} \\ &\cdots \\ w_j &= \arg \max_{\substack{w^T \mathbf{R} w_j = 0, \\ j = 1, \ldots, r-1}} \frac{\sum_{c=1}^{C} |w^T(\mathbf{R}_c - \mathbf{R})w|}{w^T \mathbf{R} w} \end{aligned} \tag{10}$$

An approximate iteration algorithm has been developed in [28] to sequentially calculate the optimal spatial filters $w_1, \ldots, w_r$. Although this MCSP is derived from the closed-form Bayesian error estimation and thus has solid theoretical foundation, it also has two shortcomings. The first is that the upper bound of the Bayesian error estimation (8) is too loose, because the estimation is amplified for several times by applying inequalities. This can also be observed from (9), maximizing the sum of the differences between all $\mathbf{R}_c$ and the total covariance matrix $\mathbf{R}$ does not necessarily maximize the differences among pairs of $\mathbf{R}_c$. The second is the assumption of Gaussian distribution for each class as mentioned above. Moreover, the equal class prior probability assumption is also too strong. For motor imagery task, it seems reasonable, but for other activities such as vigilance estimation, the prior

probabilities for alert, drowsy and sleep are apparently different. To handle the first shortcoming, we propose a new multi-class CSP with different optimization criteria. To overcome the second shortcoming, we extend the proposed new multi-class CSP to nonparametric version.

III. TWO-CLASS NONPARAMETRIC CSP

The traditional CSP and its multi-class extensions are based on the assumption of class Gaussian distribution. Therefore, these methods suffer from one limitation which is the sample covariance matrices used in CSP and MCSP are sensitive to outliers. EEG signals consist of many artifacts and noise, and outliers are also common in EEG. Moreover, EEG may exhibit non-Gaussian distributions [37], particularly, in the sleep stage. This justifies the establishment of nonparametric CSP algorithms. In this section, we describe a two-class nonparametric CSP (NCSP) algorithm.

Here, we adopt the same notations as used in the above section. Given two classes with EEG data sample matrices $\mathbf{X}^1 = [\mathbf{x}_1^1, \mathbf{x}_2^1, \ldots, \mathbf{x}_{t_1}^1]$ and $\mathbf{X}^2 = [\mathbf{x}_1^2, \mathbf{x}_2^2, \ldots, \mathbf{x}_{t_2}^2]$, for a single trial EEG data $\mathbf{x}_i^c = [x_{i,1}^c, x_{i,2}^c, \ldots, x_{i,S}^c] \in R^{D\times S}$, where $x_{i,j}^c \in R^D$ is a D-dimensional vector which stands for the EEG sample values on all channels at one time instant. We can construct a nonparametric within-class scatter matrix for this trial as

$$S_i^c = \sum_{j=1}^{S}\sum_{p=1}^{k}(x_{i,j}^c - N_p(x_{i,j}^c))(x_{i,j}^c - N_p(x_{i,j}^c))^T \quad (11)$$

$$S_i^c = \frac{S_i^c}{\mathrm{tr}(S_i^c)}, \quad c = 1, 2 \quad (12)$$

where $N_p(x)$ denotes the pth nearest neighbor among the column vectors of this trial $\mathbf{x}_i^c$, k is a tunable parameter which indicates that we consider k nearest neighbors in defining the nonparametric scatter matrices. The total within-class nonparametric scatter matrix $\mathbf{S}_c$ is obtained by averaging over all trials of each class

$$\mathbf{S}_c = \frac{1}{t_c}\sum_{i=1}^{t_c} S_i^c, \quad c = 1, 2 \quad (13)$$

In our NCSP framework, we just replace the covariance matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ by $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively. So the optimal spatial filter (projection vector) w_{ncsp} of NCSP is solved by the following optimization problem:

$$w_{\mathrm{ncsp}} = \arg\max_w \frac{w^T\mathbf{S}_2\, w}{w^T\mathbf{S}_1 w} \quad (14)$$

or

$$w_{\mathrm{ncsp}} = \arg\max_w \frac{w^T\mathbf{S}_1\, w}{w^T\mathbf{S}_2 w} \quad (15)$$

thus, w_{ncsp} is the largest generalized eigenvector of the matrix $\mathbf{S}_1^{-1}\mathbf{S}_2$ (or $\mathbf{S}_2^{-1}\mathbf{S}_1$). If more than one projection vectors are needed, one can simultaneously diagonalize $\mathbf{S}_1$ and $\mathbf{S}_2$ as

$$W^T\mathbf{S}_c W = \Lambda_c, \quad c = 1, 2 \quad (16)$$

and multiple project vectors (rows of W_{ncsp}) can be obtained by picking up appropriate number of the rows of W associated with the largest and smallest diagonal entries of Λ_c. One should notice that when $k = S$, $\mathbf{S}_c$ in (13) is reduced to the sample covariance matrix $\mathbf{R}_c$, so our NCSP is a generalization of the traditional CSP.

IV. MULTI-CLASS NONPARAMETRIC CSP

We can extend the above two-class NCSP to multi-class NCSP in the same way, that is, we can construct nonparametric within scatter matrices $\mathbf{S}_c$ through (11) to (13) for each class of multiple classes, and then use them to replace the original covariance matrices as done in the above section. In doing so, several multi-class NCSPs can be established, the multi-class CSP algorithms in [26]–[28] all have their nonparametric versions. For example, we can make joint approximate diagonalization to the nonparametric within class scatter matrices $\mathbf{S}_c$, $c = 1, 2, \ldots, C$, to obtain the nonparametric CSP extension of [26]. One can also replace $\mathbf{R}_c$ in (9) with $\mathbf{S}_c$ to get the nonparametric CSP version of [28].

However, in this section, we will first propose a new multi-class CSP for Gaussian distribution paradigms, which has a theoretical explanation of minimizing the Bayes classification error, and then we extend it to multi-class NCSP. Our new multi-class CSP is related to that in [28], but has a different objective function, and overcomes some shortcomings of [28].

A. A NEW CRITERION FOR MULTI-CLASS CSP

For better theoretical explanation, in this subsection, we assume that we are given C classes, each follows a Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_c$: $p_c(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{R}_c)$, $c = 1, 2, \ldots, C$. It is well known that the Bayes classification error with respect to class i and j can be expressed by [28], [38]

$$\varepsilon_{ij} = \int \min(P_i p_i(\mathbf{x}), P_j p_j(\mathbf{x}))d\mathbf{x} \quad (17)$$

where P_i is the *a prior* probability of class i. As in [28], by using the inequality $\min(a, b) \le \sqrt{ab}$ for two arbitrary nonnegative numbers a and b, and notice the Gaussian distribution, this Bayes error can be estimated as follows

$$\begin{aligned}\varepsilon_{ij} &\le \int \sqrt{P_i P_j p_i(\mathbf{x}) p_j(\mathbf{x})}d\mathbf{x} \\ &= \sqrt{P_i P_j}\exp\Big(-\frac{1}{2}\ln\frac{|\mathbf{R}_{ij}|}{\sqrt{|\mathbf{R}_i||\mathbf{R}_j|}}\Big) \\ &= \sqrt{P_i P_j}\Big(\frac{|\mathbf{R}_{ij}|}{\sqrt{|\mathbf{R}_i||\mathbf{R}_j|}}\Big)^{-\frac{1}{2}}\end{aligned} \quad (18)$$

where $\mathbf{R}_{ij} = \frac{1}{2}(\mathbf{R}_i + \mathbf{R}_j)$, and $|\mathbf{R}|$ denotes the determinant of matrix $\mathbf{R}$. This inequality gives an upper bound for the Bayes error ε_{ij}. When the samples are projected to one dimension

by a projection vector w, the upper bound becomes

$$\varepsilon_{ij} \leq \sqrt{P_i P_j}\Big(\frac{w^T\mathbf{R}_{ij}w}{\sqrt{(w^T\mathbf{R}_i w)(w^T\mathbf{R}_j w)}}\Big)^{-\frac{1}{2}} \tag{19}$$

So the optimal projection vector for the two classes i and j can be obtained by minimizing the error bound of the right-hand side of (19), which is equivalent to maximize

$$\max_w \frac{w^T\mathbf{R}_{ij}w}{\sqrt{(w^T\mathbf{R}_i w)(w^T\mathbf{R}_j w)}} \tag{20}$$

This is further equivalent to

$$\max_w \frac{(w^T\mathbf{R}_{ij}w)^2}{(w^T\mathbf{R}_i w)(w^T\mathbf{R}_j w)} \tag{21}$$

Let $u = w^T\mathbf{R}_i w$, $v = w^T\mathbf{R}_j w$, (21) is maximization of the ratio between $\left(\frac{u+v}{2}\right)^2$ and uv, which is further equivalent to maximize their difference $\left(\frac{u+v}{2}\right)^2 - uv = \left(\frac{u-v}{2}\right)^2$. Thus, this is finally equivalent to maximize $|u-v|$ as

$$\max_w |w^T(\mathbf{R}_i - \mathbf{R}_j)w| \tag{22}$$

For C class problems, the upper bound of the Bayes error can be estimated as $\varepsilon \leq \sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\varepsilon_{ij}$ [28], [40]. So our new CSP (denote the project vector by w_{newcsp}) for multi-class can be obtained simply by

$$w_{\text{newcsp}} = \arg\max_{w,w^Tw=1}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}|w^T(\mathbf{R}_i - \mathbf{R}_j)w| \tag{23}$$

the constraint $w^Tw = 1$ is used to handle the scale ambiguity problem in the optimization process. This new CSP is obviously different from the multi-class CSPs developed in the literature. It is neither the joint approximate diagonalization approach, nor the one versus the rest approach [26], it is also not like the one in [28] where the objective function takes the form of (9) and is obviously different from (23). Note that when $C = 2$, (23) reduces to the traditional two class CSP, but the CSP in [28] does not coincide the traditional two class CSP. So our multi-class CSP is a true generalization of the traditional two-class CSP, while the multi-class CSP in [28] is not. Our method is somewhat like the approach where the multi-class classification is divided into pairwise two-class problems [41]. But they are actually different, for the latter, the two-class CSP is used for each two-class pair such that the variance of the projections is employed as input for a LDA classifier. In our approach, a total objective function is formed by summing up the objection functions of all two-class pairs, and then a single projection vector for all classes is obtained by maximizing this objective function.

It must be pointed out that although the object function in (23) is the sum of the object function for each two class pair that is optimal for that two classes in the Bayesian sense, it is not necessarily optimal for all classes, since maximization of the sum not necessarily maximizes each term. However, it is more likely that a suboptimal solution can be achieved by maximizing the sum, our experiment results below illustrate the effectiveness of this method. The objective function in (23) is also better than (9) because in (9), the goal is to maximize the differences between $\mathbf{R}_c$ and the total covariance matrix $\mathbf{R}$, large difference between $\mathbf{R}_c$ and $\mathbf{R}$ does not necessarily imply large differences between pairs of $\mathbf{R}_c$. In contrast, our objective function directly maximizes the differences between class covariance matrix pairs.

B. AN ALGORITHM FOR NEW MULTI-CLASS CSP

In this subsection, we develop an algorithm to solve our new multi-class CSP. Let s_{ij} denote the sign of $w^T(\mathbf{R}_i - \mathbf{R}_j)w$, then (23) turns to

$$w_{\text{newcsp}} = \arg\max_{w,w^Tw=1} w^T\Big(\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}s_{ij}(\mathbf{R}_i - \mathbf{R}_j)\Big)w \tag{24}$$

if follows that w_{newcsp} is exactly the principal component of the matrix $\mathbf{A}(s) \triangleq \sum_{i=1}^{C-1}\sum_{j=i+1}^{C}s_{ij}(\mathbf{R}_i - \mathbf{R}_j)$ associated with the largest eigenvalue, which can be solved by classical methods, e.g., power method, where $s \triangleq (s_{ij})$ denotes a $c(c-1)/2$ dimensional vector of 1 and -1 which represents a sign combination of s_{ij}. Since $\mathbf{A}(s)$ depends on s, let $\mathbf{S} = \{s_1, s_2, \ldots, s_{2^{c(c-1)/2}}\}$ be the set of all possible combinations of signs of s_{ij}, then (24) is actually

$$w_{\text{newcsp}} = \arg\max_{s\in\mathbf{S}}\max_{w,w^Tw=1} w^T\mathbf{A}(s)w. \tag{25}$$

Like in [28], when the class number C is small (usual the case for most EEG classification problems), (25) can be solved by a full search method over $\mathbf{S}$, there are totally $2^{c(c-1)/2}$ different sign combinations, for example, when $c = 3$, there are 8 sign combinations, same as in [28].

(25) only calculates one CSP direction (spatial filter), if more CSP projection vectors $w_1, \ldots, w_r$ need to be calculated, we can find them by the following optimization procedures

$$\begin{aligned} w_1 &= \arg\max_{s\in\mathbf{S}}\max_{w,w^Tw=1} w^T\mathbf{A}(s)w, \\ &\cdots \\ w_r &= \arg\max_{s\in\mathbf{S}}\max_{w,w^Tw=1,w^T\mathbf{U}_{r-1}=0} w^T\mathbf{A}(s)w. \end{aligned} \tag{26}$$

where $\mathbf{U}_{r-1} = [w_1, w_2, \ldots, w_{r-1}]$.

In traditional PCA, w_i can be sequentially solved by the Hotelling's deflation method. Hotelling's deflation is a simple and popular method to eliminate the influence of a given eigenvector from a matrix. For example, when w_1 has been calculated, one can eliminate w_1 from $\mathbf{A}(s)$ by

$$\mathbf{A}(s) \leftarrow \mathbf{A}(s) - w_1w_1^T\mathbf{A}(s)w_1w_1^T. \tag{27}$$

However, the Hotelling's deflation relies on the exact eigenvectors. When applying to this situation, it might be problematic, because here $\mathbf{A}(s)$ is not a single fixed matrix, but varies with s. w_1 is the leading eigenvector of one $\mathbf{A}(s_1)$ with

certain s_1, it might not be an eigenvector of other $\mathbf{A}(s)$ with $s \neq s_1$, so the deflation (27) may fail in this case [42].

To handle this problem, we adopt new deflation schemes from [42]. Our situation here is similar to that of [42], where in sparse PCA, the eigenvector is not the exact eigenvector, but a pseudoeigenvector, an approximate to the exact eigenvector. Several novel deflation methods that overcome the drawbacks of the Hotelling's deflation have been proposed in [42], such as projection deflation, Schur complement deflation. For example, after calculating w_1, the projection deflation reads

$$\mathbf{A}(s) \leftarrow (\mathbf{I} - w_1 w_1^T)\mathbf{A}(s)(\mathbf{I} - w_1 w_1^T). \tag{28}$$

where $\mathbf{I}$ is the $D \times D$ identity matrix. Here we adopt similar deflation scheme to calculate more than one CSP projection vectors. The procedure is described in **Algorithm 1**.

Algorithm 1 A New Multi-Class CSP

Input: EEG data matrices $\mathbf{X}^c = [\mathbf{x}_1^c, \mathbf{x}_2^c, \ldots, \mathbf{x}_{t_c}^c]$, for all classes $c \in \{1, 2, \ldots C\}$. Number of CSP vectors r.

1. Calculate covariance matrix $\mathbf{R}_c$ for each class according to (1).
2. Enumerate all $2^{c(c-1)/2}$ sign combinations s of s_{ij}, denoted by $\mathbf{S} = \{s_1, \ldots, s_{2^{c(c-1)/2}}\}$
3. Calculate $\mathbf{A}(s_j)$ by $\mathbf{A}(s_j) = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} s_{ij}(\mathbf{R}_i - \mathbf{R}_j)$, for each $s_j \in \mathbf{S}$.
4. Initialize $\mathbf{B}_1 = \mathbf{I}$ and $\mathbf{A}_1(s_j) = \mathbf{A}(s_j)$, for all $s_j \in \mathbf{S}$, $j = 1, 2, \ldots, 2^{c(c-1)/2}$.

For $i = 1, 2, \ldots, r$

For $j = 1, 2, \ldots, 2^{c(c-1)/2}$

5. Calculate the eigenvector $w_j = \arg\max_{w, w\mathbf{B}_i w=1} w^T \mathbf{A}_i(s_j) w$ associate with the leading eigenvalue λ_j.

End

6. Choose $w_i = \max_{\lambda_j}\{w_j\}$ to be the eigenvector of the largest λ_j.
7. $\mathbf{q}_i = \mathbf{B}_i w_i$

For $j = 1, 2, \ldots, 2^{c(c-1)/2}$

8. $\mathbf{A}_{i+1}(s_j) = (\mathbf{I} - \mathbf{q}_i \mathbf{q}_i^T)\mathbf{A}_i(s_j)(\mathbf{I} - \mathbf{q}_i \mathbf{q}_i^T)$

End

9. $\mathbf{B}_{i+1} = \mathbf{B}_i(\mathbf{I} - \mathbf{q}_i \mathbf{q}_i^T)$
10. $w_i \leftarrow w_i / \|w_i\|$

End

Output: r CSP projection vectors $\{w_1, w_2, \ldots, w_r\}$.

C. NONPARAMETRIC MULTI-CLASS CSP

The new multi-class CSP algorithm in the above section is also based on the Gaussian distribution assumption, and is also suffering performance degradation in non-Gaussian distribution scenarios. The nonparametric multi-class CSP (NMCSP, we use w_{nmcsp} to denote the projection vector) is derived just by replacing the covariance matrices $\mathbf{R}_i$ and $\mathbf{R}_j$ by the nonparametric within class scatter matrices $\mathbf{S}_i$ and $\mathbf{S}_j$, respectively, where these within class scatter matrices are calculated using formula like (11)-(13).

$$w_{\text{nmcsp}} = \arg \max_{w, w^T w=1} w^T \Big(\sum_{i=1}^{C-1} \sum_{j=i+1}^{C} s_{ij}(\mathbf{S}_i - \mathbf{S}_j) \Big) w \tag{29}$$

Similarly, more projection vectors can be obtained by using **Algorithm 1** but with replacement of $\mathbf{R}_i$ and $\mathbf{R}_j$ by $\mathbf{S}_i$ and $\mathbf{S}_j$. Also note that as pointed out in Section III, when $k = S$, $\mathbf{S}_i = \mathbf{R}_i$, thus NMCSP is a generalization of MCSP.

V. EXPERIMENTAL RESULTS

In this section, we conduct a series of carefully designed evaluations to demonstrate the effectiveness and efficiency of our presented algorithms. The algorithms used in the comparison experiments including our proposed nonparametric CSP and multi-class nonparametric CSP, the traditional two-class CSP, the baseline multi-class CSP [28] as well as the recently proposed CSP extensions such as ACCSP in [31], RCSP in [23] *et al.*. In all the validation experiments, CSP algorithms are used to extract the discriminant feature from raw EEG first, then we resort the SVM algorithm (LibSVM software) [9] to implement the classification task, the classification accuracy are presented. One thing that should be noted is that in all of experiments, we adopt the default parameter settings, that is the value of svm_type is C-SVC, the value of the kernel_type is linear and the gamma value is 0.5 etc..

In the vigilance estimation experiment, we firstly give a detailed description of our designed EEG signal acquisition environment and acquisition procedures. Secondly, we explain our method used to extract the EEG features for vigilance detection and motor imagery recognition as well as some EEG signal preprocessing work. Finally, we evaluate the performance of our proposed algorithms and make comparisons with the baseline as well as the state-of-the-art algorithms.

A. EEG DATA COLLECTION AND PRE-PROCESSING

We build a driving simulation system to record the EEG signals of a subject when he (she) is driving a car through the driving simulation system. 4 subjects aged 18-28 participate in our experiments. Those subjects were required to wear a EEG recording cap with 64 electrodes connected to the amplifier of the NeuroScan system and sit on chair when driving a car in a simulation driving environments. They had shown a tendency to fall asleep during the driving simulation. The electrodes are arranged based on the international 10/20 standard, and the EEG signals were recorded in a computer and the sample rate of brain potential is 100Hz per second. EEG data were acquired through 64 channels including 62 EEG channels and 2 EOG channels. The room used in the experiments was designed to be a dark, quiet, isolated room, the temperature was set to 24 degrees and the humidity was between 40% and 60%. The subjects manipulated the steering wheel and the simulated driving scenes were displayed before them with a LCD screen. A DV camera was also placed in

TABLE 1. Binary classification accuracy comparisons between the classical two-class CSP and NCSP for all subjecsts (accuracy in %).

	Algorithms	$r=5$	$r=8$	$r=10$	$r=15$	$r=20$	$r=30$	$r=40$	$r=50$
Subject 1	**CSP**	94.21	96.55	96.79	97.78	98.50	**99.55**	**99.70**	99.09
	NCSP	**94.51**	**96.95**	**97.04**	**98.17**	**98.56**	99.39	99.60	**99.83**
Subject 2	**CSP**	**84.01**	85.23	86.45	88.02	90.08	92.45	94.81	95.33
	NCSP	80.87	**85.54**	**87.75**	**89.95**	**92.35**	**94.32**	**95.99**	**96.21**
Subject 3	**CSP**	**81.94**	82.55	82.52	84.00	86.15	89.64	92.10	93.84
	NCSP	79.82	**84.47**	**85.95**	**88.56**	**90.24**	**93.48**	**94.49**	**95.44**
Subject 4	**CSP**	85.82	85.88	86.54	88.98	89.54	91.56	94.18	95.23
	NCSP	**86.15**	**87.65**	**87.97**	**90.65**	**92.94**	**93.48**	**95.60**	**97.51**
Average±Std	**CSP**	**86.49±4.66**	88.00±5.26	88.08±5.29	89.70±5.03	91.07±4.55	93.30±3.75	95.20±2.79	95.87±1.95
	NCSP	85.34±5.81	**88.65±4.93**	**89.60±4.36**	**91.83±3.74**	**93.52±3.08**	**93.74±3.42**	**96.42±1.92**	**97.25±1.66**

TABLE 2. Statistics of EEG data sets for motor imageries 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 electrodes	60			118					22
(Re)sampled frequency	250 Hz			100 Hz					250 Hz
MI classified	Left vs. right hand			Right hand vs. foot					Left vs. right hand

front of the subjects to record the facial expression of the subjects for state labeling later. The whole experiment lasted about for one hour and data labeling was completed by hand according to the monitoring video [44]–[47].

In the vigilance estimation experiments, EEG data were divided into 5s epochs with 2.5s window overlap. Half of these epochs are selected to form the training set and the other half would be the test set. CSP algorithms are used to extract the discriminant features from the raw EEG data, and SVM algorithm (libsvm) with default parameter setting to implement the vigilance estimation task. Three vigilance states, e.g., alert, drowsy and sleeping will be predicted.

We make the performance comparisons between the traditional CSP and our proposed nonparametric CSP first, then compare with Zheng's algorithm [28], which inspired us to propose the new multi-class CSP. Finally, we compare our algorithms with the recently proposed new CSP extensions such as ACCSP [31], RCSP [23], CSP-L1 [34] as well as the classical MCSP method [27]. In this section, we named Zheng's CSP algorithm proposed in [28] as ZMCSP for simplicity.

To further test the feature extraction ability of our proposed algorithms to the EEG data, we design a motion imagination recognition experiment, which test data come from the BCI competitions. CSP algorithms are used to extract the feature vector first, then SVM is applied to implement the classification task and the average accuracy and result discussion are presented.

B. BINARY CLASSIFICATION EXPERIMENTS WITH BASELINE ALGORITHM OVER VIGILANCE DETECTION DATA

Let $w_1, w_2, \ldots, w_r$ be the CSP (or NCSP) projection vectors, each EEG epoch $\mathbf{x}_i^c$ is projected onto these directions by $y_j = w_j^T \mathbf{x}_i^c, j = 1, \ldots, r$ to get r feature vectors, and the linear support vector machine (SVM) [9] is then used to classify the three vigilance states on these feature vectors. After a series of pre-processing operations including EOG and EMG removal, noisy filtering, manual labeling, we extract 42 epochs from subject-1 experimental data, 45 epochs from subject-2 experimental data, 39 epochs from subject-3 experimental data and 45 epochs from subject-4 experimental data for our comparison work.

Validation tests are run 10 times and the average classification accuracy rates with different number r CSP projection vectors is given in the Table 1. From the comparisons results we can see that two methods have almost same results, but our proposed NCSP performs slightly better. That's because we adopt the nonparametric strategy to construct the covariance matrices, which make our CSP algorithm more robust to the noisy that introduced by the raw data.

C. BINARY CLASSIFICATION EXPERIMENTS WITH BASELINE ALGORITHM OVER MOTOR IMAGERY DATA SETS

To validate the performance of our CSP on other applications based on EEG signal, we conduct classification experiments on three data sets of BCI competitions, in which 17 subjects were involved. Each subject sat in a relaxing chair with arm rests and was required to do the task of motor imageries (MI) of left hand, right hand, foot or tongue movements according to a cue, during which the EEG signals were recorded [34]. The statistics of dataset we used in this experiment are given in the following Table 2.

Through the comparison results, we can learn that our proposed nonparametric CSP algorithm still get the better results than the traditional CSP, which demonstrate that in some cases, such as when the data set is not completely

TABLE 3. Binary classification accuracy on the motor imageries data from the BCI competetion (accuracy in %).

Data Set	BCI competition III										BCI competition IV					
	Data set IIIa				Data set IVa						Data set IIa					
Subject	s1	s2	s3	Mean	aa	al	av	aw	ay	Mean	A01E	A02E	A03E	A04E	A05E	Mean
CSP	95.56	66.67	93.33	85.19	67.86	98.12	71.43	85.71	75.40	79.72	90.27	53.47	91.67	71.53	61.81	73.75
RCSP	96.67	66.12	**95.00**	85.93	73.88	100	69.23	88.84	81.28	82.65	92.31	53.85	95.80	70.63	65.74	75.67
CSP-L1	93.33	**66.67**	**95.00**	85.00	70.54	94.64	67.35	88.84	83.73	81.02	91.67	58.33	**97.92**	71.53	69.44	77.78
wlCSP-L1	**96.67**	**66.67**	**95.00**	**86.11**	**75.89**	**98.21**	**76.02**	**89.29**	85.71	**85.02**	91.67	59.03	**97.92**	**74.31**	**72.22**	**79.03**
NCSP	**96.67**	**66.67**	**95.00**	**86.11**	75.22	**98.21**	73.52	89.11	**86.45**	84.52	**92.89**	**60.40**	96.52	73.39	69.88	78.62

TABLE 4. Classification rates comparisons between our NMCSP and ZMCSP with different r (accuracy in %).

r	subject1		subject2		subject3		subject4		mean±std	
	NMCSP	ZMCSP	NMCSP	ZMCSP	NMCSP	ZMCSP	NMCSP	ZMCSP	NMCSP	ZMCSP
3	71.95	**81.02**	74.67	**86.95**	**70.78**	60.21	62.60	**75.22**	70.00±4.50	**75.85±9.94**
5	79.69	**86.95**	81.00	**85.12**	**77.41**	65.38	88.01	**90.00**	81.53±3.96	**81.86±9.67**
8	**82.75**	81.83	**90.00**	84.71	**77.41**	72.54	**91.06**	90.67	**85.30±5.57**	82.44±6.54
10	**91.55**	81.43	**93.33**	83.70	**78.01**	67.22	**92.28**	90.33	**88.79± 6.56**	80.67±8.43
15	**93.31**	82.43	**89.95**	84.92	**82.33**	59.98	**92.48**	89.33	**89.52±4.33**	79.16±11.35
20	**91.55**	81.62	**93.33**	85.12	**82.33**	58.80	**92.48**	88.00	**89.92±4.43**	78.39±11.53
30	**92.49**	81.62	**93.00**	87.36	**82.03**	60.33	**93.29**	83.33	**90.20±4.73**	78.16±10.50
40	**91.78**	81.42	**92.33**	83.90	**81.53**	60.56	**94.51**	78.33	**90.04±5.02**	76.05±9.16
50	**92.96**	81.63	**91.33**	83.50	**80.82**	58.57	**94.92**	78.00	**90.00±5.45**	75.43±9.93
55	**93.54**	81.52	**93.67**	83.70	**81.63**	59.74	**93.90**	78.67	**90.66±5.23**	75.91±9.50

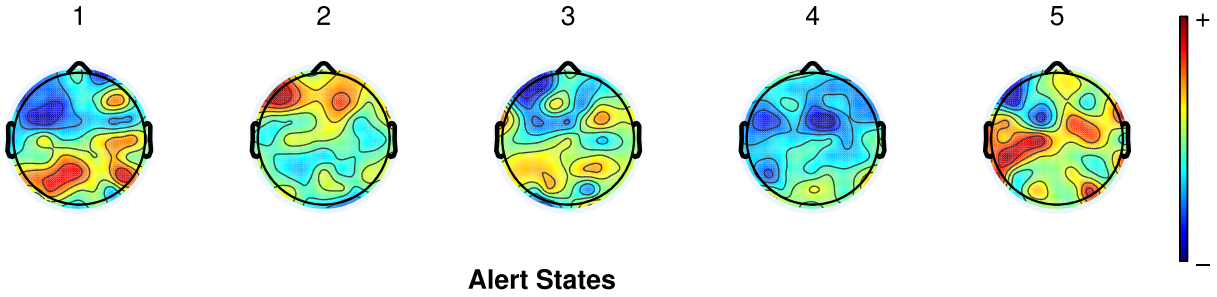


FIGURE 1. CSP patterns benefited from using ZMCSP [28] for subject 4 whose brain was in the alert states.

gaussian, the non-parametric CSP algorithm can extract more discriminative EEG characteristics.

D. MULTI-CLASS CLASSIFICATION EXPERIMENTS WITH BASELINE ALGORITHM OVER VIGILANCE DETECTION DATA

Table 4 shows the classification rates of our new nonparametric multi-class CSP and ZMCSP in [28] for three vigilance states (alert, drowsy, sleep) with different CSP projection vector number r for all subjects. It shows that ZMCSP have a better performance than our NMCSP with very few CSP vectors, for example when $r = 3, 5$. However, when the number of CSP vectors increase, our proposed NMCSP showed more significant advantage than the ZMCSP in the experiments of all four subjects. Especially in the experiment of subject 3, our NMCSP achieves much better results than ZMCSP in [28] in all of dimensionality trial, this demonstrates the better discriminant ability of criteria (23) than that of [28]. In addition, ZMCSP algorithm also shows a strange phenomenon that the performance of ZCSP is not closely related to dimensionality value, and even the performance of algorithm would decline with the increase of dimension r value, we guess that was because the increase of dimensionality introduced more noise thus leads to the performance decrease.

In Figure 1, Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6, we demonstrate the brain topographic maps of subject 4 whose brain was in alertness, drowsy and asleep states respectively, of which Figure 1, Figure 2 and Figure 3 are computed by the ZMCSP algorithm [28], while Figure 4, Figure 5 and Figure 6 are computed based on our proposed algorithm. It can be seen from the comparison results that the CSP patterns extracted from the raw EEG data for vigilance estimation by our proposed algorithm is more separable than ZMCSP [28].

To further test the performance and robustness of our proposed NMCSP, In Figure 7(a), we plot the classifcation accuracy result for three-class vigilance classification problem with respect to different number of training data on all four subject data, and in Figure 7(b) we demonstrate the classifcation accuracy result for three-class vigilance classification problem with respect to different value of dimensionality

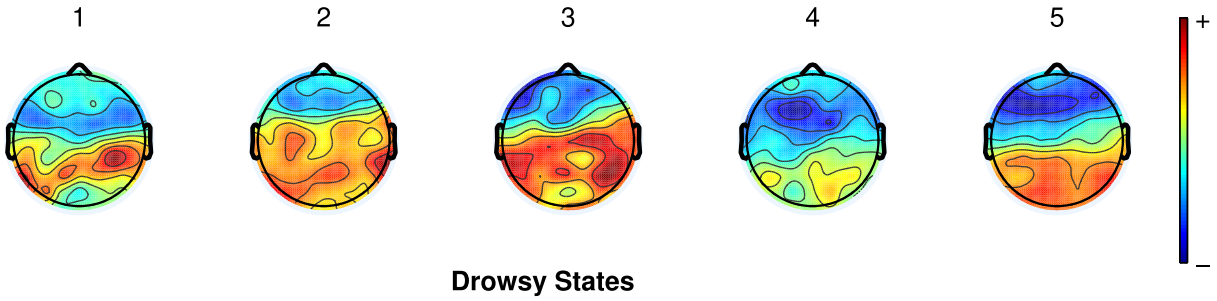


FIGURE 2. CSP patterns benefited from using ZMCSP [28] for subject 4 whose brain was in the drowsy states.

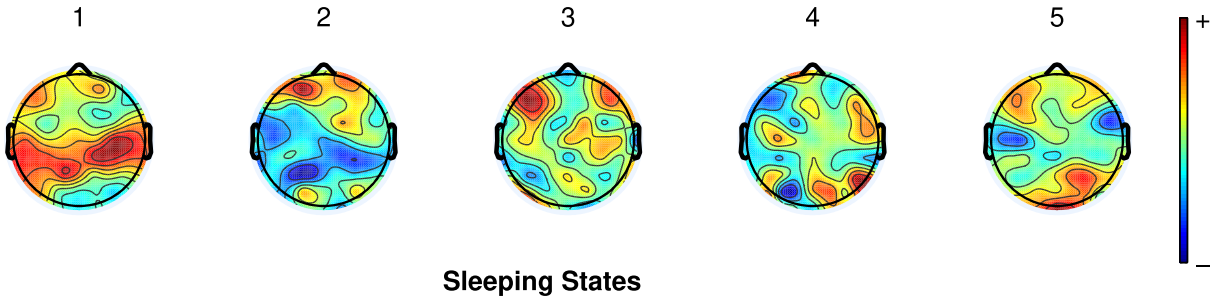


FIGURE 3. CSP patterns benefited from using ZMCSP [28] for subject 4 whose brain was in the sleep states.

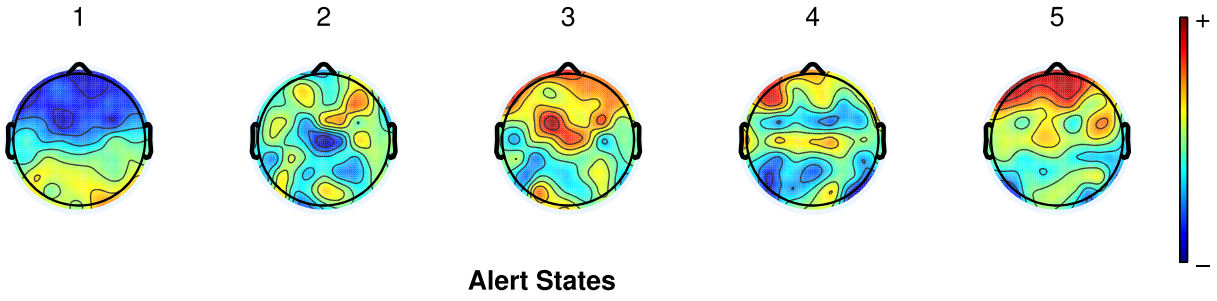


FIGURE 4. CSP patterns benefited from using Our proposed Nonparametric multi-class CSP for subject 4 whose brain was in the alert states.

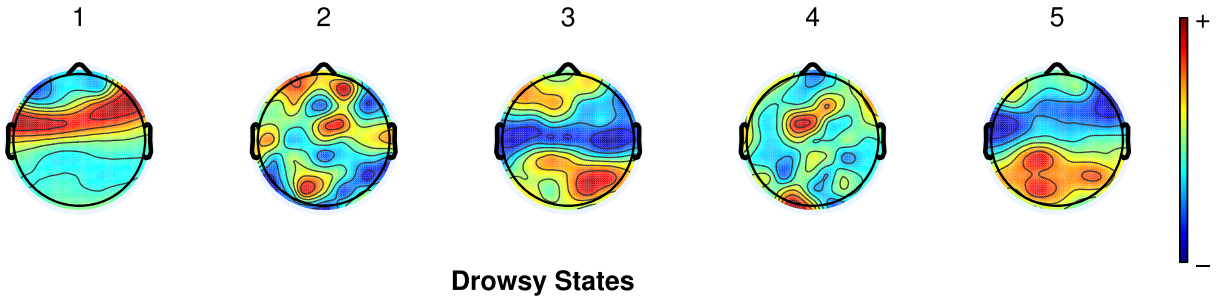


FIGURE 5. CSP patterns benefited from using Our proposed Nonparametric multi-class CSP for subject 4 whose brain was in the drowsy states.

(with fixed number ($r = 7$)) over all four subjects data. All of data of each subject were divided into two equal parts, one parts for training and the rest for test. Experimental results show good accuracy and relatively robust results for different subjects.

E. COMPARIONS WITH THE STATE-OF-THE-ART ALGORITHMS

To further test our new CSP algorithms, we compare its performance with the recently proposed CSP algorithms such as ACCSP in [31], RCSP in [23], CSP-L1 [34] as well as the

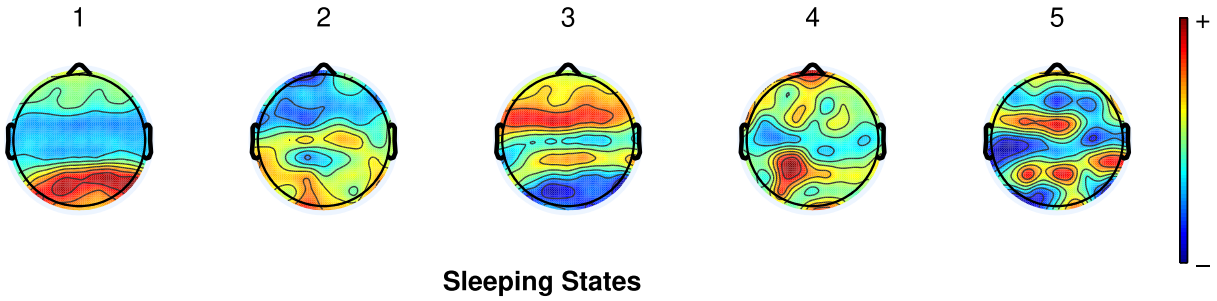


FIGURE 6. CSP patterns benefited from using Our proposed Nonparametric multi-class CSP for subject 4 whose brain was in the sleep states.

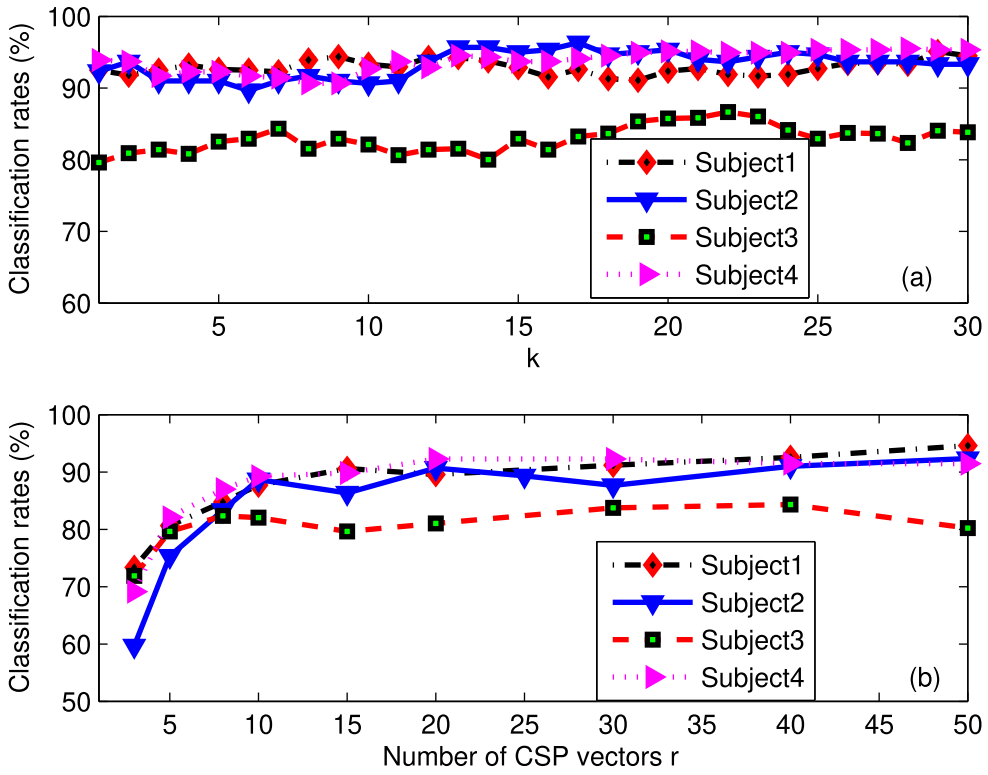


FIGURE 7. Classification rates of NMCSP: (a) with different k; (b) with different r.

TABLE 5. Classification rates of our NMCSP with different r (Accuracy in %).

r	NMCSP	ACCSP [31]	RCSP [23]	CSP-L1 [34]	MCSP [27]
3	72.38	74.50	70.72	**78.49**	69.90
5	82.25	80.95	79.31	**82.36**	75.68
8	**84.92**	80.80	80.55	82.53	78.88
10	**89.01**	81.43	83.91	85.71	82.70
15	**89.65**	83.57	84.64	86.96	83.67
20	**89.92**	85.85	86.82	87.38	84.40
30	**90.20**	85.50	85.63	88.36	85.01
40	**90.03**	83.65	87.76	87.85	84.34
50	**90.31**	85.61	89.19	89.55	85.13
55	91.69	85.50	89.64	**92.75**	84.62

classical MCSP method [27]. In the Table 5 we demonstrate the performance comparison results, the average classification accuracy over the four subjects is given. As can be seen from the comparison results in the Table 5, the NMCSP algorithm we proposed still has obvious advantages comparing with other newly proposed MCSP algorithms, thus proves the correctness of our proposed theory.

F. MULTI-CLASS CSP ALGORITHM ON THE MOTOR IMAGERIES EXPERIMENT

We then implement the proposed nonparametric multi-class CSP algorithm on the benchmark dataset of BCI competition IV namely data IIa. It is a continuous Multi-class Motor Imagery EEG data of 9 subjects. It consisted of four different motor imagery class, imagination of the left hand, right hand, both feet and tongue. The statistics of the dataset used in this experiment is given in the Table 2. In this verification experiment, we use as same setup scheme and the preprocessing method as adopted in [24]. The comparison results are presented in the Table 6. It can be seen from the experimental comparison results that all the algorithms

TABLE 6. Classification performance comparisons on the motor imagery dataset (accuracy in %).

Subject	Multiclass CSP [27]	ComplexCSP [48]	ACCSP	RCSP	CSP-L1	ZMCSP	NMCSP
A01E	48.1	61.5	60.8	58.6	54.2	60.1	62.0
A02E	27.3	32.1	36.1	39.7	43.6	65.3	48.7
A03E	70.6	68.6	66.1	65.7	66.8	72.5	76.5
A04E	21.4	27.1	23.5	29.3	31.4	38.3	40.7
A05E	22.7	34.3	29.9	32.6	32.9	33.9	35.2
A06E	32.4	35.3	36.8	35.8	40.2	37.1	38.8
A07E	52.3	48.0	53.2	55.6	56.7	60.3	61.0
A08E	65.8	65.6	62.3	66.7	67.9	60.6	63.4
A09E	34.2	41.8	46.2	52.6	59.5	58.6	65.5
Averag±Std	41.64±18.34	46.01±15.65	46.09±14.47	48.51±13.57	50.36±13.04	54.08±13.11	54.64±13.44

TABLE 7. The p-values of NMCSP and state-of-art algorithms in terms of classification accuracy on the validation datasets.

Dataset	Vigilance Data	Motor Imagery
NMCSP VS. ZMCSP	1.8e-3	2.4e-2
NMCSP VS. ACCSP	7.2e-2	9.4e-2
NMCSP VS. RCSP	1.9e-2	1.4e-2
NMCSP VS. CSPL1	1.6e-1	2.2e-2
NMCSP VS. MCSP	2.4e-2	3.2e-2

TABLE 8. The p-values of NCSP and state-of-art algorithms in terms of classification accuracy on the validation datasets.

Dataset	Vigilance Data	Dataset IIa	Dataset IIIa	Dataset IVa
NCSP VS. CSP	1.4e-3	1.7e-2	2.5e-3	3.7e-3
NCSP VS. RCSP	9.4-e2	1.7e-2	1.6e-3	3.8e-3
NCSP VS. CSPL1	4.8-e3	5.7e-3	6.5e-2	1.9e-3
NCSP VS. WICSP	5.7e-2	5.9e-2	1.7e-3	3.2e-3

based on BCI competition data set do not show good classification performance, we guess that because too much noisy introduced during the data collection procedure as well as the subjects' performance. The table shows that based on the same classification strategy, different data sets exhibit different classification accuracy performance. In addition to the algorithm itself, the performance of the subjects in the process of data collection is also an important factor affecting the recognition performance. Our proposed NMCSP get the first place by a narrow margin in this experiment.

G. STATISTICAL TEST

Tables 7 and 8 present the p-values of the classification accuracy result to assess the statistical significance of performance difference between our algorithms and the compared methods. A p-value of smaller than 0.05 is generally considered as indication of performance difference. Tables 7 and 8, we see that most of the performance differences are statistically significant.

VI. CONCLUSION

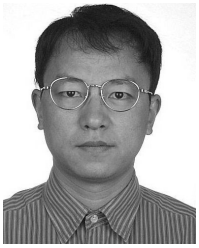
In this work, we first propose a nonparametric CSP for two-class problem, where two within class scatter matrices are calculated from the k nearest neighbors of each sample, and are then used to replace the covariance matrices in the original CSP. We also propose a novel multi-class CSP and its nonparametric version. Our multi-class CSP has an explanation of minimizing Bayes error, and is formulated as optimization problem of an objective function in terms of sample covariance matrices. We develop an algorithm base on matrix deflation to calculate multiple multi-class CSP projection vectors. This algorithm is a full search algorithm which is valid only for small class number, we will investigate more efficient algorithms that scale well for large class number in the future.

REFERENCES

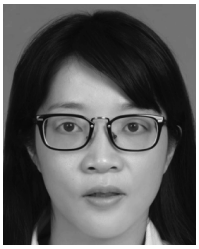
[1] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, no. 2, pp. 326–341, 2005.
[2] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.
[3] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 157–164.
[4] N. Bigdely-Shamlo, G. Ibagon, C. Kothe, and T. Mullen, "Finding the optimal cross-subject EEG data alignment method for analysis and BCI," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 1110–1115.
[5] M. Bensch, A. A. Karim, J. Mellinger, T. Hinterberger, M. Tangermann, M. Bogdan, W. Rosenstiel, and N. Birbaumer, "Nessi: An EEG-controlled Web browser for severely paralyzed patients," *Comput. Intell. Neurosci.*, vol. 2007, Jun. 2007, Art. no. 71863.
[6] S. R. Soekadar, M. Witkowski, N. Vitiello, and N. Birbaumer, "An EEG/EOG-based hybrid brain-neural computer interaction (BNCI) system to control an exoskeleton for the paralyzed hand," *Biomed. Eng./Biomedizinische Technik*, vol. 60, no. 3, pp. 199–205, 2015.
[7] R. A. Grier, J. S. Warm, W. N. Dember, G. Matthews, T. L. Galinsky, J. L. Szalma, and R. Parasuraman, "The vigilance decrement reflects limitations in effortful attention, not mindlessness," *Hum. Factors*, vol. 45, no. 3, pp. 349–359, 2003.
[8] T. Abe, T. Nonomura, Y. Komada, S. Asaoka, T. Sasai, A. Ueno, and Y. Inoue, "Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests," *Int. J. Psychophysiol.*, vol. 82, no. 3, pp. 269–274, 2011.
[9] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[10] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imag.*, vol. 8, no. 5, pp. 357–377, 2002.
[11] K. L. Hoffman, K. M. Gothard, M. C. Schmid, and N. K. Logothetis, "Facial-expression and gaze-selective responses in the monkey amygdala," *Current Biol.*, vol. 17, no. 9, pp. 766–772, 2007.
[12] L. Buck, "Reaction time as a measure of perceptual vigilance," *Psychol. Bull.*, vol. 65, no. 5, p. 291, 1966.
[13] Y. Stern, R. Mayeux, and L. Côté, "Reaction time and vigilance in Parkinson's disease: Possible role of altered norepinephrine metabolism," *Arch. Neurol.*, vol. 41, no. 10, pp. 1086–1089, 1984.
[14] E. C.-P. Chua, W.-Q. Tan, S.-C. Yeo, P. Lau, I. Lee, I. H. Mien, K. Puvanendran, and J. J. Gooley, "Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation," *Sleep*, vol. 35, no. 3, pp. 325–334, 2012.
[15] A. C. Parrott and G. Winder, "Nicotine chewing gum (2 mg, 4 mg) and cigarette smoking: Comparative effects upon vigilance and heart rate," *Psychopharmacology*, vol. 97, no. 2, pp. 257–261, 1989.

[16] J. Santamaria and K. Chiappa, "The EEG of drowsiness in normal adults," *J. Clin. Neurophysiol.*, vol. 4, no. 4, pp. 327–382, 1987.
[17] S. Makeig and T.-P. Jung, "Changes in alertness are a principal component of variance in the EEG spectrum," *Int. J. Rapid Commun. Res. Neurosci.*, vol. 7, no. 1, pp. 213–216, 1995.
[18] I. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer, 2011.
[19] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[20] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 46. Hoboken, NJ, USA: Wiley, 2004.
[21] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.
[22] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov. 2009.
[23] H. Lu, H.-W. 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.
[24] H. Meisheri, N. Ramrao, and S. Mitra, "Multiclass common spatial pattern for EEG based brain computer interface with adaptive learning classifier," 2018, *arXiv:1802.09046*. [Online]. Available: https://arxiv.org/abs/1802.09046
[25] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[26] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
[27] M. Grosse-Wentrup and M. Buss, " Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.
[28] W. Zheng and Z. Lin, "Optimizing multi-class spatio-spectral filters via Bayes error estimation for EEG classification," in *Proc. Adv. Neural Inf. Processing Syst.*, 2009, pp. 2268–2276.
[29] O. Falzon, K. P. Camilleri, and J. Muscat, "Complex-valued spatial filters for task discrimination," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug./Sep. 2010, pp. 4707–4710.
[30] O. Falzon, K. P. Camilleri, and J. Muscat, "The analytic common spatial patterns method for EEG-based BCI data," *J. Neural Eng.*, vol. 9, no. 4, 2012, Art. no. 045009.
[31] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 1–10, Jan. 2014.
[32] M. Kawanabe, W. Samek, K.-R. Müller, and C. Vidaurre, "Robust common spatial filters with a maxmin approach," *Neural Comput.*, vol. 26, no. 2, pp. 349–376, 2014.
[33] Z. Li, D. Lin, and X. Tang, "Nonparametric discriminant analysis for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell,.*, vol. 31, no. 4, pp. 755–761, Sep. 2009.
[34] H. Wang and X. Li, "Regularized filters for L1-norm-based common spatial patterns," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 2, pp. 201–211, Feb. 2016.
[35] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
[36] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 9, no. 3, pp. 283–288, Sep. 2001.
[37] M. S. Weiss, "Non-Gaussian properties of the EEG during sleep," *Electroencephalogr. Clin. Neurophysiol.*, vol. 34, no. 2, pp. 200–202, 1973.
[38] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Amsterdam, The Netherlands: Elsevier, 2013.
[39] S. Lu, Z. Lu, and Y.-D. Zhang, "Pathological brain detection based on AlexNet and transfer learning," *J. Comput. Sci.*, vol. 30, pp. 41–47, Jan. 2019.
[40] J. Chu and J. Chueh, "Error probability in decision functions for character recognition," *J. Appl. Comput. Mech.*, vol. 14, no. 2, pp. 273–280, 1967.
[41] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.
[42] L. W. Mackey, "Deflation methods for sparse PCA," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1017–1024.
[43] S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, Feb. 2001.
[44] L.-C. Shi, H. Yu, and B.-L. Lu, "Semi-supervised clustering for vigilance analysis based on EEG," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1518–1523.
[45] H.-J. Liu, Q.-S. Ren, and H.-T. Lu, "Estimate vigilance in driving simulation based on detection of light drowsiness," in *Proc. Int. Conf. Bioinformatics*, 2010, pp. 131–134.
[46] H. Yu, H. Lu, T. Ouyang, H. Liu, and B.-L. Lu, "Vigilance detection based on sparse representation of EEG," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug./Sep. 2010, pp. 2439–2442.
[47] J.-X. Ma, L.-C. Shi, and B.-L. Lu, "Vigilance estimation by using electrooculographic features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug./Sep. 2010, pp. 6591–6594.
[48] H. Zhang, H. Yang, and C. Guan, "Bayesian learning for spatial filtering in an EEG-based brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1049–1060, Jul. 2013.

HONGBIN YU received the M.S. degree in pattern recognition and intelligent system from the University of Science and Technology of China, in 2009, and the Ph.D. degree in computer science and technology from Shanghai Jiao Tong University, Shanghai, China, in 2016. He is currently a Lecturer with the School of Digital Media, Jiangnan University. His research interests include brain–computer interface and computer vision.



HONGTAO LU received the Ph.D. degree in electronic engineering from Southeast University, Nanjing, in 1997. After graduation, he became a Postdoctoral Fellow with the Department of Computer Science, Fudan University, Shanghai, China, where he spent two years. In 1999, he joined the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, where he is currently a Professor. His research interests include machine learning, computer vision, and pattern recognition. He has published more than 60 papers in international journals, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS, and in international conferences. His papers got more than 800 citations by other researchers.



SHUIHUA WANG received the B.S. degree from Southeast University, in 2008, the M.S. degree from The City University of New York, in 2012, and the Ph.D. degree from Nanjing University, in 2016. She is currently an Assistant Professor with Loughborough University, Loughborough, U.K.

KAIJIAN XIA received the M.S. degree in computer science and technology from Jiangnan University, in 2009. He is currently pursuing the the Ph.D. degree from the China University of Mining Technology. From 2009 to 2010, he was a Lecturer with the School of Computer Science and Engineering, Changshu Institute of Technology. Since 2010, he has been with the Changshu NO.1 People's Hospital. He has published nearly 20 papers in international/national journals. His research interests include biological image processing and computational medicine.

YIZHANG JIANG received the Ph.D. degree from Jiangnan University, Wuxi, Jiangsu, China, in 2016, where he is currently an Associate Professor with the School of Digital Media. He has published more than 30 papers in international journals, including IEEE TFS, IEEE TNNLS, IEEE T. Cyber., PR, KBS, and InS. His research interests include pattern recognition, intelligent computation, and their applications.

PENGJIANG QIAN received the Ph.D. degree from Jiangnan University, Wuxi, Jiangsu, China, in March, 2011, where he is currently a Full Professor with the School of Digital Media. He has authored or coauthored nearly 60 papers published in international/national journals and conferences, e.g., IEEE TNNLS, IEEE TSMC-B, IEEE T. Cyber., IEEE TFS, PR, InS, and KBS. His research interests include data mining, pattern recognition, bioinformatics, and their applications, such as analysis and processing for medical imaging, intelligent traffic dispatching, and advanced business intelligence in logistics.