

Research Article

EEG Signal Classification Using Manifold Learning and Matrix-Variate Gaussian Model

Lei Zhu , Qifeng Hu, Junting Yang, Jianhai Zhang, Ping Xu, and Nanjiao Ying

School of Automation, Hangzhou Dianzi University, Hangzhou 310000, China

Correspondence should be addressed to Lei Zhu; zhulei@hdu.edu.cn

Received 2 November 2020; Revised 1 March 2021; Accepted 17 March 2021; Published 25 March 2021

Academic Editor: Friedhelm Schwenker

Copyright © 2021 Lei Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In brain-computer interface (BCI), feature extraction is the key to the accuracy of recognition. There is important local structural information in the EEG signals, which is effective for classification; and this locality of EEG features not only exists in the spatial channel position but also exists in the frequency domain. In order to retain sufficient spatial structure and frequency information, we use one-versus-rest filter bank common spatial patterns (OVR-FBCSP) to preprocess the data and extract preliminary features. On this basis, we conduct research and discussion on feature extraction methods. One-dimensional feature extraction methods like linear discriminant analysis (LDA) may destroy this kind of structural information. Traditional manifold learning methods or two-dimensional feature extraction methods cannot extract both types of information at the same time. We introduced the bilinear structure and matrix-variate Gaussian model into two-dimensional discriminant locality preserving projection (2DDLPP) algorithm and decompose EEG signals into spatial and spectral parts. Afterwards, the most discriminative features were selected through a weight calculation method. We tested the method on BCI competition data sets 2a, data sets IIIa, and data sets collected by our laboratory, and the results were expressed in terms of recognition accuracy. The cross-validation results were 75.69%, 70.46%, and 54.49%, respectively. The average recognition accuracy of new method is improved by 7.14%, 7.38%, 4.86%, and 3.8% compared to those of LDA, two-dimensional linear discriminant analysis (2DLDA), discriminant locality property projections (DLPP), and 2DDLPP, respectively. Therefore, we consider that the proposed method is effective for EEG classification.

1. Introduction

Brain-computer interface (BCI) is a kind of real-time communication system connecting the brain and external devices. BCIs based on electroencephalogram (EEG) can convert the information sent by the brain into commands to drive the external devices, so as to realize the communication between people and the outside world [1]. There are several control signal types in BCI and, among them, motor imagery (MI) is one of the most studied applications [2]. MI is an independent BCI method that uses motor cortex as a signal source. The user imagines moving his/her limbs without any actual muscular movement. Studies on EEG signal indicate that when people perform motor imaging tasks, this will cause an event-related desynchronization (ERD) and event-related synchronization (ERS) of

oscillations in alpha band (8–13 Hz) and beta band (14–30 Hz) [3]. Due to these characteristics, researchers can process and analyze EEG signals in relevant frequency bands for the classification of motor imaging tasks.

In MI-based BCI system, a well-known problem is how to handle very large amounts of features extracted from multichannel EEG signals. Feature extraction is a commonly used approach for solving this problem. To solve the computational complexity and data storage problem caused by the high dimension of signals, many dimensionality reduction methods have been used in traditional BCI technology. Kumar et al. [4] used independent components analysis (ICA) to remove artifacts in EEG signals and used principal component analysis (PCA) for reducing high-dimensional data. Although PCA seeks to learn a projection that can preserve the main energy of data, it does not contain

discriminability. Linear discriminant analysis (LDA) uses the label information to enlarge the between-class distance and reduce the within-class distance [5]. Although LDA uses the label information in MI, it ignores the important local information in the EEG signal. Motor imaging EEG signals are collected through electrodes spread over the cerebral cortex. When a certain imaging task is performed, ERD/ERS will occur in specific local areas of the brain [6]. There are mutual influences and connections between the signals collected by adjacent electrodes. These connections exist not only in the spatial position but also in the frequency domain. As described above, the main focus is on alpha band (8–13 Hz) and beta band (14–30 Hz). The LDA method cannot effectively extract these local features. Since the manifold learning method was proposed [7, 8], some studies assumed that EEG signals are more likely to exist in low-dimensional nonlinear manifold subspace. Since then, manifold learning methods have been applied on epileptic EEG signals [9, 10], EEG-based depth of anesthesia assessment [11], emotional state classification [12], tracking dynamic EEG brain connectivity [13], and so forth. At the same time, manifold learning was applied in MI-based BCI system [14–16].

The goal of methods based on manifold learning is to keep locality-geometric structure of data in a neighboring area and successfully find the inherent features existing in nonlinear manifold [17]. Locality property projections (LPP) [18], neighborhood preserving embedding (NPE) [19], and locally linear embedding (LLE) [8] are the popular methods of manifold learning, which preserve the locality property. But they suffer from a limitation that they deemphasize discriminant information which is important in recognition problem. Discriminant locality property projection (DLPP) [20] aims to find the subspace that best discriminates different classes by minimizing the within-class distance, while maximizing the between-class distance. However, they are all vector-based methods. EEG signals, as multichannel data, contain rich spatial information. The spatial information here means that the EEG signal is in the form of a matrix, and the matrix contains two-dimensional information closely related to the spatial structure. Significant spatial channel information will be lost by using vector-based methods [21]. According to these factors, the two-dimensional dimensionality reduction method is considered. Two-dimensional discriminant locality preserving projection (2DDLPP) [22] is the 2D expansion of DLPP. 2DDLPP can extract two-dimensional information and preserve geometric structures of original data. But, for EEG signals, both spatial information and spectral information are important information that is effective for classification. 2DDLPP only calculates a single-sided projection and cannot use the spatial and spectral characteristics of the EEG signal at the same time.

In addition to the manifold learning method, sparse learning has been increasingly applied to feature extraction. For example, sparse linear discriminant analysis (SLDA) [23] and sparse two-dimensional discriminant locality preserving projection (S2DDLPP) [17] are proposed to learn a sparse discriminant subspace for feature extraction.

Researches have shown that the method based on sparse can effectively reduce the computational complexity and has good robustness to noise [24–26]. But the above methods are somewhat sensitive to the selection of the number of dimensions, since the discriminability of each projection direction is fixed [27].

To solve the above issues and combine the important local information in the EEG signal with the 2D matrix processing method, we propose an extension of 2DDLPP based on the Gaussian variable model. The main idea of matrix-variable Gaussian model [28] implies a separable structure for the covariance matrix of the vectorized data and it shows that the covariance between any two spatial-spectral features can be decomposed into two terms. On this basis, we calculate the eigenvalues and eigenvectors of spatial term and spectral term, respectively. After that, the two sets of eigenvalues are multiplied in pairs and sorted, and those features with a large weight are selected. This allows the spectral and spatial characteristics to be analyzed at the same time, thereby ensuring that the extracted features have the best discriminativeness.

This paper proposes a bilinear two-dimensional discriminant locality preserving projection (B2DDLPP) algorithm that based on a matrix-variable Gaussian model. Compared with 2DDLPP, B2DDLPP uses a bilinear structure to fully extract the connections between the channels of the EEG signal. Both spatial information and frequency information are considered at the same time, so it is more suitable for the task of feature extraction of motor imaging EEG signals.

2. Materials and Methods

2.1. Related Methods. The input data of the feature extraction algorithm introduced in this paper is the feature matrix after OVR-FBCSP, and each sample gets a feature matrix with a size of $N_f \times N_g$, where N_f is the number of frequency bands in FBCSP (9 in this article), $N_g = 2 \times m \times \text{Numclass}$, m is the number of pairs of features, and Numclass is the total number of classes. When using the vector-based feature extraction method, we vectorize the feature matrix of each sample as input. The class of each feature matrix is the same as the corresponding sample; and the OVR-FBCSP algorithm will be introduced in Section 2.3.1.

2.1.1. Discriminant Locality Property Projections (DLPP). DLPP is based on the extension of LPP, and it considers the discrimination information. This makes DLPP have better performance on classification problems compared to LPP.

The objective function of DLPP is as follows:

$$\frac{\sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{y}_i^s - \mathbf{y}_j^s)^2 \mathbf{W}_{ij}^s}{\sum_{i,j=1}^Z (\mathbf{m}_i - \mathbf{m}_j)^2 \mathbf{B}_{ij}} \quad (1)$$

where Z is the number of motor imagination classes, n_s is the number of samples in the s th class, \mathbf{y}_i^s denotes the i th weight vector in the s th class, \mathbf{m}_i and \mathbf{m}_j are separately the mean weight vectors for the i th class and j th class, respectively; that is, $\mathbf{m}_i = (1/n_i) \sum_{k=1}^{n_i} \mathbf{y}_k^i$ and $\mathbf{m}_j = (1/n_j) \sum_{k=1}^{n_j} \mathbf{y}_k^j$, where n_i and

n_j are the numbers of samples in the i th class and j th class, separately. \mathbf{W}_{ij}^s is the weight between \mathbf{y}_i^s and \mathbf{y}_j^s , and \mathbf{B}_{ij} is the weight between \mathbf{m}_i and \mathbf{m}_j . It should be noted that s in \mathbf{y}_i^s and \mathbf{W}_{ij}^s is the upper corner mark, not power operation.

Suppose that \mathbf{a} is a transformation vector; that is, $\mathbf{Y} = \mathbf{a}^T \mathbf{X}$. By simple algebra formulation, the objective function can be turned to

$$\frac{\mathbf{a}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a}}{\mathbf{a}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{a}}, \quad (2)$$

where $L = D - W$, and W can be defined as $\mathbf{W}_{ij}^s = \exp(-\|\mathbf{x}_i^s - \mathbf{x}_j^s\|^2/t)$. D is a diagonal matrix, and its entries are column (or row) sum of W s; $\mathbf{D}_{ii}^s = \sum_j \mathbf{W}_{ij}^s$. $H = E - B$, and B can be defined as $\mathbf{B}_{ij}^s = \exp(-\|\mathbf{f}_i - \mathbf{f}_j\|^2/t)$, where \mathbf{f}_i is the mean value of samples in the i th class; $F = f_1, f_2, \dots, f_s$; E is a diagonal matrix, and its entries are column (or row) sum of B ; $\mathbf{E}_{ii} = \sum_j \mathbf{B}_{ij}$.

DLPP subspace is spanned by a set of vectors \mathbf{a} , satisfying

$$\mathbf{a} = \arg \min \frac{\mathbf{a}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a}}{\mathbf{a}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{a}}. \quad (3)$$

The numerator of objective function reflects within-class distance, while the denominator reflects between-class distance. The vectors \mathbf{a}_i that minimize the objective function are given by minimum eigenvalues solutions to the generalized eigenvalues problem.

Thus, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d]$ are the solutions of equation (3), ordered according to their eigenvalues, $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$; and the embedding is as follows:

$$\mathbf{x}_i \longrightarrow \mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i. \quad (4)$$

2.1.2. Two-Dimensional Discriminant Locality Preserving Projection (2DDLPP). 2DDLPP is the 2D expansion of DLPP. The main advantage of 2DDLPP over DLPP is that it has a more accurate approximation of the original signals, which can avoid the losses of important information for recognition. The objective function of 2DDLPP is

$$J(\mathbf{Y}) = \frac{\sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{Y}_i^s - \mathbf{Y}_j^s)^T (\mathbf{Y}_i^s - \mathbf{Y}_j^s) \mathbf{W}_{ij}^s}{\sum_{i,j=1}^Z (\mathbf{M}_i - \mathbf{M}_j)^T (\mathbf{M}_i - \mathbf{M}_j) \mathbf{B}_{ij}}, \quad (5)$$

where \mathbf{Y}_i^s and \mathbf{Y}_j^s denote the projected feature matrices in class s , corresponding to the original EEG signals. Z is the number of motor imagination classes. \mathbf{W}_{ij}^s and \mathbf{B}_{ij} are the within-class weight matrix and the between-class weight matrix, separately. \mathbf{M}_i and \mathbf{M}_j represent the mean matrices of the projected signals in class i and class j ; that is, $\mathbf{M}_i = (1/n_i) \sum_{k=1}^{n_i} \mathbf{Y}_k^i$ and $\mathbf{M}_j = (1/n_j) \sum_{k=1}^{n_j} \mathbf{Y}_k^j$.

Suppose that \mathbf{X} is an $N_f \times N_g$ feature matrix signal and \mathbf{A} denotes the transformation matrix. The linear transformation is $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$.

The objective function can be reformed as follows:

$$J(\mathbf{A}) = \frac{\mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A}}{\mathbf{A}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{A}} = \frac{\mathbf{A}^T \mathbf{P}_w \mathbf{A}}{\mathbf{A}^T \mathbf{P}_b \mathbf{A}}, \quad (6)$$

where

$$\mathbf{P}_w = \frac{1}{2} \sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{X}_i^s - \mathbf{X}_j^s) (\mathbf{X}_i^s - \mathbf{X}_j^s)^T \mathbf{W}_{ij}^s, \quad (7)$$

$$\mathbf{P}_b = \frac{1}{2} \sum_{i,j=1}^Z (\mathbf{F}_i - \mathbf{F}_j) (\mathbf{F}_i - \mathbf{F}_j)^T \mathbf{B}_{ij},$$

where \mathbf{F}_i is the mean matrix of the i th class; it is defined as $\mathbf{F}_i = (1/n_i) \sum_{k=1}^{n_i} \mathbf{X}_k^i$. B is the weight matrix between any two classes' mean matrix and it is defined as $\mathbf{B}_{ij} = \exp(-\|\mathbf{F}_i - \mathbf{F}_j\|^2/t)$. \mathbf{W}_s is the weight matrix between any two samples in the s th class, and it is defined as

$$\mathbf{W}_{ij}^s = \exp\left(\frac{-\|\mathbf{X}_i^s - \mathbf{X}_j^s\|^2}{t}\right), \quad (8)$$

where t is a parameter that can be set empirically and in this paper it is set to 1.

$L = D - W$ is the Laplacian matrix, where D s is a diagonal matrix, and its entries are column (or row) sum of W s; $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ji}$; and $H = E - B$, where E is a diagonal matrix, and its entries are column (or row) sum of B ; $\mathbf{E}_{ii} = \sum_j \mathbf{B}_{ji}$.

The projection directions can be obtained by minimizing the objective function, satisfying

$$\mathbf{A} = \arg \min_{\mathbf{A}} \frac{\mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A}}{\mathbf{A}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{A}}. \quad (9)$$

The minimization problem can be turned into a generalized eigenvalue problem:

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A} = \lambda \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{A}. \quad (10)$$

Thus, $\mathbf{A} = a_1, a_2, \dots, a_d$ are the solutions of (9), ordered according to their eigenvalues, $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$.

2.2. Matrix-Variate Gaussian Model and Its Combination with Manifold Learning

2.2.1. Matrix-Variate Gaussian Model. The matrix-variate Gaussian model is the normal distribution of the matrix. Let $f(\mathbf{X}|\Omega_i)$ denote the conditional probability of matrix $\mathbf{X} \in \mathbb{R}^{N_f \times N_g}$ under class Ω_i , and the matrix-variate Gaussian model for matrix \mathbf{X} is denoted by

$$\mathbf{X}|\Omega_i \sim \mathcal{M} \mathcal{N}_{N_f \times N_g}(\mathbf{M}_i, \phi_i, \Psi_i), \quad 1 \leq i \leq Z, \quad (11)$$

where Z is the total number of classes, the matrices \mathbf{M}_i denote the mean matrix of the class Ω_i , and ϕ_i and Ψ_i denote covariance matrix of the class Ω_i . In this paper, ϕ_i and Ψ_i denote spectral covariance and spatial covariance, respectively. These matrices are defined as follows:

$$\begin{aligned} \mathbf{M}_i &= E_{\mathbf{X}|\Omega_i}(\mathbf{X}), \\ \phi_i &= \text{tr}^{-1}(\Psi_i) * E_{\mathbf{X}|\Omega_i}((\mathbf{X} - \mathbf{M}_i)(\mathbf{X} - \mathbf{M}_i)^T), \\ \Psi_i &= \text{tr}^{-1}(\phi_i) * E_{\mathbf{X}|\Omega_i}((\mathbf{X} - \mathbf{M}_i)^T(\mathbf{X} - \mathbf{M}_i)). \end{aligned} \quad (12)$$

The conditional probability of matrix X can be determined by \mathbf{M}_i , ϕ_i , and Ψ_i as follows:

$$f(\mathbf{X}|\Omega_i) = \frac{\exp\left\{-\frac{1}{2}\text{tr}\left[\phi_i^{-1}(\mathbf{X} - \mathbf{M}_i)^T \Psi_i^{-1}(\mathbf{X} - \mathbf{M}_i)\right]\right\}}{(2\pi)^{(N_f N_g)/2} \det(\phi_i)^{(N_f/2)} \det(\Psi_i)^{(N_g/2)}}, \quad (13)$$

where $\det(\cdot)$ represents the determinant of a matrix.

We define column vectorization as $\text{vec}(\cdot)$; then the mean of matrix X equals $\text{vec}(\mathbf{M}_i)$. By vectorizing the matrix Gaussian distribution, it can be converted into a multivariate Gaussian distribution of vector data as follows:

$$\mathbf{X} \sim \mathcal{M}\mathcal{N}_{N_f \times N_g}(\mathbf{M}, \phi, \Psi) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}(\text{vec}(\mathbf{M}), \Psi \otimes \phi), \quad (14)$$

where $\phi \in \mathbb{R}^{N_f \times N_f}$, $\Psi \in \mathbb{R}^{N_g \times N_g}$, and \otimes represents the Kronecker product operator. It can be seen from equation (14) that the covariance matrix of vectorized data can be transformed into a separable structure, which consists of the Kronecker product of two matrices.

The matrix-variate model in equation (11) corresponds to a specific structure for the covariance of the vectorized data. This model implies that the covariance matrix of the vectorized data can be decomposed into two parts. This separability property will be used in the algorithms proposed in this paper.

2.2.2. Bilinear Two-Dimensional Discriminant Locality Preserving Projection (B2DDLPP). Bilinear two-dimensional discriminant locality preserving projection (B2DDLPP) method is based on the matrix-variate Gaussian model. This model denotes that within-class covariance between any two spatial-spectral features can be decomposed into two parts.

Exchanging the numerator and denominator in the objective function of the 2DDLPP algorithm, the objective function of B2DDLPP can be reformulated as follows:

$$J(\mathbf{G}) = \frac{\mathbf{G}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{G}}{\mathbf{G}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{G}} = \frac{\mathbf{G}^T \mathbf{S}_b \mathbf{G}}{\mathbf{G}^T \mathbf{S}_w \mathbf{G}}, \quad (15)$$

where

$$\mathbf{S}_w = \frac{1}{2} \sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{X}_i^s - \mathbf{X}_j^s)(\mathbf{X}_i^s - \mathbf{X}_j^s)^T \mathbf{W}_{ij}^s, \quad (16)$$

$$\mathbf{S}_b = \frac{1}{2} \sum_{i,j=1}^Z (\mathbf{F}_i - \mathbf{F}_j)(\mathbf{F}_i - \mathbf{F}_j)^T \mathbf{B}_{ij}.$$

The projection directions can be obtained by maximizing the objective function, satisfying

$$\mathbf{G} = \arg \max_{\mathbf{G}} \frac{\mathbf{G}^T \mathbf{F} \mathbf{H} \mathbf{F}^T \mathbf{G}}{\mathbf{G}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{G}}. \quad (17)$$

Combining moment estimation of separable covariance matrix in separable LDA [29] and the high similarity between LDA and DLPP [30], we use the two following

equations to estimate the corresponding within-class covariance matrices. Decompose \mathbf{S}_w to get

$$\Psi = \frac{1}{2 * N_f} \sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{X}_i^s - \mathbf{X}_j^s)(\mathbf{X}_i^s - \mathbf{X}_j^s)^T \mathbf{W}_{ij}^s, \quad (18)$$

$$\phi = \frac{1}{2 * N_g} \sum_{s=1}^Z \sum_{i,j=1}^{n_s} (\mathbf{X}_i^s - \mathbf{X}_j^s)(\mathbf{X}_i^s - \mathbf{X}_j^s)^T \mathbf{W}_{ij}^s, \quad (19)$$

where \mathbf{X}_i^s denotes the i th feature matrix in class s , \mathbf{X}_j^s denotes the j th feature matrix in class s , \mathbf{W}_{ij}^s is the weight between \mathbf{X}_i^s and \mathbf{X}_j^s , and it is defined as $\mathbf{W}_{ij}^s = \exp(-\|\mathbf{X}_i^s - \mathbf{X}_j^s\|^2/t)$. n_s is the number of samples in the s th class. Z is the total number of classes. T is the transpose operation of the matrix.

Besides, the between-class scatter matrix can be considered as a separable structure. Vectorize the mean matrix in each class to get $\mathbf{SB} = \mathbf{SBR} \otimes \mathbf{SBL}$, where

$$\mathbf{S}_B = \frac{1}{2} \sum_{a,b=1}^Z (\mathbf{u}_a - \mathbf{u}_b)(\mathbf{u}_a - \mathbf{u}_b)^T \mathbf{B}_{ab}, \quad (20)$$

$$\mathbf{S}_{BL} = \frac{1}{2} \sum_{a,b=1}^Z (\mathbf{F}_a - \mathbf{F}_b)(\mathbf{F}_a - \mathbf{F}_b)^T \mathbf{B}_{ab}, \quad (21)$$

$$\mathbf{S}_{BR} = \text{tr}^{-1}(\mathbf{S}_{BL}) * \frac{1}{2} \sum_{a,b=1}^Z (\mathbf{F}_a - \mathbf{F}_b)^T (\mathbf{F}_a - \mathbf{F}_b) \mathbf{B}_{ab}, \quad (22)$$

where \mathbf{u} denotes the vectorization of \mathbf{F} ; $u = \text{vec} \mathbf{F}$. \mathbf{F}_s is the mean matrix of the s th class: $\mathbf{F}_s = (1/n_s) \sum_{i=1}^{n_s} \mathbf{X}_i^s$. \mathbf{B} is the weight matrix between any two classes' mean matrix and it is defined as

$$\mathbf{B}_{ab} = \exp\left(\frac{-\|\mathbf{F}_a - \mathbf{F}_b\|^2}{t}\right). \quad (23)$$

Using this assumption, we obtain the eigenvalues and eigenvectors of ϕ -1SBL, denoted by λ_l and \mathbf{u}_l , respectively. Similarly, we obtain the eigenvalues and the eigenvectors of Ψ -1SBR, denoted by γ_j and \mathbf{v}_j . Then, sort the two eigenvalues λ_l and γ_j in a descending order. The corresponding projection matrix can be constructed, respectively: $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{N_f}]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_g}]$. Thus, the feature matrix \mathbf{Y} is defined as

$$\mathbf{Y} = \mathbf{U}^T \mathbf{X} \mathbf{V}. \quad (24)$$

Finally, in order to get the d -dimensional features, we choose the γ_j elements of \mathbf{Y} which correspond to the d largest $\lambda_l \gamma_j$ values.

2.3. Materials and Experiments

2.3.1. Preprocessing and the Flow Chart of the Experiment. We conducted feature extraction using the proposed B2DDLPP in three databases. B2DDLPP algorithm is compared with LDA, two-dimensional linear discriminant analysis (2DLDA), DLPP, and 2DDLPP. It should be noted

that the same preprocessing method was applied to three data sets in this experiment: Before using the three feature extraction methods mentioned above, we apply the one-versus-rest (OVR) multiclass extension of the filter bank common spatial patterns (FBCSP) [31, 32] method to process the data. Taking four classes as an example, we take one class of samples as positive samples and the remaining samples as negative samples and perform FBCSP operations on the data to obtain a set of features. By analogy, a total of four sets of features can be obtained. Combine the four groups of features, and finally get a feature matrix of $N_f \times N_g$ size, where N_f is the number of frequency bands in FBCSP (9 in this article), $N_g = 2 \times m \times \text{Numclass}$, m is the number of pairs of features, and Numclass is the total number of classes. This method can reduce the dimensionality of the EEG data beforehand to ensure that the feature dimension of the data is less than the number of samples. This allows the subsequent feature extraction method to proceed smoothly. At the same time, effective spatial structure and frequency information are retained to obtain better classification results. As alpha band (8–13 Hz) and beta band (14–30 Hz) contain rich information for MI task, we divide all EEG signals into frequency subbands. The FBCSP employs a filter bank that covers 4–40 Hz, which comprises 9 bandpass filters that cover 4 Hz each [31]; and, in order to get a flatter delay response and low signal distortion, we use 6-order Chebyshev type II filter in this paper. The specific process is shown in Figure 1.

Considering that SVM classifier has been widely used in EEG classification [33–35], we use SVM classifier for classification in this paper. We divide the data set into two sets: training set and test set. The performance of BCI algorithms highly depends on the dimensionality of the feature space at the classifier’s input, denoted by d . For each method, the optimal dimensionality d_{op} of the feature space in training set is determined based on the average performance of each subject over all the validation runs. The feature space dimensionality for each method in test set is set based on the value of d_{op} in the validation phase. In order to make different methods more suitable for the data, so as to get the optimal dimension d_{op} , the selection range of parameter m in the preprocessing method is from 1 to 4. The value of the parameter m affects the data dimension after preprocessing.

The flow chart of the experiment is shown in Figure 2. All experiments are performed on MATLAB R2017a and Windows 10, with AMD core 2600X CPU and 16 GB RAM.

2.3.2. Data Set

(1) *Set 1: BCI competition IV, data sets 2a (Exp.1)*. The ultimate purpose of this experiment is to classify the following motor-imagery tasks: left hand, right hand, feet, and tongue movement. This data set contains EEG signals of nine healthy subjects. It is recorded in two sessions and the signals are recorded using 22 Ag/AgCl electrodes at 250 Hz sampling rate. Each session consists of six runs and each of which includes 48 trials of a length of 3 seconds, yielding a total of 288 trials per session. These two sessions are used as

training set and test set, respectively. It should be mentioned that there are three Electrooculogram (EOG) channel recordings in this data set and they can be used as a reference for denoising. In this paper, we chose the time period of 3 s to 6 s and in order to preserve complete spatial information, all channels are reserved.

(2) *Set 2: BCI competition III, data sets IIIa (Exp.2)*. This data set consists of recordings from three healthy subjects (k3b, k6b, and 11b). Each subject sat in a relaxing chair with armrests and was asked to perform imagery movements with four different tasks: left hand, right hand, foot, and tongue. Each subject completed 60 trials per class. Recordings were made with a 60-channel EEG amplifier from Neuroscan with the left mastoid for reference and the right mastoid as ground. EEG signals were recorded with a sampling rate of 250 Hz and filtered between 1 and 50 Hz with the notch filter on. In this paper, we chose the time period of 3 s to 7 s and all channels are reserved.

(3) *Set 3: data sets 3 (Exp.3)*. The third data set used in this paper was obtained by our laboratory. This data set consists of recordings from 10 subjects and each subject sat in a relaxing chair with armrests and was asked to perform imagery movements with three different tasks: left hand, right hand, and the idle state. Motor imaging duration is 4 seconds. The signals are recorded using 62 Ag/AgCl electrodes at 1000 Hz sampling rate and each subject completed 125 trials per class, yielding a total of 375 trials. Among them, 300 trials are used as the training set, and the remaining data are used as the test set. In order to reduce data redundancy and improve the efficiency of data processing, we performed downsampling to reduce the frequency to 250 Hz.

3. Results and Discussion

In this section, we show the results of different methods on training set and the test set. We used a 5-fold cross-validation method to process the training set and then use the optimal parameters obtained in the training set for the classification of the test set. All results are displayed by the classification accuracy.

The pseudocode for training the B2DDLPP feature extractor is presented in Table 1 and the validation results of training set for Exp.1, Exp.2, and Exp.3 are presented in Tables 2–4, respectively. For each subject, the highest average recognition rate over all the cross-validation runs and the corresponding feature dimension of different algorithms are reported. We also studied the case where no feature extraction method is utilized and the FBCSP features are directly passed to the classifier.

Table 2 shows that the average recognition accuracy of B2DDLPP is improved by 2.81%, 8.47%, 7.07%, 4.86%, and 3.8% compared to those of None, LDA, 2DLDA, DLPP, and 2DDLPP, respectively. From the results, B2DDLPP has the highest accuracy rate for each subject; and whether to use 2D method or manifold learning method, it can be seen from the comparison result that these two factors have an

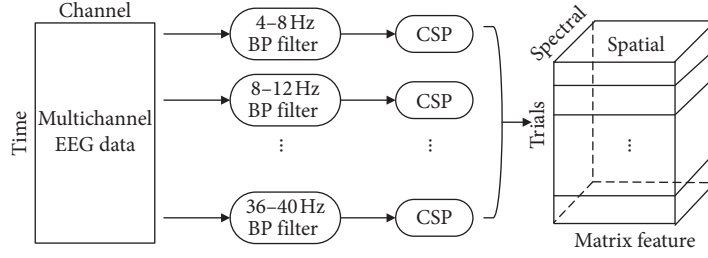


FIGURE 1: Operation steps of FBCSP. Each trial in matrix feature contains spatial-spectral information.

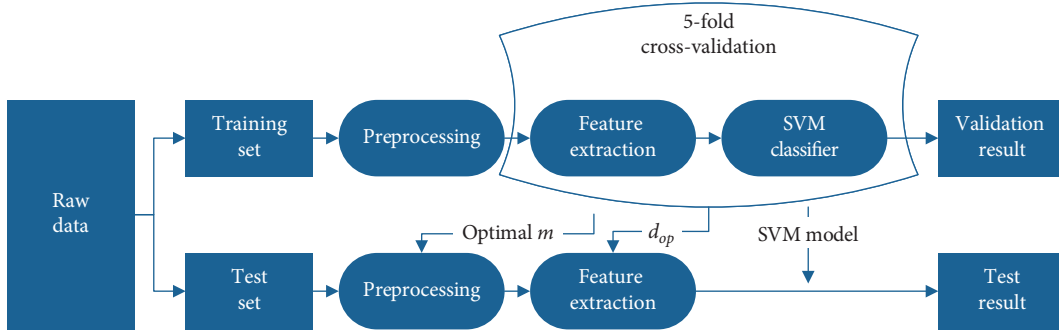


FIGURE 2: The flow chart of the experiment.

TABLE 1: The pseudocode for training the B2DDLPP feature extractor.

Algorithm: B2DDLPP

Inputs:

- Training sample $X_{N_f \times N_g}$. The total number of samples is N . The number of training samples in each class is N_i , $1 \leq i \leq Z$.

Outputs:

- The feature extraction operators $U_{N_f \times N_f}$ and $V_{N_g \times N_g}$ - the corresponding λ_l and γ_j values which determine the priority in selecting the elements in feature matrix.

Procedure:

1. Calculate the spatial covariance matrix ψ and the spectral covariance matrix ϕ according to (18) and (19).
2. Calculate S_{BL} and S_{BR} according to (21) and (22).
3. Calculate the eigenvalues λ_l and the corresponding eigenvectors u_l of $\phi^{-1}S_{BL}$, $1 \leq l \leq N_f$. And calculate the eigenvalues γ_j and the corresponding eigenvectors v_j of $\psi^{-1}S_{BR}$. Construct U and V .
4. Calculate the feature matrix Y according to (24).
5. Calculate the feature matrix Y according to (24).
6. Choose the y_{ij} elements of Y which correspond to the d largest $\lambda_l \gamma_j$ values. D is the dimension after feature extraction.

important influence on the accuracy of EEG classification. In the operation of LDA and DLPP, vectorization of the extracted matrix features will cause the loss of important information. There is an indispensable connection in the matrix structure, which has a great influence on the accuracy of classification. The accuracy of 2DLDA and 2DDLPP using the 2D method has improved compared with the method without 2D processing; and, compared to 2DLDA, the accuracy of 2DDLPP increased by 3.27%, which shows that keeping locality-geometric structure of EEG data in a neighboring area plays an effective role in improving the classification accuracy.

When compared with 2DDLPP, the result of B2DDLPP has greatly improved. 2DDLPP only transforms the rows or columns of matrix data, which inevitably leads to unnecessary information loss. B2DDLPP uses bilinear structure to decompose the covariance matrix into row part and column part, which has a better grasp of the internal interconnections of spatial data. Therefore, B2DDLPP can extract more discriminative information. We can draw the same conclusion in Table 3.

The overall recognition accuracy in Table 4 is not as high as that in the other two tables. The reason may be that the subject of this data set is not a professionally trained person

TABLE 2: Cross-validation performance results for different algorithms in Exp.1.

Feature extraction	Subj.1 (%)	Subj.2 (%)	Subj.3 (%)	Subj.4 (%)	Subj.5 (%)	Subj.6 (%)	Subj.7 (%)	Subj.8 (%)	Subj.9 (%)	Average (%)
None	84.72 m = 3	57.70 m = 1	87.23 m = 2	54.93 m = 4	63.81 m = 2	50.71 m = 4	88.61 m = 1	87.84 m = 4	81.05 m = 1	72.88
LDA	79.22 m = 1	51.78 m = 1	79.57 m = 1	50.38 m = 1	63.94 m = 1	44.81 m = 1	85.91 m = 1	75.04 m = 1	74.34 m = 1	67.22
	$d_{op} = 72$	$d_{op} = 64$	$d_{op} = 53$	$d_{op} = 47$	$d_{op} = 7$	$d_{op} = 6$	$d_{op} = 10$	$d_{op} = 22$	$d_{op} = 58$	
2DLDA	83.27 m = 1	55.19 m = 2	78.44 m = 2	50.33 m = 1	67.32 m = 2	44.56 m = 1	86.10 m = 1	77.43 m = 2	74.97 m = 1	68.62
	$d_{op} = 6$	$d_{op} = 5$	$d_{op} = 4$	$d_{op} = 4$	$d_{op} = 3$	$d_{op} = 5$	$d_{op} = 4$	$d_{op} = 7$	$d_{op} = 5$	
DLPP	80.55 m = 2	55.56 m = 1	84.38 m = 1	54.17 m = 2	63.20 m = 1	50.69 m = 2	88.19 m = 1	80.55 m = 2	80.21 m = 2	70.83
	$d_{op} = 134$	$d_{op} = 65$	$d_{op} = 72$	$d_{op} = 106$	$d_{op} = 69$	$d_{op} = 122$	$d_{op} = 51$	$d_{op} = 132$	$d_{op} = 128$	
2DDLPP	84.03 m = 1	56.40 m = 1	86.60 m = 1	52.71 m = 3	65.89 m = 2	49.66 m = 2	88.97 m = 1	81.79 m = 2	80.98 m = 1	71.89
	$d_{op} = 6$	$d_{op} = 7$	$d_{op} = 7$	$d_{op} = 9$	$d_{op} = 9$	$d_{op} = 6$	$d_{op} = 9$	$d_{op} = 7$	$d_{op} = 9$	
B2DDLPP	86.80 m = 2	61.11 m = 1	89.61 m = 1	58.64 m = 2	71.91 m = 2	53.48 m = 1	90.61 m = 1	85.38 m = 4	83.68 m = 1	75.69
	$d_{op} = 142$	$d_{op} = 58$	$d_{op} = 44$	$d_{op} = 126$	$d_{op} = 70$	$d_{op} = 67$	$d_{op} = 63$	$d_{op} = 167$	$d_{op} = 52$	

For each method and each subject, optimal m related to FBCSP's output and the optimal dimension (d_{op}) are presented.

TABLE 3: Cross-validation performance results for different algorithms in Exp.2.

Feature Extraction	Subj.1 (% m, d_{op})	Subj.2 (% m, d_{op})	Subj.3 (% m, d_{op})	Average (%)
None	81.67 m = 1	56.67 m = 1	58.33 m = 2	65.56
LDA	78.89 m = 1, $d_{op} = 59$	58.33 m = 2, $d_{op} = 37$	53.33 m = 2, $d_{op} = 103$	63.52
2DLDA	79.89 m = 2, $d_{op} = 8$	58.50 m = 2, $d_{op} = 3$	55.17 m = 2, $d_{op} = 3$	64.52
DLPP	85.00 m = 1, $d_{op} = 58$	57.50 m = 1, $d_{op} = 22$	52.50 m = 1, $d_{op} = 30$	65.00
2DDLPP	85.44 m = 3, $d_{op} = 5$	54.33 m = 1, $d_{op} = 7$	57.67 m = 1, $d_{op} = 7$	65.81
B2DDLPP	87.22 m = 1, $d_{op} = 37$	60.00 m = 2, $d_{op} = 79$	64.17 m = 1, $d_{op} = 26$	70.46

For each method and each subject, optimal m related to FBCSP's output and the optimal dimension (d_{op}) are presented.

having a long-term specialized motor imaging training. This insufficient effective information in the original data leads to the result, whereas, in Exp.1 and Exp.2, the subject is the specially trained group of people, which generates less noise when performing a motor imaging task. Table 4 shows that the average recognition accuracy of B2DDLPP is improved by 2.34%, 6.75%, 5.7%, 4.66%, and 3.66% compared to those of None, LDA, 2DLDA, DLPP, and 2DDLPP, respectively. This result also fully proves the effectiveness of the 2D method and manifold learning method for the extraction of spatial and structural information from EEG signals.

It can be seen from the results of the three tables that although B2DDLPP has the highest accuracy rate, the classification accuracy of other feature extraction methods listed in the table is a bit lower than the accuracy rate without feature extraction. The reason can be attributed to the two following points: First, the methods of LDA and DLPP are not very robust to noise. Among the acquisition channels, the channels related to motor imagery are only a part of them. In this experiment, in order to ensure the integrity of the channel structure, we use the data of all channels, which brings redundant data and noise. Second, the extraction of effective information is insufficient. 2DLDA and 2DDLPP methods only transform the rows or

columns of matrix data and this one-sided compression is not sufficient for the extraction of effective features in the spatial-spectral matrix.

It should be noted that the parameter t related to the calculation of the weight matrix in DLPP, 2DDLPP, and B2DDLPP is set to 1.

The following shows the results of the test set. The results of test set for Exp.1 and Exp.2 are presented in Figures 3 and 4, respectively. Note that the feature space dimensionality for each method in test set is set based on the value of d_{op} in the validation phase.

We can see from the figures that the performance results on the test data show a trend that is very similar to that of the performance results during the cross-validation phase.

In addition, we studied the influence of different feature dimensions on the effects of these methods. Figure 5 shows recognition accuracy of all the methods on different number of dimensions under three subjects in Exp.2. The results in the figure are shown by the average accuracy of 5-fold cross-validation. We can see from the figure that the overall trend of various methods rises slightly with the increase of the dimension and finally tends to be flat. Combining the curve results of the three subjects, the accuracy of the B2DDLPP method is the highest, followed by 2DDLPP. The results of

TABLE 4: Cross-validation performance results for different algorithms in Exp.3.

Feature extraction	Subj.1 (%)	Subj.2 (%)	Subj.3 (%)	Subj.4 (%)	Subj.5 (%)	Subj.6 (%)	Subj.7 (%)	Subj.8 (%)	Subj.9 (%)	Subj.10 (%)	Average (%)
None	43.33 m = 1	79.67 m = 1	58.31 m = 2	50.66 m = 2	48.33 m = 2	41.00 m = 2	57.60 m = 2	42.36 m = 1	39.33 m = 1	84 m = 1	54.45
LDA	39.33 m = 1	73.67 m = 1	50.00 m = 1	45.00 m = 1	43.34 m = 1	40.33 m = 1	54.77 m = 1	39.67 m = 1	39.00 m = 1	75.33 m = 1	50.04
	$d_{op} = 23$	$d_{op} = 18$	$d_{op} = 39$	$d_{op} = 49$	$d_{op} = 28$	$d_{op} = 41$	$d_{op} = 11$	$d_{op} = 50$	$d_{op} = 52$	$d_{op} = 24$	
2DLDA	41.13 m = 1	72.33 m = 1	52.67 m = 1	47.52 m = 1	45.67 m = 1	39.83 m = 1	54.93 m = 1	40.33 m = 1	39.13 m = 1	77.33 m = 1	51.09
	$d_{op} = 5$	$d_{op} = 3$	$d_{op} = 4$	$d_{op} = 6$	$d_{op} = 2$	$d_{op} = 2$	$d_{op} = 1$	$d_{op} = 4$	$d_{op} = 2$	$d_{op} = 5$	
DLPP	41.67 m = 2	75.33 m = 1	52.67 m = 2	46.00 m = 2	49.33 m = 2	39.33 m = 1	56.00 m = 2	39.67 m = 2	39.67 m = 1	81.67 m = 1	52.13
	$d_{op} = 98$	$d_{op} = 11$	$d_{op} = 44$	$d_{op} = 8$	$d_{op} = 20$	$d_{op} = 29$	$d_{op} = 23$	$d_{op} = 98$	$d_{op} = 20$	$d_{op} = 7$	
2DDLPP	41.67 m = 1	78.00 m = 1	55.33 m = 2	47.33 m = 2	45.33 m = 1	40.67 m = 1	55.33 m = 2	41.67 m = 2	40.33 m = 2	79.67 m = 1	53.13
	$d_{op} = 2$	$d_{op} = 2$	$d_{op} = 7$	$d_{op} = 9$	$d_{op} = 2$	$d_{op} = 7$	$d_{op} = 7$	$d_{op} = 9$	$d_{op} = 5$	$d_{op} = 3$	
B2DDLPP	45.67 m = 1	80.67 m = 2	59.33 m = 2	51.67 m = 2	52.82 m = 2	45.00 m = 2	59.70 m = 2	44.40 m = 2	42.33 m = 1	86.33 m = 1	56.79
	$d_{op} = 35$	$d_{op} = 82$	$d_{op} = 51$	$d_{op} = 45$	$d_{op} = 94$	$d_{op} = 14$	$d_{op} = 19$	$d_{op} = 79$	$d_{op} = 49$	$d_{op} = 54$	

For each method and each subject, optimal m related to FBCSP's output and the optimal dimension (d_{op}) are presented.

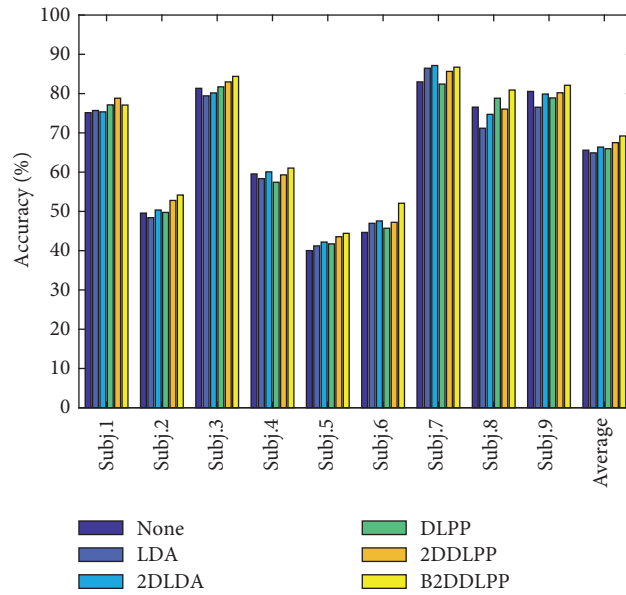


FIGURE 3: The results of the test set for Exp.1. The figure shows the accuracy of 9 subjects and the average accuracy of them.

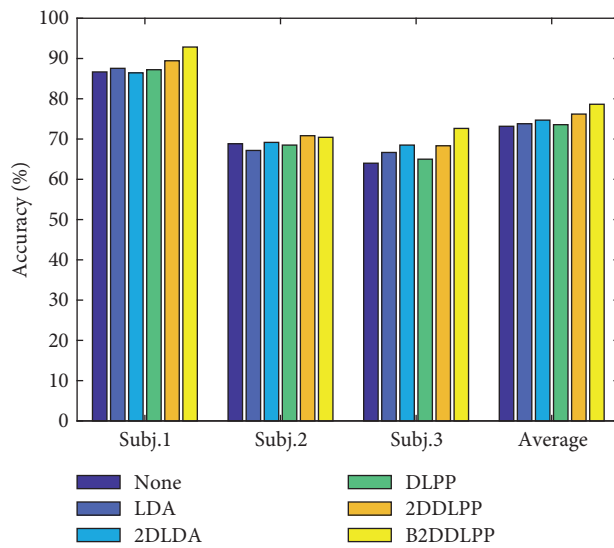


FIGURE 4: The results of the test set for Exp.2.

DLPP and 2DLDA are not much different, and the worst is the LDA method. Different methods have different extraction capabilities for spatial information and structural information and the effect of this extraction determines the

final accuracy. The results also show that B2DDLPP can achieve a high accuracy rate in the case of low feature dimensions, which can greatly help reduce the complexity of data and extract effective features.

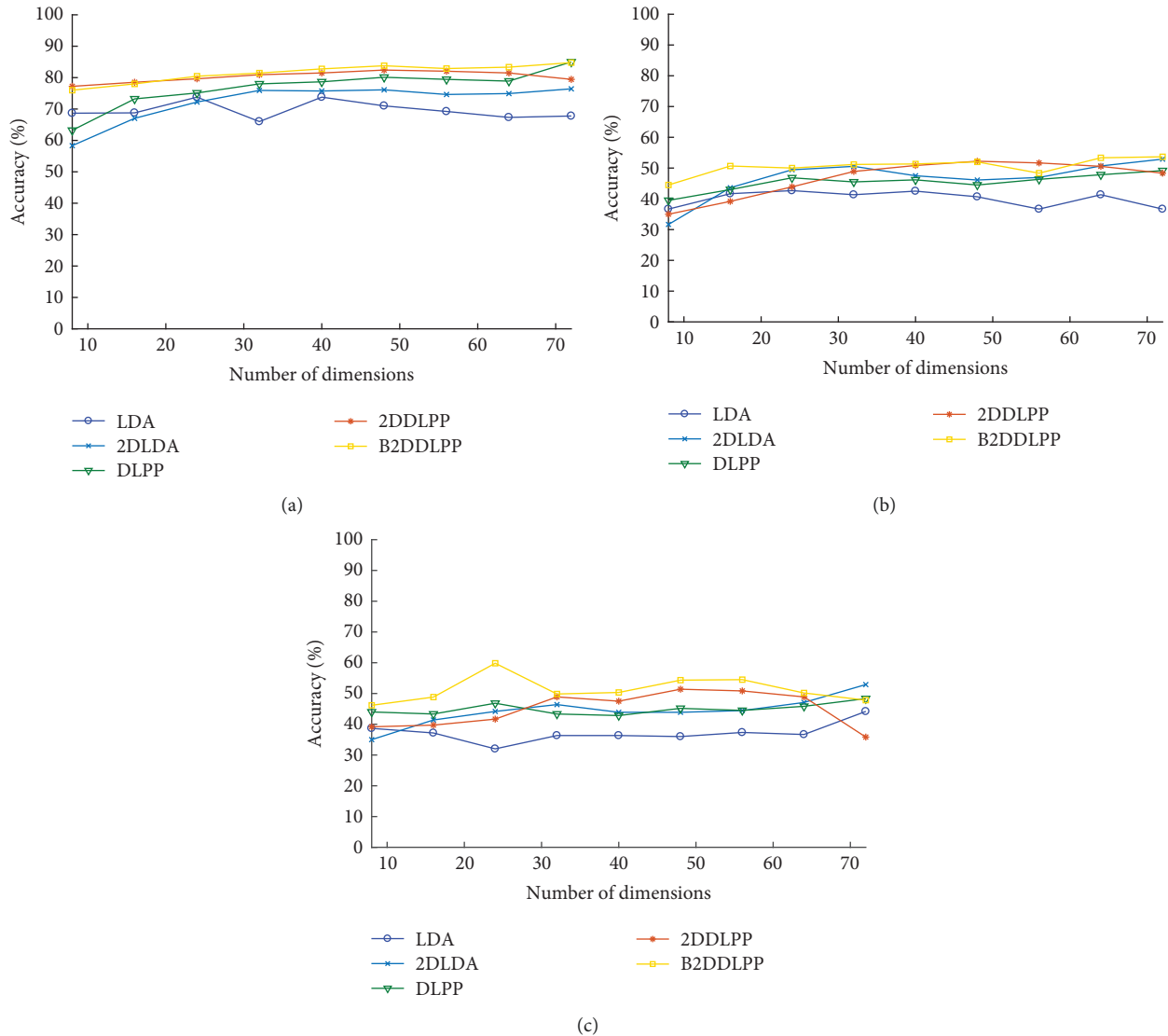


FIGURE 5: Five-fold cross-validation performance on different number of dimensions on Exp.2: (a) the accuracy of different dimensions in subj.1; (b) the accuracy of different dimensions in subj.2; (c) the accuracy of different dimensions in subj.3.

4. Conclusions

In this paper, we propose a B2DDLPP algorithm. By adding the matrix-variate Gaussian model and a bilinear structure, B2DDLPP decomposes the within-class covariance matrix and the between-class scatter matrix to obtain the optimal projection matrix. As a matrix-based method, B2DDLPP is more effective for extracting spatial information than vector-based methods. At the same time, the bilinear structure further enhances this effect.

In order to fully verify the effectiveness of the algorithm, we apply the B2DDLPP algorithm to three EEG data sets in this paper. The results show that B2DDLPP has a higher feature extraction performance compared to other methods.

Data Availability

The Exp1 and Exp2 data used to support the findings of this study are available at <http://www.bbci.de/competition/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the Joint Funds of the National Natural Science Foundation of China (U1609218) and NSFC (61633010) and Key Research and Development Project of Zhejiang Province (2020C04009).

References

- [1] D. Mattia, L. Astolfi, J. Toppi, M. Petti, F. Pichiorri, and F. Cincotti, "Interfacing brain and computer in neuro-rehabilitation," in *Proceedings of the 2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1-2, Gangwon-do, Korea, February 2016.

- [2] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [3] G. Pfurtscheller and F. H. Lopes Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [4] P. N. Kumar and H. Kareemullah, "EEG signal with feature extraction using SVM and ICA classifiers," in *Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES2014)*, pp. 1–7, S.A. Engineering College, Chennai, India, February 2014.
- [5] M. I. Gursoy and A. Subast, "A comparison of PCA, ICA and LDA in EEG signal classification using SVM," in *Proceedings of the 2008 IEEE 16th Signal Processing, Communication and Applications Conference*, pp. 1–4, Aydin, Turkey, April 2008.
- [6] H.-Gi Yeom and S. Kwee-Bo, "ERS and ERD analysis during the imaginary movement of arms," in *Proceedings of the 2008 International Conference on Control, Automation and Systems*, pp. 2476–2480, Seoul, Korea, October 2008.
- [7] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, p. 2319, 2000.
- [8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, p. 2323, 2000.
- [9] P. Ataee, A. Yazdani, S. K. Setarehdan, and H. A. Noubari, "Manifold learning applied on EEG signal of the epileptic patients for detection of normal and pre-seizure states," in *Proceeding of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5489–5492, Lyon, France, September 2007.
- [10] A. M. Álvarez-Meza, G. Daza-Santacoloma, and G. Castellanos-Domínguez, "Biomedical data analysis by supervised manifold learning," in *Proceeding of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 41–44, San Diego, CA, USA, August 2012.
- [11] J. Kortelainen, E. Väärinen, and T. Seppänen, "Isomap approach to EEG-based assessment of neurophysiological changes during anesthesia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 2, pp. 113–120, 2011.
- [12] R. Yuvaraj, M. Murugappan, N. M. Ibrahim et al., "Optimal set of EEG features for emotional state classification and trajectory visualization in Parkinson's disease," *International Journal of Psychophysiology*, vol. 94, no. 3, pp. 482–495, 2014.
- [13] M. Xing, *Thought Chart: Tracking Dynamic EEG Brain Connectivity with Unsupervised Manifold Learning*, pp. 149–157, Springer International Publishing, New York, NY, USA, 2016.
- [14] M. Li, X. Luo, J. Yang, Y. Sun, and E. Martinelli, "Applying a locally linear embedding algorithm for feature extraction and visualization of MI-EEG," *Journal of Sensors*, vol. 2016, Article ID 7481946, 9 pages, 2016.
- [15] K. Sadatnejad, S. Ghidary, and Shiry, "Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application," *Neurocomputing*, vol. 179, pp. 152–160, 2016.
- [16] E. Krivov and M. Belyaev, "Dimensionality reduction with isomap algorithm for EEG covariance matrices," in *Proceedings of the 2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, Gangwon-do, Korea, February 2016.
- [17] M. Wan, G. Yang, C. Sun, and M. Liu, "Sparse two-dimensional discriminant locality-preserving projection (S2DDLPP) for feature extraction," *Soft Computing*, vol. 23, no. 14, pp. 5511–5518, 2019.
- [18] X. He and P. Niyogi, "Locality preserving projections," *NIPS*, vol. 16, pp. 234–241, 2003.
- [19] X. He, D. Cai, S. Yan, and H. J. Zhang, "Neighborhood preserving embedding," *IEEE International Conference on Computer Vision*, pp. 1208–1213, 2005.
- [20] W. Yu, X. Teng, and C. Liu, "Discriminant locality preserving projections: a new method to face representation and recognition," in *Proceedings of the 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 201–207, October 2005.
- [21] M. S. Mahanta, A. S. Aghaei, and K. N. Plataniotis, "A Bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals," in *Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3955–3958, San Diego, CA, USA, August 2012.
- [22] R. Zhi and Q. Ruan, "Facial expression recognition based on two-dimensional discriminant locality preserving projections," *Neurocomputing*, vol. 71, no. 7-9, pp. 1730–1734, 2018.
- [23] Z. Qiao, L. Zhou, and J. Z. Huang, "Sparse linear discriminant analysis with applications to high dimensional low sample size data," *Iaeng International Journal of Applied Mathematics*, vol. 39, no. 1, pp. 48–60, 2009.
- [24] A. Jiang, Q. Wang, J. Shang, and X. Liu, "Sparse common spatial pattern for EEG channel reduction in brain-computer interfaces," in *Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pp. 1–4, Shanghai, China, November 2018.
- [25] D. Wen, P. Jia, Q. Lian, Y. Zhou, and C. Lu, "Review of sparse representation-based classification methods on EEG signal processing for epilepsy detection, brain-computer interface and cognitive impairment," *Frontiers in Aging Neuroscience*, vol. 8, p. 2016, 2016.
- [26] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *International Journal of Neural Systems*, vol. 27, no. 2, Article ID 1650032, 2017.
- [27] J. Wen, X. Fang, J. Cui et al., "Robust sparse linear discriminant analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 390–403, 2019.
- [28] A. Gupta and D. Nagar, *Matrix Variate Distributions*, Vol. 104, CRC Press, Boca Raton, FL, USA, 1999.
- [29] J. Zhao, P. L. H. Yu, L. Shi, and S. Li, "Separable linear discriminant analysis," *Computational Statistics and Data Analysis*, vol. 56, no. 12, pp. 4290–4300, 2012.
- [30] R. Zhi and Q. Ruan, "Facial expression recognition based on two-dimensional discriminant locality preserving projections," *Neurocomputing*, vol. 71, no. 7-9, pp. 1730–1734, 2008.
- [31] K. K. Ang, "filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proceedings of the 2008 IEEE International Joint Conference on Neural Networks*, pp. 2390–2397, IEEE World Congress on Computational Intelligence, Hong Kong, China, June 2008.
- [32] 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 Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 993–1002, 2004.

- [33] Y. Zhang, J. Liu, J. Liu, J. Sheng, and J. Lv, "EEG recognition of motor imagery based on SVM ensemble," in *Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI)*, pp. 866–870, IEEE, Nanjing, China, November 2018.
- [34] M. Riyadi, T. Prakoso, F. Whailan, M. David, and A. Hidayatno, "Classification of EEG-based brain waves for motor imagery using support vector machine," in *Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp. 422–425, IEEE, Bandung, Indonesia, October 2019.
- [35] Y. Park and W. Chung, "Optimal channel selection using covariance matrix and cross-combining region in EEG-based BCI," in *Proceedings of the 2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, IEEE, Gangwon, South Korea, February 2019.