

## Research Article

# Embedding Tangent Space Extreme Learning Machine for EEG Decoding in Brain Computer Interface Systems

Mingwei Zhang,<sup>1</sup> Yao Hou ,<sup>2</sup> Rongnian Tang,<sup>2</sup> and Youjun Li<sup>2</sup>

<sup>1</sup>Physics and Electronic Engineering, Hainan Normal University, Haikou 571158, China

<sup>2</sup>Mechanical and Electrical Engineering College, Hainan University, Haikou 570228, China

Correspondence should be addressed to Yao Hou; [mrhouyao@163.com](mailto:mrhouyao@163.com)

Received 31 March 2021; Revised 7 August 2021; Accepted 11 August 2021; Published 17 September 2021

Academic Editor: Radek Matušů

Copyright © 2021 Mingwei Zhang 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 motor imagery brain computer interface system, the spatial covariance matrices of EEG signals which carried important discriminative information have been well used to improve the decoding performance of motor imagery. However, the covariance matrices often suffer from the problem of high dimensionality, which leads to a high computational cost and overfitting. These problems directly limit the application ability and work efficiency of the BCI system. To improve these problems and enhance the performance of the BCI system, in this study, we propose a novel semisupervised locality-preserving graph embedding model to learn a low-dimensional embedding. This approach enables a low-dimensional embedding to capture more discriminant information for classification by efficiently incorporating information from testing and training data into a Riemannian graph. Furthermore, we obtain an efficient classification algorithm using an extreme learning machine (ELM) classifier developed on the tangent space of a learned embedding. Experimental results show that our proposed approach achieves higher classification performance than benchmark methods on various datasets, including the BCI Competition IIA dataset and in-house BCI datasets.

## 1. Introduction

Brain computer interface (BCI) systems have been developed to provide a novel and promising alternative method for humans to interact with their environment. Among currently operational BCI systems, many have been based on electroencephalography (EEG) signals, owing to the non-invasive nature of the technique and affordable data acquisition equipment, both of which also facilitate real-time operation [1]. Although many EEG signals such as P300 waves or steady-state visual evoked potentials (SSVEP) could be used for BCI [2], in this study, we focus on motor imagery-based BCI systems, with which trained subjects can voluntarily generate EEG control signals by imagining movements of different parts of their bodies [3]. With the advantage not requiring external stimulation, motor imagery-based BCI has proven effective in practice.

The common spatial pattern (CSP) algorithm has been shown to be one of the most effective methods for feature

extraction from motor imagery signals [4]. The success of CSP techniques in BCI applications depends heavily on learning a proper spatial filter from a training dataset. Some extensions of CSP were proposed to improve the performance of feature extraction in motor imagery classification. In [5], a novel regularized CSP combined with correlation-based channel selection was proposed to extract effective features of EEG. In [6], a discriminative canonical pattern matching combined with CSP was used to extract the movement-related features. In [7], to capture the shared salient information across multiple related spatial patterns, a multiview learning-based sparse optimization was proposed to jointly extract robust CSP features with L1L2-norm regularization. In [8], a generic model training by weighted linear discriminant analysis was proposed to reduce the correction time of the BCI system. In [9], a fusion method for internal feature selection was proposed to fuse different feature selection rules. It can consume less computational cost and obtain high performance of motor imagery

classification. Typically, in such motor imagery classification system, some discriminant information contained in the testing dataset remains unutilized. However, graph-based models are able to utilize such information because they naturally incorporate diverse types of information, such as the relationship between training and testing datasets [10].

Graph construction is an essential step in graph-based models. Two conventional methods are commonly used, including the  $k$ -nearest neighbor method and the  $\varepsilon$ -ball-based method [11]:

- (1)  $k$ -nearest neighbor methods consider data samples  $x_i$  and  $x_j$  as neighbors if  $x_i$  is among the  $k$  nearest neighbors of  $x_j$  or  $x_j$  is among the  $k$  nearest neighbors of  $x_i$ , where the nearest neighbors are determined as the samples with the smallest Euclidean distance
- (2)  $\varepsilon$ -ball neighborhood methods consider data samples  $x_i$  and  $x_j$  as neighbors if  $\|x_i - x_j\|^2 < \varepsilon$ , where  $\|\cdot\|^2$  is the Euclidean norm and  $\varepsilon \in \mathbb{R}$

The above traditional graph construction methods are relatively simple to implement and have been applied effectively in practice. However, if the data samples considered do not lie in a Euclidean space, the  $k$ -nearest neighbor and  $\varepsilon$ -ball-based methods cannot guarantee the connectivity of the whole graph and often result in several separated subgraphs because neighbors defined by Euclidean distance cannot characterize the real geometrical relations among samples well [12]. An obvious example is the covariance feature data used in motor imagery-based BCI. Because covariance features of EEG signals with the form of symmetric positive-definite (SPD) matrices lie on a Riemannian manifold [13], we propose an extension of the graph construction approach based on Riemannian geodesic distance to construct a proper graph for EEG signals, namely, a Riemannian graph. In fact, EEG classification techniques on high-dimensional Riemannian manifolds have recently received increased attention [14]. For example, an efficient classification algorithm called tangent space linear discriminant analysis (TS+LDA), was proposed in [14]. TS+LDA mapped all the data points in a Riemannian manifold into its tangent space, which is known as the best hyperplane for classification tasks [15], and then applied linear discriminant analysis (LDA) for classification in the tangent space.

After constructing a Riemannian graph for an EEG signal, many graph-based algorithms can be exploited on the constructed Riemannian graph. Such graph embedding algorithms are among the most successful in graph-based data processing on high-dimensional Riemannian manifolds because the most important information is always contained in a low-dimensional embedding of a high-dimensional Riemannian manifold. Many state-of-the-art semisupervised and unsupervised algorithms have been proposed. Semisupervised learning methods can utilize both training and test data to boost the algorithmic performance. Many related works, such as [16–18], were proposed to classify EEG based on training and test data in the context of semisupervised learning. In [16], a self-

training support vector machine (SVM) method was proposed to use the test data for updating the parameter of SVM model. In [17], an iterative semisupervised SVM method was proposed for channel selection and classifier training. In [18], a semisupervised expectation maximization (EM) was incorporated into CSP to extract more effective features. On the contrary, many unsupervised dimensionality reduction methods also are proposed to learn a low-dimensional embedding preserving similarities between vertex pairs on such graphs, including locally linear embedding [19], locality-preserving projection [20], and Laplacian eigenmaps [21].

Considering locality preservation, in this study, we propose a bilinear locality-preserving embedding learning algorithm on a constructed Riemannian graph. Because the learned embedding was also an SPD matrix space, inspired by the TS+LDA algorithm, we applied an extreme learning machine (ELM) [22, 23] method on the embedding tangent space for classification. Many related works of graph embedding methods and ELM-based methods have been recently proposed. In [24], a hybrid-graph discriminant learning method was proposed to reveal the complex high-order relationship of the original data and to extract the discriminant features. In [25], a local constrained manifold structure collaborative preserving embedding was proposed to enhance the intraclass compactness and the interclass separability. In [26], to improve the classification performance of motor imagery, a sparse Bayesian ELM algorithm was proposed to automatically control the model complexity and exclude redundant hidden neurons. In [27], a multi-kernel ELM method integrating two types of kernels was proposed to effectively explore the supplementary information from multiple nonlinear feature spaces.

In [28], a clustering-based multitask feature learning method was proposed to improve EEG pattern decoding. Most of these studies have achieved great success in graph embedding and classification. However, the above methods learn embedding or classifier on Euclidean space and ignore that the covariance matrices lie on Riemannian manifold. The contributions of this study are threefold.

- (1) A graph model of the covariance features of motor imagery signals based on Riemannian geodesic distance was constructed. The neighborhood relationships among data samples were characterized by Riemannian geodesics because the EEG covariance features endowed with a Riemannian metric formed a Riemannian manifold.
- (2) A novel framework of graph embedding is proposed for the constructed Riemannian graph. The calculation of graph embedding is formulated as an eigenvalue problem.
- (3) A classification algorithm, called an embedding tangent space extreme learning machine (ETS+ELM), is proposed to learn low-dimensional embeddings. Experimental results show that the

proposed algorithm achieved higher classification performance than prior methods used for comparison

The remainder of this paper is organized as follows. In Section 2, some basic concepts concerning SPD matrix spaces and BCI data models are briefly reviewed. In Section 3, we introduce the construction of Riemannian graphs and propose a bilinear locality-preserving graph embedding algorithm and an embedding tangent space ELM for classification. Extensive experimental results are presented in Section 4 to demonstrate the effectiveness of the proposed method. Finally, some conclusions are presented in Section 5.

## 2. Concepts of Riemannian Geometry and BCI Data Models

**2.1. Riemannian Geometry.** In mathematics, Riemannian manifold  $\mathcal{M}$  is defined as a smooth manifold equipped with a Riemannian metric  $\langle \cdot, \cdot \rangle$ , i.e., a smoothly varying inner product  $(\langle \cdot, \cdot \rangle_p, p \in \mathcal{M})$ , on its tangent space [29]. A smooth manifold is a topological manifold with a smooth structure locally similar to Euclidean space. Every point on this topological manifold has a neighborhood for which there exists a homeomorphism mapping the neighborhood to a Euclidean space. The materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed subsections if several methods are described.

For many applications, the space of SPD matrices endowed with a Riemannian metric is one of most important Riemannian manifolds [30]. Owing to the properties of SPD matrices, explicit formulae exist for the major concepts of Riemannian geometry. Therefore, these concepts can be applied relatively easily in the context of pattern recognition.

**2.2. Space of Symmetric Positive-Definite Matrices.** Denoting the space of symmetric matrices by

$$\mathcal{S}(N) = \{\mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{P} = \mathbf{P}^T\}, \quad (1)$$

and the space of positive-definite matrices by

$$\mathcal{P}(N) = \{\mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{u}^T \mathbf{P} \mathbf{u} > 0, \forall \mathbf{u} \in \mathbb{R}^N\}, \quad (2)$$

the space of SPD matrices is defined as

$$\mathcal{SPD}(N) = \mathcal{S}(N) \cap \mathcal{P}(N). \quad (3)$$

The space of SPD matrices lies on a differentiable Riemannian manifold [30]. Hence, all the mathematical tools defined in Riemannian geometry can be applied to any  $\mathcal{SPD}(N)$ .

The Riemannian distance between two matrices  $\mathbf{P}_1, \mathbf{P}_2 \in \text{calSPD}(N)$  is the minimum length of the curves connecting  $\mathbf{P}_1$  and  $\mathbf{P}_2$  on a Riemannian manifold [31]. There are many possible mathematical definitions of the Riemannian distance. In this study, we adopt the following formula for Riemannian distance:

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \|\log(\mathbf{P}_1^{-1} \mathbf{P}_2)\|_F = \left[ \sum_{i=1}^N \log^2 \eta_i \right]^{(1/2)}, \quad (4)$$

where  $\|\cdot\|_F$  is the Frobenius norm of a matrix and  $\eta_i$  is the  $i$ th real eigenvalue of  $\mathbf{P}_1^{-1} \mathbf{P}_2$ .

The associated Euclidean tangent space is critical in the analysis of a Riemannian manifold, as illustrated in Figure 1. As a preliminary step to define the tangent space, we first, respectively, introduce the logarithmic mapping and exponential mapping operators as

$$\begin{aligned} \text{Log}_P(\mathbf{P}_j) &= \mathbf{S}_j = \mathbf{P}^{(1/2)} \log(\mathbf{P}^{-(1/2)} \mathbf{P}_j \mathbf{P}^{-(1/2)}) \mathbf{P}^{(1/2)}, \\ \text{Exp}_P(\mathbf{S}_j) &= \mathbf{P}_j = \mathbf{P}^{(1/2)} \exp(\mathbf{P}^{-(1/2)} \mathbf{S}_j \mathbf{P}^{-(1/2)}) \mathbf{P}^{(1/2)}, \end{aligned} \quad (5)$$

where  $\text{Log}_P(\mathbf{P}_j)$  is a mapping from a point on the manifold to the tangent space at  $P$ , while  $\text{Exp}_P(\mathbf{S}_j)$  is a mapping from the tangent space at  $P$  to the manifold. These two operators are both one-to-one mapping operators between the Riemannian manifold and the tangent space.

Because the tangent space  $\mathcal{T}_P(N) = \{\text{Log}_P(\mathbf{P}_j), \mathbf{P}_j \in \mathcal{SPD}(N)\}$  is the space of symmetric matrices, it comprises only  $N(N+1)/2$  independent elements. A minimal representation of the tangent space  $\mathcal{T}_P(N)$  [10] is defined as

$$\mathcal{T}(N) = \{\mathbf{s}_j = \text{upper}(\mathbf{P}^{-(1/2)} \text{Log}_P(\mathbf{P}_j) \mathbf{P}^{-(1/2)}) \in \mathbb{R}^{N(N+1)/2}\}, \quad (6)$$

where the upper  $(\cdot)$  operator is used to retain the upper triangular portion of the symmetric matrix and obtain its vectorization.

The mean of the SPD matrices plays an important role in classification. Since the neighborhood of Riemannian manifold is local homeomorphic to its tangent space, the trinational classifier can be performed on tangent space to obtain high classification performance [14]. However, large neighborhood will lead to large distortion between neighborhood and its tangent space, particularly when the selection of tangent point is not representative. The mean of SPD matrices is the best representation of all data points. It is defined as the point  $\mathbf{P}_m \in \mathcal{SPD}(N)$  with the minimum sum of squared distances to all SPD matrices in a dataset  $\mathcal{C}$ :

$$\mathbf{P}_m = \arg \min_{\mathbf{P} \in \text{calSPD}(N)} \sum_{\mathbf{P}_i \in \mathcal{C}} \delta_R^2(\mathbf{P}, \mathbf{P}_i). \quad (7)$$

**2.3. BCI Data Model.** The recorded EEG signal was a multilag/multichannel signal:

$$\mathbf{X}(t) = [\mathbf{x}(t), \dots, \mathbf{x}(t+L-1)] \in \mathbb{R}^{N \times L}, \quad (8)$$

where  $N$  and  $L$  denote the number of channels and sampled points, respectively, and  $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T \in \mathbb{R}^{N \times 1}$  is the snapshot vector. In motor imagery BCIs, the second-order statistical information (covariance feature) of  $\mathbf{X}(t)$  often provides discriminative information for classification. The covariance feature of  $\mathbf{X}(t)$  is represented by its sample covariance matrix (SCM) [32] as

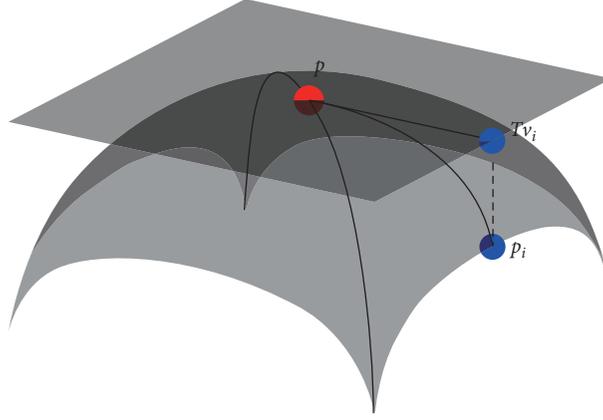


FIGURE 1: Illustration of Riemannian manifold and tangent space.

$$\mathbf{P}(t) = \frac{1}{L-1} \mathbf{X}(t) \mathbf{X}^T(t). \quad (9)$$

### 3. Proposed Bilinear Locality-Preserving Embedding and ELM Classification

**3.1. Riemannian Graph Construction.** In SPD matrix spaces, because the data lie on a Riemannian manifold, it is natural to construct a Riemannian graph  $G_R = (\mathcal{V}, \mathbf{W})$  via Riemannian geodesic distance, as illustrated in Figure 2.  $\mathcal{V}$  represents the set of graph vertices, and  $\mathbf{W}$  represents the graph weight matrix. For a given dataset, the Riemannian graph is constructed as follows.

First, to construct an adjacency graph, a Riemannian distance-based  $k$  nearest neighbor method is used. The points  $P_i, P_j \in \mathcal{V}$  are considered neighbors if  $P_i$  is among the  $k$  nearest neighbors of  $P_j$  or  $P_j$  is among the  $k$  nearest neighbors of  $P_i$ , where  $k$  is a positive integer and  $k$  nearest neighbors are measured by the Riemannian geodesic distance  $d_{ij} = \delta_R(P_i, P_j)$ .

Second, to calculate graph weights, many approaches may be used, such as heat kernels [33], inverse Euclidean distance [34], and local linear reconstruction coefficients [19]. In this study, we use a heat kernel based on the Riemannian geodesic distance to construct graph weights:

$$\mathbf{W}_{ij} \begin{cases} e^{-d_{ij}^2/2\sigma^2}, & \text{if } P_i \text{ and } P_j \text{ are neighbors,} \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where  $\mathbf{W}_{ij}$  is the  $(i, j)$ th element of  $\mathbf{W}$  and the parameter  $\sigma$  controls the width of the neighborhoods.

**3.2. Bilinear Locality-Preserving Embedding and ELM on Tangent Space.** Because the covariance features studied in this paper lie in SPD matrix spaces, we propose a bilinear mapping with  $\mathbf{V} \in \mathbb{R}^{M \times N}$  and  $\mathbf{V}^T$  under the constraint  $\mathbf{V}\mathbf{V}^T = \mathbf{I}_M$ , to construct a system learning a low-dimensional embedding. The low-dimensional embedding  $\mathbf{P}_e = \mathbf{V}\mathbf{P}\mathbf{V}^T \in \mathbb{R}^{M \times M}$  was expected to preserve the locality geometry of a high-dimensional Riemannian manifold. Locality preservation attempts to ensure that if  $P_i$  and  $P_j$  in a

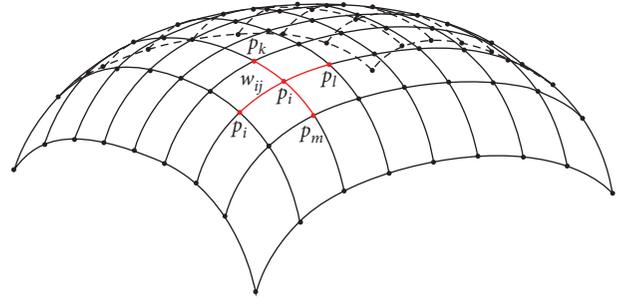


FIGURE 2: Riemannian graph from Riemannian manifold.

high-dimensional manifold are “close,” then  $\mathbf{V}\mathbf{P}_i\mathbf{V}^T$  and  $\mathbf{V}\mathbf{P}_j\mathbf{V}^T$  in low-dimensional embedding are also “close.” A reasonable criterion for locality preservation is to minimize the following objective function:

$$\min \sum_{i,j=1}^{|\mathcal{C}|} \|\mathbf{V}\mathbf{P}_i\mathbf{V}^T - \mathbf{V}\mathbf{P}_j\mathbf{V}^T\|_F^2 \mathbf{W}_{ij}, \quad (11)$$

where the  $F$ -norm  $\|\cdot\|_F$  is used to measure the “closeness” of data points in the embedding.

There is no closed-form optimal solution for the optimization problem (11). It should be noted that the objective function (11) is the trade-off model between optimized solution and Riemannian locality preserving. In this study, we propose an iterative strategy to obtain a local optimal solution of (11). In the proposed strategy, the matrix  $\mathbf{V}$  is first initialized as  $\mathbf{V}_0 = [\mathbf{I}_M, 0]$ , and the new optimal  $\mathbf{V}_{t+1}$  of the  $(t+1)$ th iteration can be obtained by solving the following eigendecomposition problem, given the  $t$ th iteration of  $\mathbf{V}$  and  $\mathbf{V}_t$ :

$$\max -\text{tr}(\mathbf{V}_{t+1} \mathbf{L} \mathbf{V}_{t+1}^T), \quad (12)$$

where  $\mathbf{L} = \sum_{i,j=1}^{|\mathcal{C}|} (\mathbf{P}_i - \mathbf{P}_j) \mathbf{V}_t^T \mathbf{V}_t (\mathbf{P}_i - \mathbf{P}_j)^T \mathbf{W}_{ij}$ . Negating (11), the minimization problem becomes a maximization problem. Because the optimization problem in (12) is a generalized eigenvector problem [35, 36], the optimal transformation matrix  $\mathbf{V}_{t+1}$  is the first  $M$  generalized eigenvectors of  $\mathbf{L}$  corresponding to the  $M$  largest eigenvalues. The iteration is repeated until the solution converges, and a

local optimal solution of (11) is obtained. Further details are provided in Algorithm 1.

After learning the optimized bilinear mapping matrix  $\mathbf{V}_{\text{opt}}$ , the proposed approach maps all points in the embedding into the tangent space according to (6). It has been reported that the tangent space located at the Riemannian mean is the best hyperplane for classification, as it is associated with the minimal distance loss [15]. A three-layer ELM network was constructed on a tangent space for classification because the ELM classifier is more flexible to handle complex and multivariate data than LDA and SVM classifier. The mapping function from the  $j$ th input tangent vector  $s_j$  to the  $i$ th hidden node was defined as

$$h_i(s_j) = g(s_j; \theta) = \frac{1}{1 + \exp(-(\mathbf{a}_i^T s_j + b_i))}, \quad (13)$$

where  $\theta = \{\mathbf{a}_i, b_i\}$  denotes the parameters of the sigmoid mapping function. Based on the principles of ELMs, we randomly chose parameters  $\mathbf{a}_i$  and  $b_i$  within the range  $[-1, 1]$ . We denote by  $\mathbf{h}(s_j) \in \mathbb{R}^{1 \times N_h}$  the output vector of the hidden layer with respect to  $s_j$  and by  $\boldsymbol{\beta} \in \mathbb{R}^{N_h \times N_o}$  the output weights that connect the hidden layer to the output layer, where  $N_h$  and  $N_o$  are the number of hidden neurons and output neurons, respectively. Moreover, the output weights  $\boldsymbol{\beta}$  can be learned by solving an optimization problem [22]:

$$\min \|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2, \quad (14)$$

where  $\mathbf{H} = [\mathbf{h}(s_1)^T, \dots, \mathbf{h}(s_j)^T, \dots, \mathbf{h}(s_{|C|})^T]^T$  and  $\lambda$  parameters are used to balance the loss and  $L_2$  is a regularizer. The parameter  $\mathbf{Y}$  is a matrix of the target output, and the  $j$ th row of  $\mathbf{Y}$  is a binary vector with only one entry (corresponding to the class to which  $s_j$  belongs) equal to one.

The framework of ETS + ELM algorithm is shown in Figure 3; the proposed ETS + ELM algorithm includes a Riemannian graph construction for data representation, locality preserving with bilinear mapping for dimensionality reduction, tangent space mapping for feature extraction, and 3-layer extreme learning machine network for classification.

## 4. Experimental Results

### 4.1. Experimental Setup

**4.1.1. Data Description.** In this study, the BCI Competition 2008 motor imagery data [37] and our in-house experimental data are used for performance evaluation.

- (1) The BCI Competition 2008 IIA EEG motor imagery data were recorded from nine subjects who performed four kinds of motor imagery actions (imagined right hand, left hand, foot, and tongue movements) with 22 EEG channels. For each subject, there were 144 trials per class, including training and testing trials. The EEG signals were filtered by an 8–30 Hz bandpass filter.
- (2) Our in-house EEG data were recorded from 12 subjects with 64 EEG channels. For each subject, there were 234 trials per class. Our in-house

experiments tested right and left hand imagined movements.

The sampling rate of the EEG signal was 250 Hz. The bandpass filter was the same as the Competition dataset.

**4.1.2. Algorithms Evaluated.** We evaluate the proposed algorithms (ETS + ELM) against five competing algorithms.

- (1) CSP + LDA: the CSP in [4] and linear discriminant analysis (LDA) in [38] were applied for evaluation
- (2) CSP + SVM: the number of CSP spatial filters was set to 8, as suggested in [4], for feature extraction, and a support vector machine (SVM) [39] was used for classification
- (3) MDRM: minimum distance to Riemannian mean [14] was used for evaluation
- (4) TS + LDA :LDA was used to classify signals on a tangent space [14]

**4.1.3. Parameter Setting.** In all experiments, we empirically set  $k = 40$  and  $\sigma = 1$  during Riemannian graph construction. In ETS + ELM, the number of iterations was set to  $N_{\text{ite}} = 8$ , and the iterative stop threshold to  $\tau = 0.01$ . In fact, most information of the original space was lost during the process, while the number of embedding dimensions was too small. However, all of the methods evaluated suffer from computational cost problems under conditions of excessively high dimensionality. Hence, the dimension  $M$  was empirically set to 10. Because the ELM method suffers from an overfitting problem in the case of excessively hidden neurons, a regularizer parameter was used to control the overfitting problem. Therefore, the regularizer parameter  $\lambda$  was empirically set to 0.001, while the number of hidden neurons was 2000.

**4.2. Experiments for EEG Classification.** To evaluate the performance of the proposed algorithms, ETS + ELM, some competing algorithms (CSP + LDA, CSP + SVM, MDRM [14], and TS + LDA [14]) were compared on the EEG datasets from the 2008 BCI Competition and our in-house BCI system.

First, we evaluated the convergence performance of the iterative algorithm for optimization problem (12). Because the Kappa coefficient [40] scales linearly from 0 to 1 in the range between random and perfect classification, it was selected as a performance measure [40] for the classification algorithms in this study. In each iteration step of (12), we were able to obtain a learned mapping matrix  $\mathbf{V}_t$  and a corresponding low-dimensional embedding. The performance of the ETS + ELM classification algorithm under the same parameter settings was evaluated on these embeddings. The average performance of nine subjects on the BCI Competition IIA dataset is shown in Figure 4. It is clear that the mapping matrix  $\mathbf{V}_t$  learned from the proposed algorithm converged quickly. Normally, it required only two iteration steps.

**Input:** given SPD matrix set  $C$  with samples  $P_i \in \mathbb{R}^{N \times N}$ ,  $i = 1, \dots, |C|$ , dimensions of embedding  $M$ , iterative number  $N_{ite}$ , stop threshold  $\tau$ , and ELM parameter  $a_i, b_i, N_h$ , and  $N_o$ ;

- (1) Construct a Riemannian graph  $G_R$  over all samples based on Riemannian geodesic distance (Section 3.1);
- (2) Initialize:  $V_0 = [I_M, 0]$ ;
- (3) **For**  $t = 1 : 1:(N_{ite})$ , **do**;  
     Calculate the matrices  $L$  (Section 3.2);  
     Obtain mapping matrix  $V_{t+1} \in \mathbb{R}^{M \times N}$  by (12);  
     **If**  $\|V_t - V_{t-1}\|_F^2 \leq \tau$   
         **break**;  
     **end if**;  
   **end for**;
- (4) Construct  $M \times M$ -dimensional embedding  $P_{ei} = V_{opt} P_i V_{opt}^T$ ,  $i = 1, \dots, |C|$ , where  $V_{opt}$  is convergent matrix in step 4;
- (5) Calculate the Riemannian mean of all data on embedding:  $P_{cm} = \operatorname{argmin}_{P_c} \sum_i \delta_R^2(P_c, P_{ei})$
- (6) Project embedding point onto tangent space on Riemannian mean by (6)  $T(N) = \{s_j\}$ ;
- (7) Apply a three-layer ELM in the tangent space, learning the weight  $\beta$  from (14), using the training data;
- (8) The label of the testing data may be obtained by  $H_t \beta$ , where  $H_t$  is the mapping matrix of the testing data;

**Output:** the label of the testing data.

ALGORITHM 1: Embedding tangent space extreme learning machine (ETS + ELM).

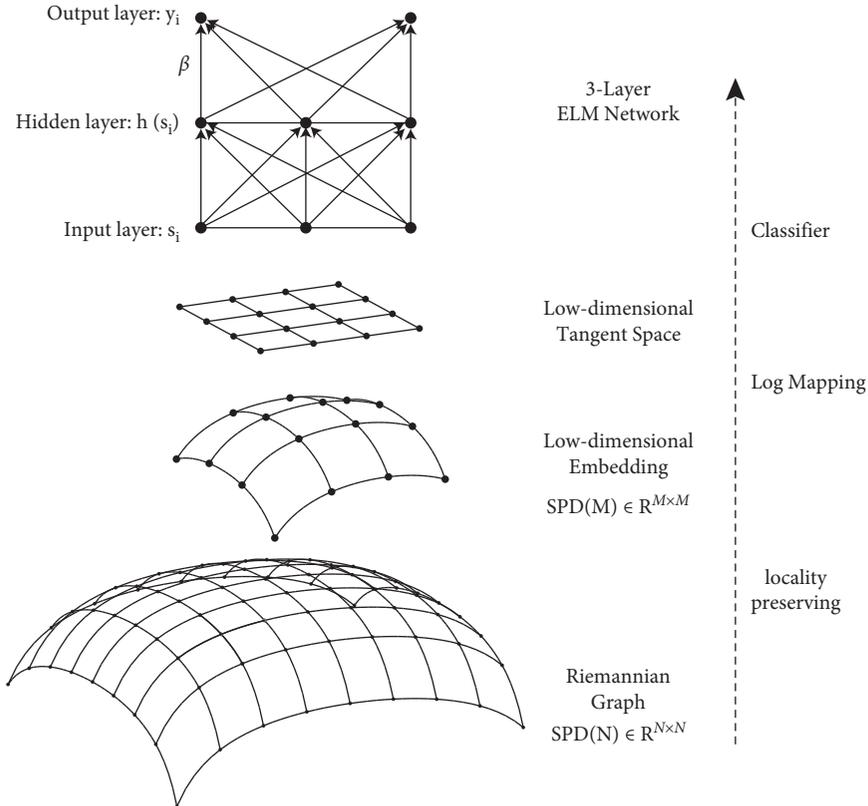


FIGURE 3: The framework of ETS + ELM algorithm.

We also compared the performance of various methods on classification of the BCI Competition 2008 data. In Table 1, the performance of the proposed methods, MDRM, TS+LDA, and the top three Competition winners for dataset Ila of BCI Competition 2008 (1st of BCI08, 2nd of BCI08, and 3rd of BCI08) are presented for comparison. From the results shown in Table 1, it may be observed that the proposed ETS+ELM method achieves a mean

performance of 0.577, which was the best performance of all the methods compared.

We also tested the performance of the proposed algorithms in terms of classification accuracy on the BCI Competition 2008 Ila dataset with a 30-fold cross-validation procedure. The cross-validation reflects the generalization performance of the proposed method. From the results in Table 2, it is clear that, for all subjects, ETS+ELM

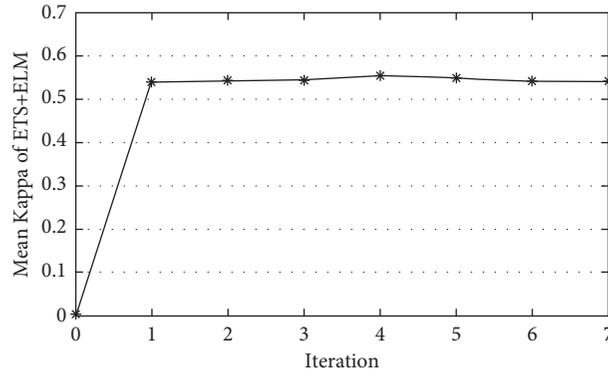


FIGURE 4: Iteration vs. mean Kappa value of ETS + ELM for BCI competition 2008 Iia (the Kappa value is the mean of nine subjects).

TABLE 1: Classification results of BCI competition 2008 Iia in kappa values.

Methods	Mean Kappa	Subjects								
		S01	S02	S03	S04	S05	S06	S07	S08	S09
<b>ETS + ELM</b>	<b>0.577</b>	0.74	0.27	0.79	0.52	0.31	0.31	0.69	0.77	0.79
TS + ELM	0.573	0.73	0.32	0.77	0.51	0.32	0.31	0.67	0.76	0.77
1st of BCI08	0.570	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61
TS + LDA [14]	0.567	0.74	0.38	0.72	0.50	0.26	0.34	0.69	0.71	0.76
MDRM [14]	0.521	0.75	0.37	0.66	0.53	0.29	0.27	0.56	0.58	0.68
2nd of BCI08	0.520	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69
3rd of BCI08	0.310	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44

outperformed the reference methods. This indicates that, by finding a low-dimensional embedding with the property of locality preservation, the classification method ETS + ELM exhibited the best performance.

Moreover, we evaluated the performance of all algorithms on the dataset recorded from our in-house BCI systems. As shown in Table 3, the proposed ETS + ELM showed a significant performance improvement over the prior methods (an approximately 2% improvement in classification accuracy), even when compared with the benchmark algorithm CSP [4]. In conclusion, the experimental results on the competition dataset and in-house dataset demonstrate that the proposed ETS + ELM approached achieved a higher performance than the other methods. In additional, we provide the ANOVA and  $t$ -test analysis of Tables 1–3. The repeated measures ANOVA results for Tables 1–3 are  $p = 0.2 \times 10^{-10}$ ,  $0.7 \times 10^{-7}$ , and 0.03, respectively. It indicates that all studied methods have statistically significant difference on classification performance. The  $t$ -test analysis of Tables 1–3 are shown in Table 4.

We also compared the computational loads of the proposed algorithm with the competing methods. In the BCI competition dataset, the training time of proposed algorithm is 24.7 s, which is far longer than the training time of MDRM (0.12 s), TSLDA (1.17 s), CSPLDA (0.18 s),

and CSPSVM (0.19 s). It is because that the proposed algorithm is a semisupervised learning model where the training and test dataset are simultaneously used for learning features, but the competing methods only used training dataset to learn feature. Other reason is that the proposed algorithm takes a lot of time in iterative strategy of optimization (equation (11)), but the competing methods solve the optimization problem by quick eigenvalue decomposition. However, the test time of proposed algorithm is 0.06 s, which is close to the test time of MDRM (0.17 s), TSLDA (0.03 s), CSPLDA (0.04 s), and CSPSVM (0.03 s). The computational loads on in-house dataset are similar with those of competition dataset.

Finally, we compared our bilinear locality-preserving embedding with other state-of-the-art manifold learning algorithms. We show the 2-dimensional embedding learning from ISOMAP, LLE, and the proposed algorithm in Figure 5. The motor imagery EEG data are a spatio-temporal signal, and the correlations of EEG signals in high-dimensional space are expected to be destroyed in 2-dimensional embedding. However, compared with feature distributions extracted by other manifold learning methods, the proposed method can also obtain a moderately good separability of classes in the feature space.

TABLE 2: Classification accuracy of BCI competition 2008 Ila in 30-fold cross-validation (%).

Subject	Methods					
	ETS + ELM	TS + ELM	TS + LDA [14]	MDRM [14]	CSP + LDA [35]	CSP + SVM [36]
S01	<b>81.5</b>	80.5	80.5	77.8	78.3	76.3
S02	<b>60.3</b>	58.0	51.3	44.1	44.7	50.7
S03	<b>88.1</b>	86.8	87.5	76.8	82.2	85.1
S04	<b>64.1</b>	60.0	59.3	54.9	59.1	52.9
S05	<b>55.4</b>	59.3	45	43.8	39.7	48.8
S06	<b>57.1</b>	53.2	55.3	47.1	50.1	49.2
S07	<b>82.8</b>	80.7	82.1	72	81	78.1
S08	<b>85.0</b>	80.7	84.8	75.2	68.5	77.4
S09	<b>87.3</b>	87.4	86.1	76.6	77.4	82.2
Mean $\pm$ std	<b>73.5 <math>\pm</math> 13.2</b>	71.8 $\pm$ 13.8	70.2 $\pm$ 17.1	63.2 $\pm$ 15.2	64.6 $\pm$ 16.6	66.7 $\pm$ 15.7

TABLE 3: Classification accuracy on in-house BCI dataset (%).

Subject	Methods					
	ETS + ELM	TS + ELM	TS + LDA [14]	MDRM [14]	CSP + LDA [45]	CSP + SVM [36]
A01	<b>100</b>	100	100	100	98.2	96.4
A02	<b>98.4</b>	93.7	97.4	93.5	98.7	96.1
A03	<b>96.6</b>	89.0	93.5	96.7	87.0	83.8
A04	<b>91.0</b>	85.0	91.2	84.2	89.4	82.4
A05	<b>87.1</b>	80.1	84.2	80.7	78.9	68.4
A06	<b>90.5</b>	91.2	73.2	67.8	82.1	75.0
A07	<b>73.2</b>	71.2	66.0	67.8	67.8	66.0
A08	<b>76.3</b>	71.0	78.9	71.9	70.1	<b>73.6</b>
A09	<b>82.2</b>	75.5	72.6	67.5	79.4	79.4
A10	<b>94.4</b>	97.7	89.6	90.9	93.5	96.1
A11	<b>100</b>	100	100	100	100	100
A12	<b>72.9</b>	85.3	88.7	85.4	77.4	74.1
Mean $\pm$ std	<b>88.4 <math>\pm</math> 10.3</b>	86.6 $\pm$ 10.4	86.2 $\pm$ 11.3	83.8 $\pm$ 12.6	85.2 $\pm$ 10.5	82.6 $\pm$ 11.3

TABLE 4: The  $t$ -test results of Tables 1–3.

	Table 1	Table 2	Table 3
ETS + ELM vs. TS + LDA	~	*	~
ETS + ELM vs. MDRM	~	**	~
ETS + ELM vs. CSP + LDA	—	**	*
ETS + ELM vs. CSP + SVM	—	**	*
ETS + ELM vs. 1st winner	~	—	—
ETS + ELM vs. 2nd winner	*	—	—
ETS + ELM vs. 3rd winner	**	—	—

Note. ~, nonsignificant; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; —, nonexistent.

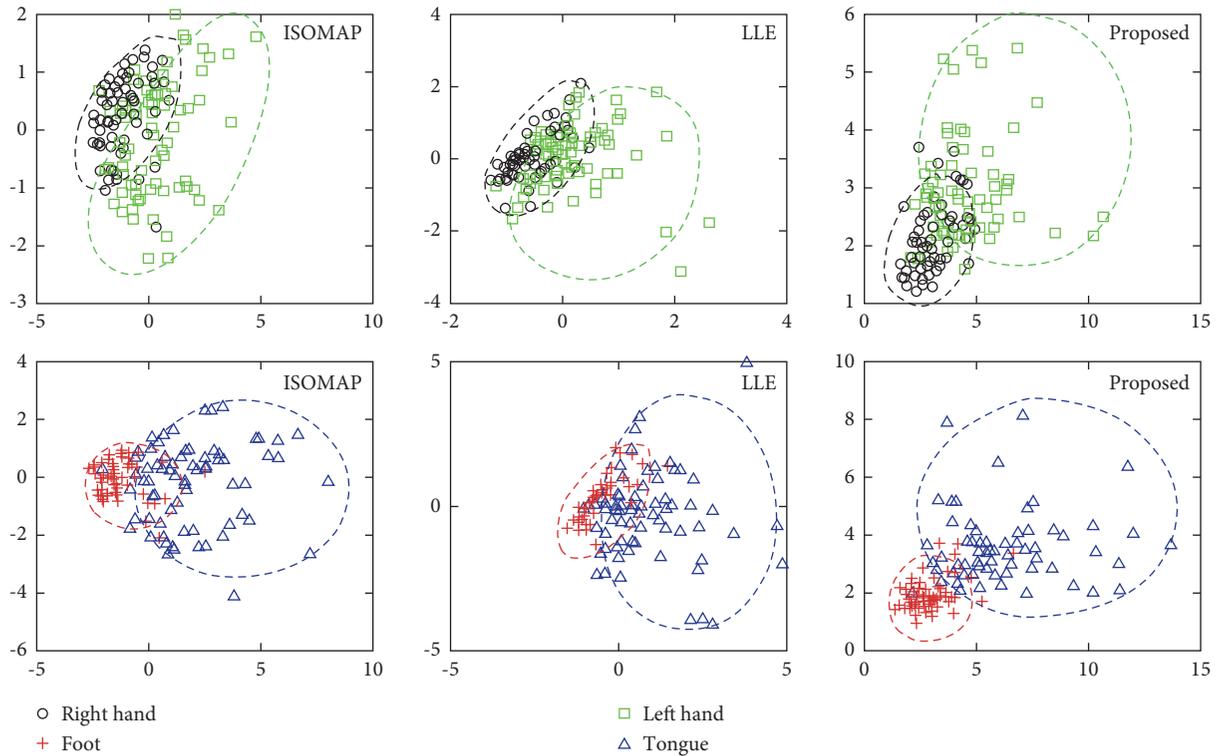


FIGURE 5: 2-dimensional embedding learning from ISOMAP, LLE, and proposed algorithm for subject 1 of BCI competition IV Ila. Right hand vs. left hand show in the first row, and foot vs. tongue show in the second row.

## 5. Conclusions

Inspired by the fact that the covariance feature of EEG signals in the form of SPD matrices lies on a Riemannian manifold, in this study, we have presented an approach to construct a Riemannian graph for motor imagery classification. A bilinear locality preserving method has been proposed to find low-dimensional embeddings carrying discriminant information. A classification algorithm, called ETS + ELM, was further proposed on a tangent space of the learned embedding. Experimental results on BCI competition and our in-house datasets show that the proposed bilinear locality-preserving embedding was able to capture the useful locality and discriminant information of the original manifold, and the proposed classification algorithms have higher performance than the other competing algorithms compared. The proposed methods can also be applied to many other pattern-recognition problems with input data in the form of SPD matrices, such as pedestrian detection problem [15] or hyperspectral image processing [41]. In future, we will further study weight calculation method to construct preferable Riemannian graph and study the interpretation of the feature learned by Riemannian-based methods.

## Data Availability

Previously reported BCI competition EEG data used to support this study are available at <https://www.bbci.de/competition/iv/#dataset2a>. These prior studies (and

datasets) are cited at relevant places within the article as references. The in-house EEG data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

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

## Acknowledgments

This work was supported in part by the Natural Science Foundation of Hainan, under Grant 2019RC165, and the Young Talents Science and Technology Innovation Project of Hainan Association for Science and Technology, under Grant QCXM202011.

## References

- [1] A. Akrami, S. Solhjo, A. Motie-Nasrabadi, and M.-R. Hashemi-Golpayegani, "EEG-based mental task classification: linear and nonlinear classification of movement imagery," in *Proceedings of the IEEE 27th International Conference on Engineering in Medicine and Biology Society*, pp. 4626–4629, Shanghai, China, January 2006.
- [2] Y. Yuanqing Li, J. Jiahui Pan, F. Fei Wang, and Z. Zhuliang Yu, "A hybrid bci system combining p300 and ssvep and its application to wheelchair control," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 11, pp. 3156–3166, 2013.

- [3] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, 1997.
- [4] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.
- [5] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Networks*, vol. 118, pp. 262–270, 2019.
- [6] J. Jin, L. Shurui, I. Daly et al., "The study of generic model set for reducing calibration time in P300-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 3–12, 2019.
- [7] Y. Jiao, T. Zhou, and L. Yao, "Multi-view multi-scale optimization of feature representation for EEG classification improvement," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 99, 2020.
- [8] K. Wang, M. Xu, and Y. Wang, "Enhance decoding of pre-movement EEG patterns for brain-computer interfaces," *Journal of Neural Engineering*, vol. 17, no. 1, p. 11, Article ID 016033, 2020.
- [9] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and dempster-shafer theory," *IEEE Transactions on Neural Networks and Learning Systems*, no. 99, pp. 1–12, 2020.
- [10] O. Chapelle and B. S. A. Zien, *Semi-supervised Learning*, MIT Press Cambridge, Cambridge, MA, USA, 2006pp. 1–13, Chapter 1.
- [11] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with 11-graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.
- [12] S. Yan and H. Wang, "Semi-supervised learning by sparse representation," in *Proceedings of the SIAM International Conference on Data Mining*, pp. 792–801, Sparks, NV, USA, April 2009.
- [13] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," *Latent Variable Analysis and Signal Separation*, in *Proceedings of the International Conference on Latent Variable Analysis and Signal Separation*, pp. 629–636, St. Malo, France, September 2010.
- [14] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.
- [15] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.
- [16] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1285–1294, 2008.
- [17] J. Long, Y. Li, and Z. Yu, "A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces," *Cognitive Neurodynamics*, vol. 4, no. 3, pp. 207–216, 2010.
- [18] Y. Li and C. Guan, "An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces," *Neural Computation*, vol. 18, no. 11, pp. 2730–2761, 2006.
- [19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [20] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160, 2004.
- [21] X. Xiaofei He, S. Shuicheng Yan, Y. Niyogi, and P. Hong-Jiang Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [23] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [24] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–18, 2020.
- [25] G. Shi, F. Luo, Y. Tang, and Y. Li, "Dimensionality reduction of hyperspectral image based on local constrained manifold structure collaborative preserving embedding," *Remote Sensing*, vol. 13, 2021.
- [26] Z. Jin, G. Zhou, and D. Gao, "EEG classification using sparse Bayesian extreme learning machine for brain-computer interface," *Neural Computing & Applications*, vol. 32, no. 6, 2020.
- [27] Y. Zhang, Y. Wang, G. Zhou et al., "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," *Expert Systems with Applications*, vol. 96, pp. 302–310, 2018.
- [28] Y. Zhang, T. Zhou, and W. Wu, "Improving EEG decoding via clustering-based multitask feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, no. 99, , 2021.
- [29] J. M. Lee, "Smooth manifolds," *Introduction to Smooth Manifolds, Chapter 5*, pp. 93–105, Springer, Berlin, Germany, 2012.
- [30] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*, pp. 299–309, Springer, Berlin, Germany, 2003.
- [31] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005.
- [32] S. T. Smith, "Covariance, subspace, and intrinsic Cram er/Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [33] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [34] C. Cortes and M. Mohri, "On transductive regression," *Advances in Neural Information Processing Systems*, vol. 19, pp. 305–312, 2007.
- [35] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," *Advances in Neural Information Processing Systems*, vol. 18, pp. 499–506, 2006.
- [36] Y. Liu, Y. Liu, and K. C. Chan, "Multilinear isometric embedding for visual pattern analysis," in *Proceedings of the IEEE 12th International Conference on Computer Vision Workshops*, pp. 212–218, Kyoto, Japan, October 2009.

- [37] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, *BCI Competition 2008-Graz Data Set a*, Graz University of Technology, Graz, Austria, 2008.
- [38] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [39] H. Sun, Y. Xiang, Y. Sun, H. Zhu, and J. Zeng, "On-line EEG classification for brain-computer interface based on CSP and SVM," in *Proceedings of the IEEE 3rd International Congress on Image and Signal Processing*, vol. 9, pp. 4105–4108, Yantai, China, October 2010.
- [40] G. Townsend, B. Graimann, and G. Pfurtscheller, "A comparison of common spatial patterns with complex band power features in a four-class BCI experiment," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 642–651, 2006.
- [41] F. Cai, T. Wang, and J. Wu, "Handheld four-dimensional optical sensor," *Optik*, vol. 203, Article ID 164001, 2020.