

## Research Article

# Motor Imagery EEG Classification Based on Decision Tree Framework and Riemannian Geometry

Shan Guan , Kai Zhao , and Shuning Yang 

School of Mechanical Engineering, Northeast Electric Power University, 132012 Jilin, China

Correspondence should be addressed to Shan Guan; [guanshan1970@163.com](mailto:guanshan1970@163.com)

Received 24 September 2018; Revised 20 December 2018; Accepted 3 January 2019; Published 21 January 2019

Guest Editor: Anastassia Angelopoulou

Copyright © 2019 Shan Guan 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.

This paper proposes a novel classification framework and a novel data reduction method to distinguish multiclass motor imagery (MI) electroencephalography (EEG) for brain computer interface (BCI) based on the manifold of covariance matrices in a Riemannian perspective. For method 1, a subject-specific decision tree (SSDT) framework with filter geodesic minimum distance to Riemannian mean (FGMDRM) is designed to identify MI tasks and reduce the classification error in the nonseparable region of FGMDRM. Method 2 includes a feature extraction algorithm and a classification algorithm. The feature extraction algorithm combines semisupervised joint mutual information (*semi*-JMI) with general discriminate analysis (GDA), namely, SJGDA, to reduce the dimension of vectors in the Riemannian tangent plane. And the classification algorithm replaces the FGMDRM in method 1 with k-nearest neighbor (KNN), named SSDT-KNN. By applying method 2 on BCI competition IV dataset 2a, the kappa value has been improved from 0.57 to 0.607 compared to the winner of dataset 2a. And method 2 also obtains high recognition rate on the other two datasets.

## 1. Introduction

Brain computer interface (BCI) based on motor imagery (MI) is used to analyze human intention by electroencephalogram (EEG) signals generated by human brain electrophysiological activity [1, 2]. Based on BCI technology, exoskeletons can be used to help people with physical disabilities regain their motor ability, and BCI also has wide applications in smart home, entertainment, military, and other fields [3–6].

Common spatial pattern (CSP) is widely used in motor imagery to extract EEG features [7]. CSP has excellent performance in two classification tasks, but the drawback is that it needs a lot of electrodes [8].

Despite its short history, the use of the Riemannian geometry in BCI decoding is currently attracting increasing attention [9–13]. Covariance matrices lie in the space of symmetric positive definite (SPD) matrices, which can be formulated as a Riemannian manifold [14]. In the BCI field, the connections of the CSP algorithm and the tools of information geometry have been investigated, considering

several divergence functions in alternative to the Riemannian distance [15–18]. Barachant et al. proposed a simple data augmentation approach for improving the performance of the Riemannian mean distance to mean (MDM) algorithm [13]. Kumar et al. propose a single band CSP framework for MI-BCI that utilizes the concept of tangent space mapping in the manifold of covariance matrices, and the proposed method obtains good results when compared to other competing methods [19]. A hierarchical MDM classifier for multiclass problem has been tested in [20].

Advanced classifiers based on the tangent space on the Riemannian manifold of positive matrices are also receiving increasing attention. Barachant et al. map the covariance matrices in the tangent space and apply feature selection and linear discriminate analysis (LDA) in the tangent space [10]. For the application of the classifier in the tangent space, the problem is that the curse of dimensionality. Traditional data dimensionality reduction methods include two categories: linear dimensionality reduction (LDR) and nonlinear dimensionality reduction (NLDR). Since most of the actual

data are nonlinear, NLDR techniques such as locally linear embedding (LLE) [21], isometric mapping (ISOMAP) [22], maximum variance unfolding (MVU) [23], and t-distributed stochastic neighbor embedding (t-SNE) [24, 25] are used to tackle problems widely. Lee et al. used discrete wavelet transform (DWT) and continuous wavelet transform (CWT) to extract features of MI tasks, and Gaussian mixture model (GMM) was used to construct GMM supervectors; this method accelerates the speed of training and improves the accuracy of motor imagery [26]. Sadatnejad et al. propose a new kernel to preserve the topology of data points in the feature space, and the proposed kernel is strong, particularly in the cases where data points have a complex and nonlinear separable distribution [8]. Xie et al. proposed a framework for intrinsic submanifold learning from a high-dimensional Riemannian manifold; the proposed method exhibited strong robustness against a small training dataset [27].

There is still another approach for overcoming the problem of high dimensionality in SPD manifolds. And this method maps from a high-dimensional SPD manifold to a lower dimensional one while the geometry of SPD manifolds is preserved. And there are only two works of this way. Davoudi et al. [14] proposed distance preservation to local mean (DPLM) as dimensionality reduction technique, combined with FGMDM, the best performance of this article in terms of kappa value is 0.60. Harandi et al. [28] learned a mapping that maximizes the geodesic distances between interclass and simultaneously minimizes the distances between intraclass, and it is done via an optimization on Grassmann manifolds.

In this paper, we proposed a novel SSDT-FGMDRM and SSDT-KNN for the classification of multiclass MI tasks by designing a simple yet efficient subject-specific decision tree framework. Method 1 contains SSDT-FGMDRM to improve the performance of FGMDRM. For each individual, method 1 first separates the two most discriminative classes from the group. Furthermore, the remaining categories including the misclassification samples of the previous nodes are reclassified in the last node. Method 2 contains SSDT-KNN and a NLDR method named SJGDA. SJGDA combines the advantage of *semi*-JMI and GDA, and method 2 performed well on different datasets. The aims of this article are as follows:

- (1) To verify the effectiveness of the proposed SSDT framework through dataset 1
- (2) To verify the superiority of SJGDA in feature extraction, compared with *semi*-JMI and GDA
- (3) To validate the generalization ability of method 2 through different datasets, in this paper

The rest of the paper is organized as follows: Section 2 introduced the mathematical preliminaries of the Riemannian geometry. Section 3 discussed the proposed methods in detail. Three datasets are introduced in Section 4. The results of our work are discussed in Section 5. And in Section 6, we compared our methods with the state of the art. This paper concludes in Section 7.

## 2. Geometry of SPD Matrices

Let  $X_i$  represent a short segment of continuous EEG signals, and  $X_i$  can be denoted as follows:

$$X_i = [X_{t+T_i} \cdots X_{t+T_i+T_s-1}] \in \mathbb{R}^{n \times T_s}, \quad (1)$$

where  $X_i$  corresponds to the  $i$ th trail of imaged movement starting at time  $t = T_i$ .  $T_s$  denotes the number of sampled points of the selected segment.

For the  $i$ th trail, the spatial covariance matrix (SCM)  $P_i \in \mathbb{R}^{n \times n}$  can be calculated as follows:

$$P_i = \frac{1}{T_s - 1} X_i X_i^T. \quad (2)$$

Based on the SCM, there are two ways to classify MI tasks in the Riemannian manifold.

**2.1. Filter Geodesic Minimum Distance to the Riemannian Mean.** The Riemannian distance between two SPD matrices  $P_1$  and  $P_2$  in  $P(n)$  is given by [29]

$$\delta_R(P_1, P_2) = \left\| \log(P_1^{-1} P_2) \right\|_F = \left[ \sum_{i=1}^n \log^2 \lambda_i \right]^{1/2}. \quad (3)$$

Given  $m$  SPD matrices  $P_1, \dots, P_m$ , the geometric mean in the Riemannian sense is defined as

$$\mathfrak{M}(P_1, \dots, P_m) = \arg \min_{P \in P(n)} \sum_{i=1}^m \delta_R^2(P, P_i). \quad (4)$$

For algorithm mean Riemannian distance to Riemannian mean (MDRM), we compute the Riemannian distance between unknown class  $P$  to the Riemannian mean point of each class and classify the unknown class into categories corresponding to the shortest distance. Inspired by the principal geodesics analysis (PGA) method [30], the literature [31] finds a set of filters by applying an extension of Fisher linear discriminant analysis (FLDA) named Fisher geodesic discriminant analysis (FGDA). And then, apply these filters to MDRM to form filter geodesic minimum distance to Riemannian mean (FGMDRM). More details can be seen from [31].

**2.2. Tangent Space Mapping.** As shown in Figure 1, the SPD matrix of  $P$  is denoted by a differentiable Riemannian manifold  $Z$ . Each tangent vector  $S_i$  can be seen as the derivative at  $t=0$  of the geodesic  $\Gamma(t)$  between  $P$  and the exponential mapping  $P_i = \text{EXP}_P(S_i)$ , defined as follows:

$$\text{Exp}_P(S_i) = P_i = P^{1/2} \exp(P^{-1/2} S_i P^{-1/2}) P^{1/2}. \quad (5)$$

The inverse mapping is given by the logarithmic mapping and can be defined as follows:

$$\log_P(P_i) = S_i = P^{1/2} \log(P^{-1/2} P_i P^{-1/2}) P^{1/2}. \quad (6)$$

Using the Riemannian geodesic distance, the Riemannian mean of  $I > 1$  SPD matrices by

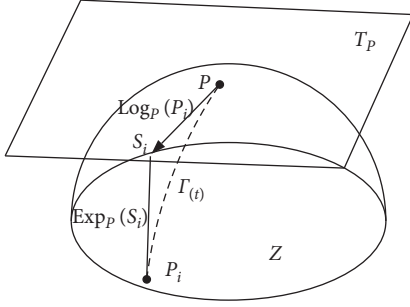


FIGURE 1: The tangent space at point  $P$ , and the geodesic  $\Gamma(t)$  between  $P$  and  $P_i$ .

$$\mathfrak{P}(P_1, \dots, P_I) = \arg \min_{P \in P(n)} \sum_{i=1}^I \delta_R^2(P, P_i). \quad (7)$$

Using the tangent space located at the geometric mean of the whole set trials,  $P_{\mathfrak{P}} = \mathfrak{P}(P_i, i = 1 \dots I)$ , and then, each SCM  $P_i$  is mapped into this tangent space, to yield the set of  $m = n(n+1)/2$  dimensional vectors:

$$S_i = \text{upper}(P_{\mathfrak{P}}^{-1/2} \log(P_i) P_{\mathfrak{P}}^{-1/2}). \quad (8)$$

Many efficient classification algorithms can be implemented in the Riemannian space [10].

### 3. Methods

**3.1. Subject-Specific Decision Tree Framework.** Decision tree is a common machine learning method. Each node of decision tree can be defined as a rule. Guo and Gelfand [32] proposed classification trees with neural network, and this method embeds multilayer neural networks directly in nodes. In the decision tree, one of the most important things is to construct a proper binary tree structure; the upper nodes have the greater impact of the accuracy of the whole samples [33]. In order to solve the multiclassification problem in this paper, we constructed a subject-specific decision tree (SSDT) classification framework as shown in Figure 2 according to the best separating principle [34]. As can be seen from Figure 2, the SSDT proposed in this paper trains a different classification model at different nodes of the decision tree.

The advantages of the SSDT framework are as follows:

- (1) This model separates the two MI tasks (e.g., C.1 and C.2) with the highest recognition rate as far as possible
- (2) At the last node, we reclassify some samples to enhance the classification ability of the classifier

**3.2. Method 1: A Direct Classification Method Based on SSDT-FGMDRM.** Firstly, we point out one problem of the multi-class FGMDRM by using an example. Figure 3 gives a three-class classification problem. Figure 3(a) shows the classification progress by FGMDRM. We can see that three Riemannian mean points (RMPs) are located on the manifold. Since the classification criterion is decided by the distance calculated

between the test point and the RMP, it caused a wrong classification. Figure 3(b) shows the example of the classification results obtained by using the first node of the SSDT-FGMDRM framework. It can be seen that the error classification is corrected by using the decision tree framework.

Method 1 is used to classify four types of MI tasks directly. The training and testing diagram is shown in Figure 4.

**3.3. Feature Extraction Algorithm Based on the Riemannian Tangent Space.** In this paragraph, we propose a novel data reduction method which combines *semi*-JMI and GDA, namely SJGDA, to solve the dimension disaster problem after tangent space mapping.

#### 3.3.1. Semisupervised Joint Mutual Information.

Semisupervised dataset  $D = D\{D_L \cup D_U\}$  consists of two parts,  $D_L = \{x^i, y^i\}_{i=1}^{N_L}$  are labelled data and  $D_U = \{x^{N_L+i}\}_{i=1}^{N_U}$  are unlabelled data. A binary random variable  $S$  is introduced to determine the distribution of labelled dataset and unlabelled dataset. When  $s = 1$ , we record the value of  $y$ , otherwise not. In this way, the labelled set  $D_L$  comes from the joint distribution  $p(x, y|s=1)$ , while the unlabelled set  $D_U$  comes from the distribution  $p(x|s=0)$ . The underlying mechanism  $S$  turns out to be very important for feature selection.

Feature selection method based on mutual information theory is a common feature selection method [35]. In these methods, we rank the features according to the score and select the features with higher scores. For example, by ranking the features according to their mutual information with the labels, we get the sort of correlation that is related to class labels. The characteristics of the score are defined as follows:

$$J_{\text{JMI}}(X_k) = \sum_{X_j \in X_\theta} \hat{I}\left(X_k; \frac{Y}{X_j}\right), \quad (9)$$

where  $X_\theta$  represents the set of the features already selected and  $X_k$  is the feature ranked by scores.  $Y$  represents the label corresponding to feature  $X_k$ .

*Semi*-JMI is a method of using a semisupervised dataset as a training set for JMI. More details can be seen from Reference [36]. In this paper, the missingness mechanism is class-prior-change semisupervised scenario (MAR-C) [37]. After feature ranking, we can obtain a feature vector as follows:

$$f = [f_1, f_2, \dots, f_n], \quad (10)$$

where  $n$  is the length of the tangent vectors  $S_i$ . Since information redundancy exists in  $f$ , we select the best vector length  $m$  ( $m < n$ ) of each subject by the classification recognition rate:

$$f_{\text{SJ}} = [f_1, f_2, \dots, f_m]. \quad (11)$$

**3.3.2. Generalized Discriminant Analysis.** After variable selection, this paper uses generalized discriminant analysis

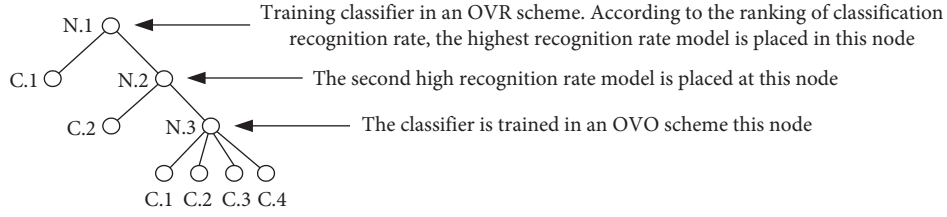


FIGURE 2: SSDT based on the best separating principle for four class.  $N. i$  represents node  $i$ , and  $C. i$  represents class  $i$ .

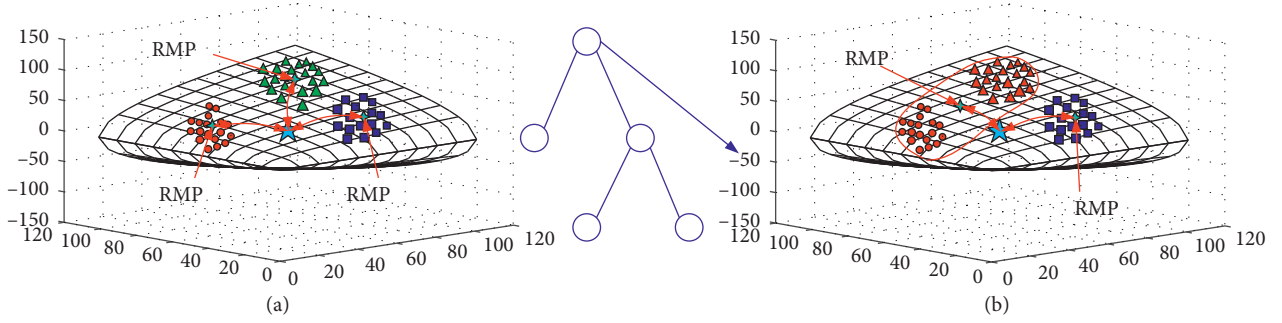


FIGURE 3: Three classification problems classified by FGMDRM (a); and a subjectspecific decision tree FGMDRM model (b).

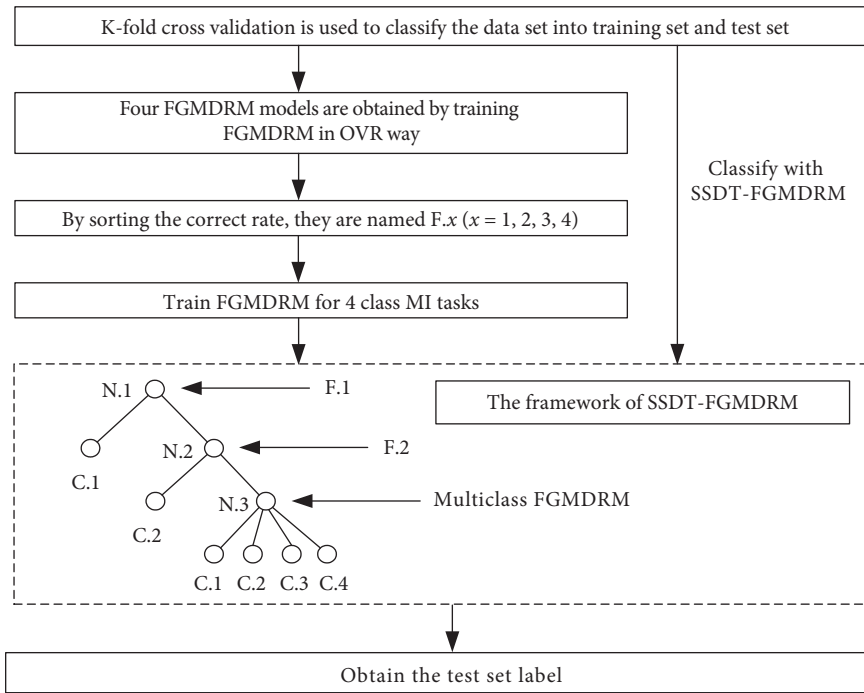


FIGURE 4: Block diagram for method 1.

(GDA) [38, 39], which is a nonlinear feature reduction technique based on kernels to reduce the length of the feature vectors  $f_{Sj}$  and their redundancies. Mapping  $X$  ( $f_{Sj}$ ) into a high-dimensional space  $F$  through a kernel function  $\Phi$ :

$$\begin{aligned} \Phi: R^d &\longrightarrow F, \\ x^\Phi &\longrightarrow (x). \end{aligned} \quad (12)$$

The linear Fisher decision is performed in the  $F$  space, and the criterion function for its extension is

$$J(W^\Phi) = \arg \max_{W^\Phi} \left( \frac{|(W^\Phi)^T S_B^\Phi W^\Phi|}{|(W^\Phi)^T S_T^\Phi W^\Phi|} \right), \quad (13)$$

where  $W^\Phi \in F$  and  $S_B$  and  $S_W$  are between-class scatter and within-class scatter, respectively.



For the convenience of the numerical calculation, kernel functions are introduced to solve the problem:

$$k(x, y) = (\Phi(x) \cdot \Phi(y)). \quad (14)$$

Gauss kernel, poly kernel, and sigmoid kernel are widely used in GDA [40]. For test data  $z$ , its image  $\Phi(z)$  in  $F$  space projects on  $W^\Phi$  is as follows:

$$(W_i^\Phi \cdot \Phi(z)) = \sum_{j=1}^N \alpha_{ij} (\Phi(x_j) \cdot \Phi(z)) = \sum_{j=1}^N \alpha_{ij} k(x_j, z). \quad (15)$$

This paper uses ploy kernel to reduce the dimension. After GDA, we can get a vector  $f_G$  as follows:

$$f_G = [f_1, f_2, \dots, f_d], \quad (16)$$

where  $d$  of  $f_G$  is decided by the actual needs, and in this paper we set  $d = 1$ . And then, SJGDA is applied to the dataset of this paper, and the final feature vectors are constructed as follows:

$$f_{\text{SJGDA}} = [f_G, f_{\text{SJ}}]. \quad (17)$$

**3.4. Method 2: SJGDA and Subject-Specific Decision Tree  $k$ -Nearest Neighbor.** Method 2 is used to classify four types of MI tasks after tangent space mapping. The training and testing diagram is shown in Figure 5.

## 4. Description of Data

**4.1. Dataset 1.** BCI competition IV dataset 2a is used to evaluate the performance of the proposed two methods [41]. Dataset 2a collects 22 channel EEG data and 3 EOG channel data. Four types of motor imagery were collected: left hand, right hand, foot, and tongue. The dataset contains nine healthy subjects and each subject has two sessions, one training session and one test session. Each session has 288 trails of MI data with 72 trails for each MI task. The EEG signals are bandpass filtered by a 5-th order Butterworth filter in the 8–30 Hz frequency band. The selection of trial period is important in MI classification; we select 2 s data (0.5 s and 2.5 s) after the cue, instructing the user to perform the MI tasks by the winner of the competition.

**4.2. Dataset 2.** BCI competition III dataset IIIa is used to evaluate the performance of method 2. BCI III dataset IIIa contains 3 subjects: K3b, K6b, and L1b, and collects 64 channel EEG data. The EEG was sampled with 250 Hz. Four types of motor imagery were collected: left hand, right hand, foot, and tongue. More details about this dataset can be seen at Reference [42].

**4.3. Dataset 3.** In our own dataset, Emotiv Epoc+ is used to collect EEG data of motor imagery. It is a portable EEG acquisition device with a sampling rate of 128 Hz. It has fourteen electrode channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4), two inference electrodes (CMS and

DRL), and the electrode placement follows the international 10–20 standard. Equipment and the Emotiv 14 electrodes are located over 10–20 international system positions as shown in Figure 6. This experiment collected three kinds of EEG signals of one joint: imagination of shoulder flexion (F), extension (E), and abduction (A), as shown in Figure 7.

Seven subjects participated in this experimental study. These subjects were in good health. During the experiment, subjects were naturally placed with both hands, trying to avoid body or head movement. During the experiment, subjects carried out motor imagery under the outside cue, a single experiment collected EEG signal for 5 seconds, and then took 5–7 seconds to have rest, each action repeated acquisition 20 times. The experimental process is shown in Figure 8.

## 5. Results

**5.1. Results of Method 1.** We use SSDDT-FGMDRM to classify multiclass MI tasks as introduced in Section 3.1. Since there are four classes, we can have four pairs of MI tasks: left vs rest (L/RE), right vs rest (R/RE), foot vs rest (F/RE), and tongue vs rest (T/RE). For each subject, the pair with the highest accuracy is used to train N.1, and the pair with the second highest accuracy is to train N.2. Table 1 gives the ten-folder cross-validation results obtained using FGMDRM in OVR scheme.

Table 2 displays the kappa values obtained by method 1. Compared with other methods, five subjects (A03, A06, A07, A08, and A09) achieved higher kappa value of nine without exploring the frequency domain information by method 1. In the case of fixed frequency window, we have improved the mean kappa value of 0.069 than MDRM ( $p = 0.4683$ ), and 0.139 than FGMDM\_fixed ( $p = 0.1423$ ). Our approach also shows significant improvement than FGMDM ( $p = 0.6607$ ), which has exploited subject-specific frequency information, in terms of the kappa value of 0.039.

**5.2. Results of Method 2.** The results in Figure 9 show the T/RE feature distribution of the five features of subject A09. Figure 9(a) shows the first five ranked features with *semi*-JMI. After applying the *semi*-JMI, the first five best features extracted have shown statistically significant improvement in the separability with  $p$  values  $< 0.05$  except feature 2 with  $p$  value 0.77. In Figure 9(b), the first five features extracted from primitive feature vectors with  $p$  value of 0.13, 0.05, 0.87, 0.05, and 0.13. The  $p$  values indicate that the pair T/RE have no significance in the primitive feature vectors. The results show that with our semisupervised feature ranking algorithm, the separable degree of the feature has been greatly improved.

Figure 10 shows the evolution of the classification accuracy with KNN ( $k = 5$  in this paper) against the number of ranked variables in OVR scheme. L/RE and T/RE are the two pairs with the highest recognition rate, and they achieved the highest recognition rate in 100 variables. But this is still a curse of dimensionality for classifiers; GDA is used to analyze the first 100 sorted variables in our study.

As the separation of characteristics cannot meet our requirements, GDA is used to get more obvious variables.

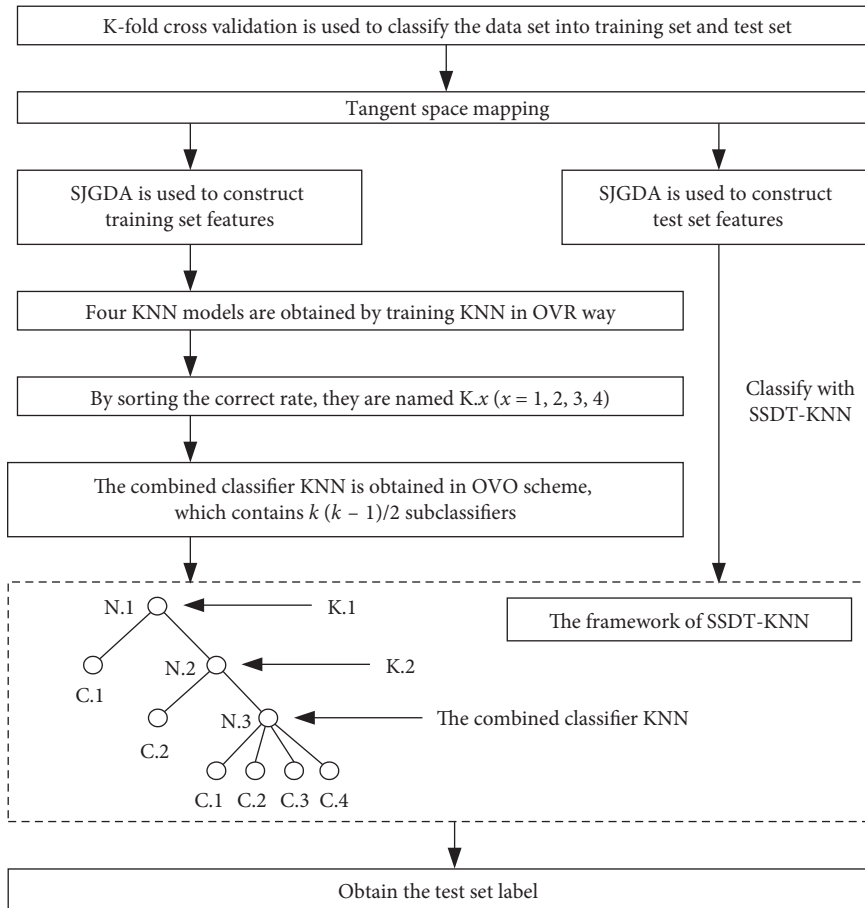


FIGURE 5: Block diagram for method 2.

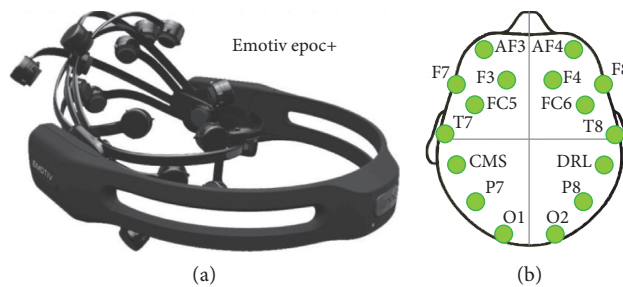


FIGURE 6: (a) Emotiv Epoc+ and (b) Emotiv 14 electrodes located over 10–20 international system positions.

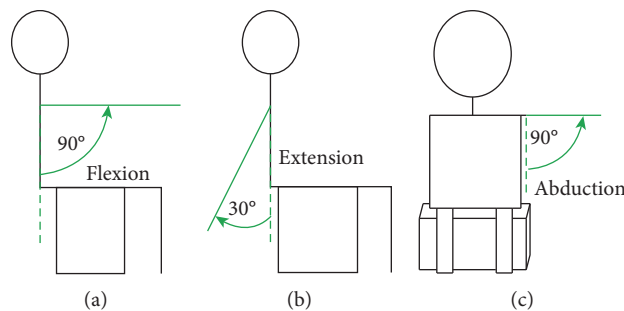


FIGURE 7: Three movements of shoulder joint: (a) flexion, (b) extension, and (c) abduction.

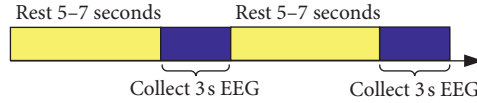


FIGURE 8: Timing for experimental process.

TABLE 1: Ten-folder cross-validation classification accuracy (%) for FGMDRM with OVR scheme applied on BCI competition dataset 2A.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	Std
L/RE	84.40	67.39	93.05	77.64	66.96	74.34	86.33	93.83	94.09	82.00	10.90
R/RE	<b>89.26</b>	75.33	<b>95.16</b>	78.56	63.92	71.81	81.18	93.39	82.92	81.28	10.24
F/RE	77.80	<b>83.44</b>	89.20	<b>80.88</b>	71.56	<b>78.22</b>	88.89	79.18	84.4	81.51	5.65
T/RE	88.21	68.76	90.59	79.95	<b>71.85</b>	76.72	<b>90.64</b>	<b>94.43</b>	<b>94.38</b>	83.95	9.82

TABLE 2: Kappa value comparison by SSDT-FGMDRM with other published results.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	<i>p</i> value
Method 1	0.66	0.39	<b>0.78</b>	0.47	0.25	<b>0.41</b>	<b>0.72</b>	<b>0.79</b>	<b>0.83</b>	<b>0.589</b>	
MDRM [10]	<b>0.75</b>	0.37	0.66	<b>0.53</b>	<b>0.29</b>	0.27	0.56	0.58	0.68	0.52	0.4683
FGMDM [14]	0.72	<b>0.50</b>	0.64	0.38	0.28	0.34	0.64	0.68	0.75	0.55	0.6607
FGMDM_fixed [14]	0.69	0.35	0.60	0.28	0.21	0.30	0.46	0.62	0.53	0.45	0.1423

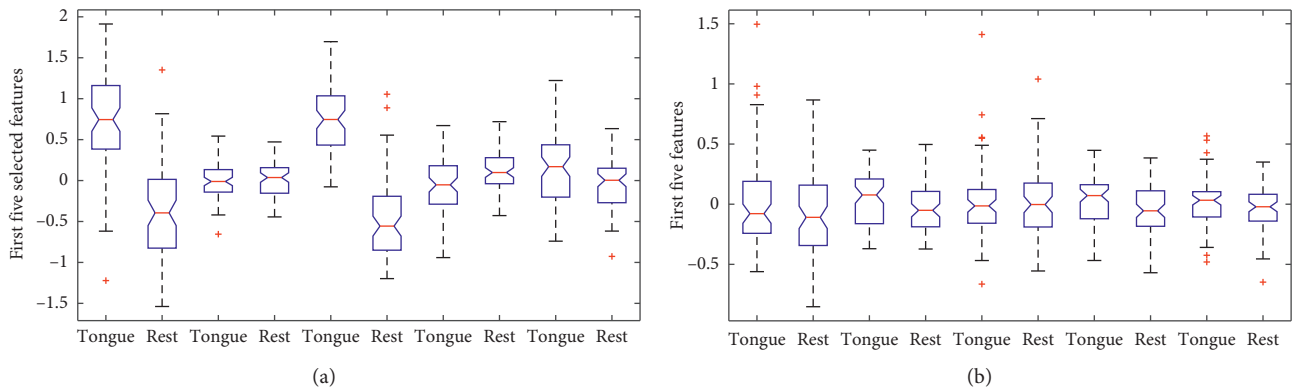


FIGURE 9: The box plot of five features of subject A09. (a) One-way ANOVA analysis on the first five features after applying *semi*-JMI. (b) One-way ANOVA analysis on the first five features from primitive vectors.

Figure 11 illustrates distributions for the first five most discriminant variables with GDA and *semi*-JMI. It can be seen from Figure 11 that L/RE is separated equally well by using GDA.

Table 3 displays ten-folder cross-validation results obtained using SJGDA and KNN in OVR scheme. It can be seen that the vectors which are mapped to the tangent space have better classification performance than that in the Riemannian manifold directly.

Table 4 presents the results obtained by SJGDA in pairwise way for multiclass MI tasks. We have six pairs of MI tasks: left and right (L/R), left and foot (L/F), left and tongue (L/T), right and foot (R/F), right and tongue (R/T), and foot and tongue (F/T).

Table 5 displays the comparison of classification accuracy using SJGDA and KNN for L/R task in 10-folder cross validation. References [8, 43–45] contain the classification of

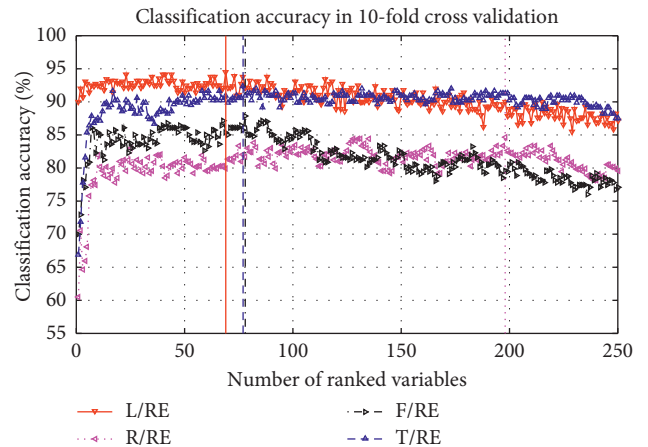


FIGURE 10: Classification accuracy corresponding the number of selected variables of Subject A09.

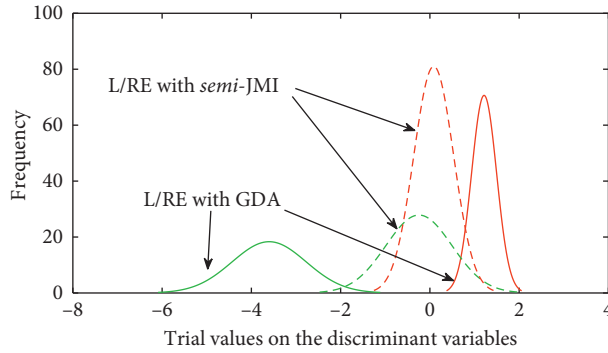


FIGURE 11: Feature distribution for the most discriminant variables in tangent space of Subject A09.

TABLE 3: Ten-folder cross-validation results (%) obtained using SJGDA and KNN in OVR scheme applied on BCI competition dataset 2A.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	Std
L/RE	90.29	74.62	93.05	75.02	75.31	<b>78.15</b>	85.08	92.35	<b>95.14</b>	84.33	8.12
R/RE	<b>91.34</b>	77.11	<b>94.47</b>	77.07	73.61	75.74	81.58	93.41	87.53	83.54	7.75
F/RE	85.06	<b>83.70</b>	84.77	77.11	75.36	77.09	88.17	82.30	85.44	82.11	4.24
T/RE	88.89	76.72	89.25	<b>80.91</b>	<b>78.51</b>	76.04	<b>89.27</b>	<b>94.08</b>	93.41	85.23	6.76

other publications. We have improved the accuracy compared with Reference [44] ( $p=0.85$ ) and Reference [45] ( $p=0.45$ ). Gaur et al. [43] ( $p=0.95$ ) explored the specific frequency information for each subject, and Sadatnejad and Shiry Ghidary [8] ( $p=0.90$ ) used a novel kernel for dimensionality reduction which is similar to SJGDA. Although the results in the paper are not as high as those in Reference [43], it can be concluded that there is no difference between the results in Reference [43] and those in this paper because of  $p=0.95$ .

Table 6 presents the results in terms of the kappa value. The proposed method 1 achieved a mean performance of 0.589 which ranks this method to the first place of the competition. And with our proposed method 2, we have achieved a mean performance of 0.607, which makes method 2 to acquire the best performance of the state of the art.

Dataset 2 is used to verify the effect of method 2, and the classification results are given directly in this paper. The results are shown in Table 7. As can be seen from Table 7, method 2 obtained the second highest recognition rate in the comparative literature. Compared with the recent reference [47], method 2 achieved good classification results.

**5.3. Results of Dataset 3.** Dataset 3 is used to evaluate the performance of method 2. Figure 12 shows the classification error with KNN against the number of ranked variables in OVR scheme. A/RE and F/RE are the two pairs with the lowest classification error, and they all achieved the highest recognition rate within 60 variables. In this paper, the first 60 ranked variables are used for the next analysis.

Figure 13 displays 5-folder cross-validation results obtained by using SJGDA and KNN in OVR and OVO scheme. This Figure 13(a) illustrates three possible pairs of MI tasks (F/RE, E/RE, and A/RE) for each subject. It can be learned

from the figure that flexion and abduction are the easiest movement to distinguish in six subjects of seven, and the six subjects are S1, S3, S4, S5, S6, and S7. However, due to individual differences, the highest recognition rate of each subject is different.

We also compared three possible pairs (F/E, F/A, and E/A) in OVO scheme of seven subjects. Figure 13(b) depicts the comparison results for each subject, and it can be seen that the pair of F/A obtained the highest recognition rate in seven subjects. Combined with the analysis results of Figures 13(a) and 13(b), it can be considered that flexion and extension are more obvious in the three MI tasks.

As SJGDA is a new method proposed in this paper, we also compared the feature distribution of SJGDA, GDA, and *semi*-JMI to illustrate the effectiveness of SJGDA. Figure 14 depicts the feature distribution of F/E MI tasks of seven subjects. The blue and red circles represent the two different feature classes. As shown in Figure 14, the F/E MI tasks learned by SJGDA have high separability than GDA and *semi*-JMI.

The performance of the proposed method 2 is evaluated by using classification accuracy. Since there are three classes, the chance level is 33.33%. Figure 15 demonstrates that the proposed method achieves higher performance for six subjects (S1, S2, S3, S4, S5, and S6) out of seven except S7 compared to *semi*-JMI and GDA methods. In addition, it also can be seen that GDA obtains a better classification accuracy for four subjects of seven (S1, S2, S5, and S7) compared with *semi*-JMI. The reasons for this phenomenon can be attributed to as follows: In the process of feature selection, we manually select feature dimensions suitable for classifiers, which results in partial information loss. As a feature dimensionality reduction technique, GDA is suitable for the preservation of useful information from the primitive vectors. And the proposed method SJGDA in this paper not only preserves the advantages of GDA but also adds some



TABLE 4: Ten-folder cross-validation results (%) obtained using SJGDA and KNN in OVO scheme applied on BCI competition dataset 2A.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	Std
L/R	90.95	67.29	94.52	63.87	64.48	70.33	70.29	<b>97.95</b>	95.00	79.41	13.85
L/F	95.19	87.43	93.93	81.84	68.10	<b>77.14</b>	<b>98.57</b>	88.29	93.81	87.14	9.29
L/T	96.52	64.24	<b>96.57</b>	83.24	<b>73.67</b>	69.91	97.95	97.23	<b>99.29</b>	86.51	13.16
R/F	95.90	<b>87.52</b>	95.05	<b>84.84</b>	66.71	72.88	97.24	93.18	87.43	86.75	10.01
R/T	<b>99.33</b>	79.82	96.48	79.12	72.22	71.48	97.24	95.05	92.29	87.00	10.61
F/T	83.94	84.17	86.62	75.10	62.62	74.21	88.31	93.10	90.33	82.04	9.10

TABLE 5: Comparison of classification accuracy (%) for L/R task with other published results using OVO scheme applied on BCI competition dataset 2A.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	Std	<i>p</i> value
Method 2	90.95	<b>67.29</b>	94.52	63.87	<b>64.48</b>	70.33	70.29	<b>97.95</b>	<b>95.00</b>	79.41	13.85	
Reference [43]	<b>91.49</b>	60.56	94.16	76.72	58.52	68.52	78.57	97.01	93.85	<b>79.93</b>	14.13	0.95
Reference [8]	88.89	59.03	90.28	<b>78.47</b>	62.50	<b>75.00</b>	72.92	93.06	87.50	78.63	11.63	0.90
Reference [44]	88.89	51.39	<b>96.53</b>	70.14	54.86	71.53	<b>81.25</b>	93.75	93.75	78.01	16.04	0.85
Reference [45]	90.28	54.17	93.75	64.58	57.64	65.28	62.50	90.97	85.42	73.84	15.02	0.45

TABLE 6: Kappa value comparison with other published results.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	<i>p</i> value
Method 2	0.77	0.38	0.76	0.47	0.27	<b>0.42</b>	0.73	<b>0.81</b>	<b>0.85</b>	<b>0.607</b>	
Method 1	0.66	0.39	<b>0.78</b>	0.47	0.25	0.41	0.72	0.79	0.83	0.589	0.8632
Reference [43]	<b>0.86</b>	0.24	0.70	<b>0.68</b>	0.36	0.34	0.66	0.75	0.82	0.60	0.9586
TSLDA [10]	0.74	0.38	0.72	0.50	0.26	0.34	0.69	0.71	0.76	0.567	0.6894
Winner 1 [46]	0.68	0.42	0.75	0.48	<b>0.40</b>	0.27	0.77	0.75	0.61	0.57	0.7051
Reference [8]	0.71	0.46	0.76	0.44	0.26	0.37	<b>0.79</b>	0.75	0.61	0.57	0.7245
Reference [14]	0.75	<b>0.49</b>	0.76	0.49	0.34	0.36	0.68	0.76	0.76	0.60	0.9353

TABLE 7: Five-folder cross validation by method 2 applied on BCI III dataset IIIa.

Subject	Method 2	Reference [47]	Reference [48]	Reference [49]	Reference [50]
k3b	91.67	90.00	86.67	94.20	<b>94.44</b>
k6b	75.00	76.25	<b>81.67</b>	69.00	62.50
l1b	81.67	77.91	<b>85.00</b>	78.60	78.33
Mean	82.78	81.38	<b>84.44</b>	80.60	78.42
<i>p</i> value		0.96	0.42	0.94	0.73

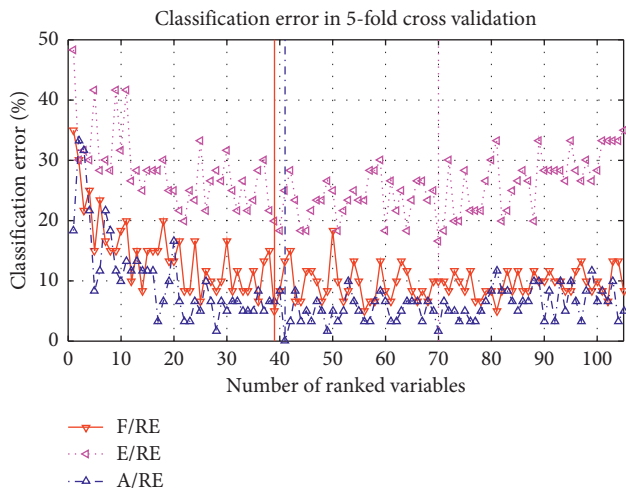


FIGURE 12: Classification accuracy corresponding to the number of selected variables of S01.

high ranking features to strengthen the expressive ability of the features.

## 6. Discussions

In this paper, we proposed a novel SSDT framework combined with classifiers to improve the performance of classifiers for multiclass MI tasks. We also proposed a novel NLDR method named SJGDA, and this NLDR method performs better than both *semi*-JMI and GDA on different datasets. In the following paragraphs, we have discussed the two methods in detail.

Method 1 indicates the drawback of FGMDRM, and then the novel SSDT framework is used to improve the accuracy for each individual. As shown in Table 2, compared with other published results, method 1 gets a quite good result in the case of processing the EEG signals of fixed frequency segment (8–30 Hz).

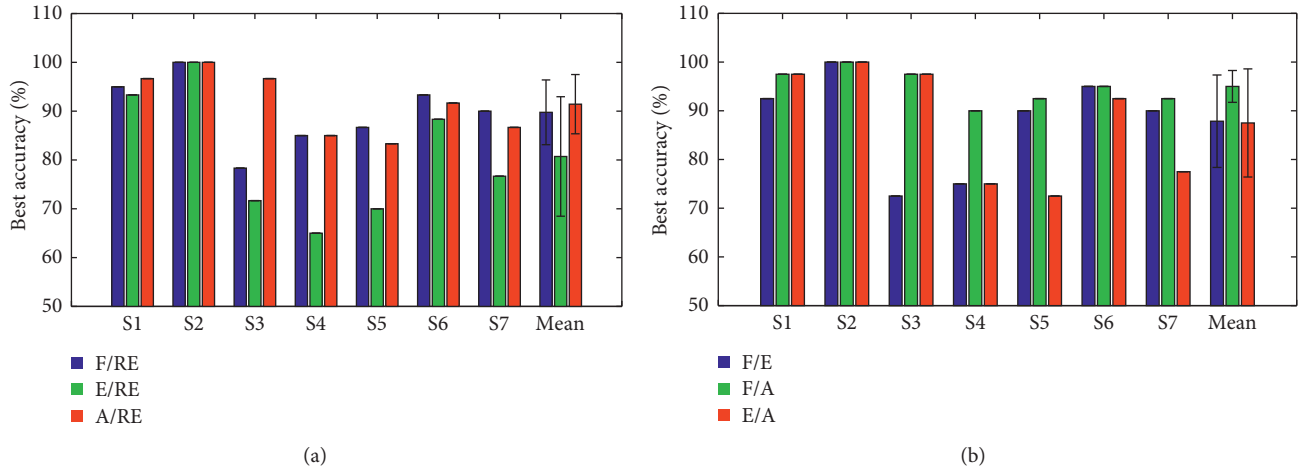


FIGURE 13: The comparison of three MI tasks in OVR scheme of seven subjects: (a) F/RE, E/RE, and A/RE; (b) F/E, F/A, and E/A.

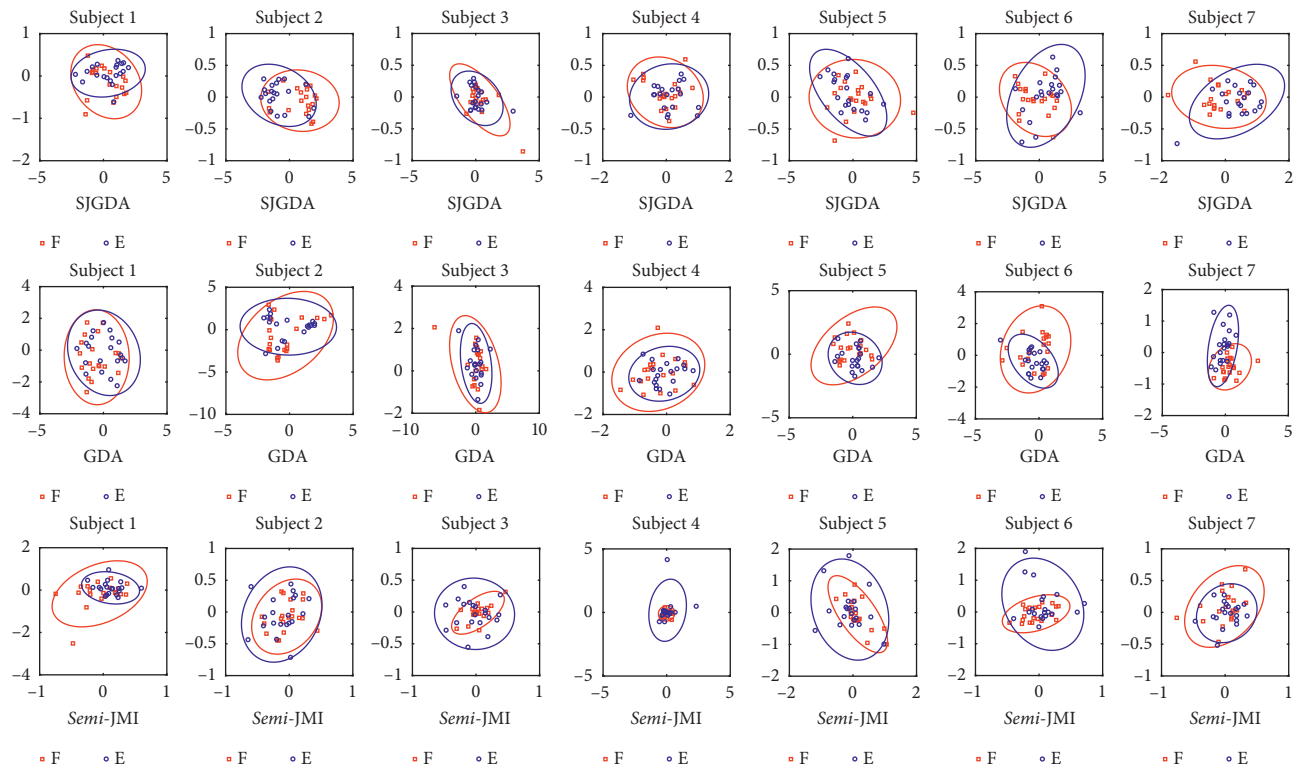


FIGURE 14: Feature distribution of F/E MI tasks extracted by SJGDA, GDA, and *semi*-JMI in dataset 2.

As shown in Table 6, Gaur et al. [43] proposed SS-MEMDBF to select the subject-specific frequency to obtain enhanced EEG signals which represent MI tasks related to  $\mu$  and  $\beta$  rhythms, then classification with the Riemannian distance directly. TSLDA was proposed by Barachant et al. [10], and the covariance matrices are mapped onto a higher dimensional space where they can be vectorized and treated as Euclidean objects. Ang et al. [46] is the winner of the competition, FBCSP and multiple OVR classifiers were used for MI tasks, and achieved the mean kappa value of 0.57. Sadatnejad and Shiry Ghidary [8] proposed a new kernel for NLDR over the manifold of SPD matrices, the

kappa value is 0.576. Davoudi et al. [14] considered the geometry of SPD matrices and provides a low-dimensional representation of the manifold with high-class discrimination, and the best result of this method in terms of the kappa value is 0.60.

In method 2, SJGDA is used to get more obvious vectors from the tangent vectors, and a SSDT-KNN classifier is used to identify different MI tasks. Combined with SJGDA and SSDT-KNN, we have achieved a better performance compared with method 1 ( $p = 0.8632$ ), Reference [43] ( $p = 0.9586$ ), TSLDA ( $p = 0.6894$ ), winner 1 ( $p = 0.7051$ ), Reference [8] ( $p = 0.7245$ ), and Reference [14] ( $p = 0.9353$ ).

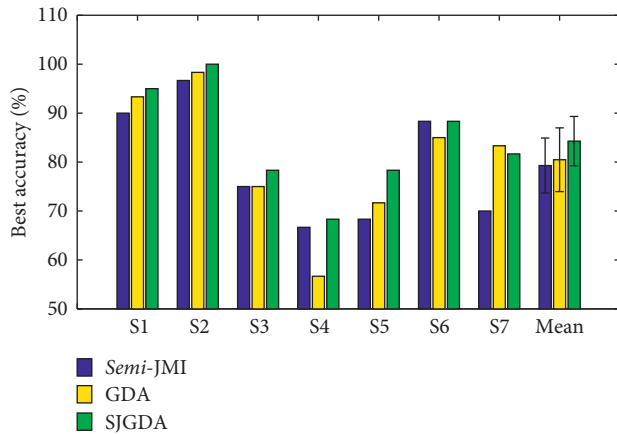


FIGURE 15: The comparison of the three different feature selection methods (*semi*-JMI, GDA, and SJGDA) with the same classifier (DT-KNN) of three MI tasks classification accuracy for each subject is depicted. The results are obtained by 5-folder cross validation.

It is clear that the proposed method in this paper is effective for MI tasks in a BCI system.

In order to prove the effectiveness of the proposed method 2, we tested it on two other datasets. As shown in Table 7 and Figure 15, method 2 achieves good classification results on two datasets.

## 7. Conclusion

The experimental results of method 1 show that the proposed classification framework significantly improves the classification performance of the classifier. The experimental results of method 2 show that the SJGDA algorithm proposed in this paper is superior to GDA and *semi*-JMI in feature extraction, and method 2 has the highest recognition rate in this paper. However, as the classifiers in the SSDT framework is substitutable, the focus of the next work is to combine more advanced classifiers with SSDT to increase the recognition rate of the BCI systems.

## Data Availability

The dataset 1 and dataset 2 used to support the findings of this study are available from <http://bnci-horizon-2020.eu/database/data-sets>. The dataset 3 used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

Northeast Electric Power University (Grant number BSJXM-201521) and Jilin City Science and Technology Bureau (Grant number 20166012).

## References

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [2] N. Birbaumer, "Breaking the silence: brain-computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.
- [3] K. K. Ang, C. Guan, K. S. Phua et al., "Transcranial direct current stimulation and EEG-based motor imagery BCI for upper limb stroke rehabilitation," in *Proceedings of 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4128–4131, IEEE, San Diego, CA, USA, August 2012.
- [4] B. Xu, S. Peng, A. Song, R. Yang, and L. Pan, "Robot-aided upper-limb rehabilitation based on motor imagery EEG," *International Journal of Advanced Robotic Systems*, vol. 8, no. 4, p. 40, 2011.
- [5] F. Wang, X. Zhang, R. Fu, and G. Sun, "Study of the home-auxiliary robot based on BCI," *Sensors*, vol. 18, no. 6, p. 1779, 2018.
- [6] F. Wang, H. Wang, and R. Fu, "Real-Time ECG-based detection of fatigue driving using sample entropy," *Entropy*, vol. 20, no. 3, 2018.
- [7] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topography*, vol. 2, no. 4, pp. 275–284, 1990.
- [8] K. Sadatnejad and S. Shiry Ghidary, "Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application," *Neurocomputing*, vol. 179, pp. 152–160, 2016.
- [9] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017.
- [10] 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.
- [11] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Common spatial pattern revisited by riemannian geometry," in *Proceedings of 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pp. 472–476, IEEE, Saint Malo, France, October 2010.
- [12] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "A brain-switch using riemannian geometry," in *Proceedings of 5th International Brain-Computer Interface Conference 2011 (BCI 2011)*, pp. 64–67, Graz, Austria, September 2011.
- [13] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.
- [14] A. Davoudi, S. S. Ghidary, and K. Sadatnejad, "Dimensionality reduction based on distance preservation to local mean (DPLM) for spd matrices and its application in BCI," 2016, <https://arxiv.org/abs/1608.00514>.
- [15] S. Brandl, K.-R. Müller, and W. Samek, "Robust common spatial patterns based on Bhattacharyya distance and gamma divergence," in *Proceedings of 2015 3rd International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, IEEE, Gangwon-Do, South Korea, January 2015.
- [16] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms,"

- IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.
- [17] W. Samek and K.-R. Müller, “Information geometry meets BCI spatial filtering using divergences,” in *Proceedings of 2014 International Winter Workshop on Brain-Computer Interface (BCI)*, pp. 1–4, IEEE, Gangwon, South Korea, February 2014.
- [18] W. Samek and M. Kawanabe, “Robust common spatial patterns by minimum divergence covariance estimator,” in *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2040–2043, IEEE, Florence, Italy, May 2014.
- [19] S. Kumar, K. Mamun, and A. Sharma, “CSP-TSM: optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-BCI,” *Computers in biology and medicine*, vol. 91, pp. 231–242, 2017.
- [20] C. Lindig-León, N. Gayraud, L. Bougrain, and M. Clerc, “Comparison of hierarchical and non-hierarchical classification for motor imagery based BCI systems,” in *Proceedings of The Sixth International Brain-Computer Interfaces Meeting*, Pacific Grove, CA, USA, May-June 2016.
- [21] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [22] O. Kramer and D. Lückehe, “Visualization of evolutionary runs with isometric mapping,” in *Proceedings of 2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1359–1363, IEEE, Sendai, Japan, May 2015.
- [23] K. Q. Weinberger, F. Sha, and L. K. Saul, “Learning a kernel matrix for nonlinear dimensionality reduction,” in *Proceedings of the Twenty-First International Conference on Machine Learning*, p. 106, ACM, Banff, Canada, July 2004.
- [24] L. Van Der Maaten, “Learning a parametric embedding by preserving local structure,” *RBM*, vol. 500, p. 26, 2009.
- [25] L. Van Der Maaten, “Accelerating t-SNE using tree-based algorithms,” *Journal of Machine Learning Research*, vol. 15, pp. 3221–3245, 2014.
- [26] D. Lee, S.-H. Park, and S.-G. Lee, “Improving the accuracy and training speed of motor imagery brain-computer interfaces using wavelet-based combined feature vectors and Gaussian mixture model-supervectors,” *Sensors*, vol. 17, no. 10, p. 2282, 2017.
- [27] X. Xie, Z. L. Yu, H. Lu, Z. Gu, and Y. Li, “Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 504–516, 2017.
- [28] M. T. Harandi, M. Salzmann, and R. Hartley, “From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices,” in *Proceedings of European Conference on Computer Vision*, pp. 17–32, Springer, Zurich, Switzerland, September 2014.
- [29] 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.
- [30] P. T. Fletcher and S. Joshi, “Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors,” in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, M. Sonka, I. A. Kakadiaris, and J. Kybic, Eds., pp. 87–98, Springer, Berlin, Heidelberg, 2004.
- [31] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, “Riemannian geometry applied to BCI classification,” in *Proceedings of International Conference on Latent Variable Analysis and Signal Separation*, pp. 629–636, Springer, St. Malo, France, September 2010.
- [32] H. Guo and S. B. Gelfand, “Classification trees with neural network feature extraction,” *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 923–933, 1992.
- [33] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [34] Y.-H. Shao, W.-J. Chen, W.-B. Huang, Z.-M. Yang, and N.-Y. Deng, “The best separating decision tree twin support vector machine for multi-class classification,” *Procedia Computer Science*, vol. 17, pp. 1032–1038, 2013.
- [35] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, “Conditional likelihood maximisation: a unifying framework for information theoretic feature selection,” *Journal of Machine Learning Research*, vol. 13, pp. 27–66, 2012.
- [36] K. Sechidis and G. Brown, “Simple strategies for semi-supervised feature selection,” *Machine Learning*, vol. 107, no. 2, pp. 357–395, 2017.
- [37] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, “A unifying view on dataset shift in classification,” *Pattern Recognition*, vol. 45, no. 1, pp. 521–530, 2012.
- [38] G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach,” *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [39] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, “Cloudid: trustworthy cloud-based and cross-enterprise biometric identification,” *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.
- [40] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, Berlin, Germany, 2013.
- [41] C. Brunner, R. Leeb, G. Müller-utz, A. Schlögl, and G. Pfurtscheller, BCI Competition 2008–Graz Data Set A, vol.16, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces) Graz University of Technology, Graz, Austria, 2008.
- [42] B. Blankertz, K. R. Muller, D. J. Krusienski et al., “The BCI competition III: validating alternative approaches to actual BCI problems,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [43] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, “A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and riemannian geometry,” *Expert Systems with Applications*, vol. 95, pp. 201–211, 2018.
- [44] F. Lotte and C. Cuntai Guan, “Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.
- [45] H. Raza, H. Cecotti, Y. Li, and G. Prasad, “Adaptive learning with covariate shift-detection for motor imagery-based brain-computer interface,” *Soft Computing*, vol. 20, no. 8, pp. 3085–3096, 2015.
- [46] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b,” *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.
- [47] H. Baali, A. Khorshidtalab, M. Mesbah, and M. J. E. Salami, “A transform-based feature extraction approach for motor imagery tasks classification,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 3, pp. 1–8, 2015.
- [48] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, “Characterization of four-class motor imagery EEG data for the

- BCI-competition 2005,” *Journal of Neural Engineering*, vol. 2, no. 4, pp. L14–L22, 2005.
- [49] 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.
- [50] I. Koprinska, “Feature selection for brain-computer interfaces,” in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 106–117, Springer, Bangkok, Thailand, April 2009.