Transfer Learning Based on Hybrid Riemannian and Euclidean Space Data Alignment and Subject Selection in Brain-Computer Interfaces

YAN LI, QINGGUO WEI [iD], YUEBIN CHEN, AND XICHEN ZHOU
Department of Electronic Information Engineering, School of Information Engineering, Nanchang University, Nanchang 330031, China
Corresponding author: Qingguo Wei (wqg07@163.com)

ABSTRACT Transfer learning is a promising approach for reducing training time in a brain-computer interface (BCI). However, how to effectively transfer data from previous users to a new user poses a huge challenge. This paper presents a novel transfer learning approach that combines data alignment and source subject selection for motor imagery (MI) based BCIs. The former is achieved by a reference matrix from the regularization of the two reference matrices estimated in Riemannian and Euclidean space respectively, whereas the latter is implemented by a modified sequential forward floating-point search algorithm. The aligned training data from chosen source subjects are used for creating a classification model based on either spatial covariance matrices in Riemannian space or common spatial pattern algorithm in Euclidean space. The proposed algorithms were evaluated on two MI based BCI data sets with different subjects and compared with existing transfer learning algorithms with sole data alignment or subject selection. The experimental results show that the hybrid-space data alignment methods for reducing the differences among subjects significantly outperform two single-space alignment methods, and the source subject selection method can substantially enhance the similarity between source subjects and the target subject. The combination of the two methods achieves superior classification performance compared to either one. The proposed algorithms will greatly facilitate the real-world applications of MI based BCIs.

INDEX TERMS Brain-computer interfaces, transfer learning, data alignment, hybrid Riemannian and Euclidean space data alignment, source subject selection.

I. INTRODUCTION

A brain-computer interface (BCI) system provides a new non-muscular channel for sending messages from the brain to the external world by analyzing electrical brain activity or other electrophysiological measures of brain functions [1]. According to the generation of EEG signals, BCI systems can be divided into two types: spontaneous and evoked [2]. As a typical kind of spontaneous systems, motor imagery (MI) based BCIs do not need additional stimulation devices. Instead, their input signals are generated by users via imagining their own limb movements or observing the movements of others [3], [4]. Among the various imaging models of brain signals, electroencephalography (EEG) can be recorded by electrodes placed on the scalp and thus

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma [iD].

provides the advantages such as non-invasive, easy to acquire and high time resolution.

Common spatial pattern (CSP) is the most commonly used algorithm for spatial filtering and subsequent feature extraction [5], [6]. Linear discriminant analysis (LDA) [7] is a typical feature discrimination method. LDA is widely applied in MI based BCIs because it is robust to noise and has no parameters to tune. The combination of CSP and LDA is an efficient method for creating a classification model in Euclidean space. In MI based BCIs, the second-order statistics of EEG signals contain the separable information of brain states [8]. Spatial covariance matrix (SCM) is the most commonly used second-order statistic of EEG signals [9]. The spatial interaction information among channels is completely embedded into its SCM. A optimization algorithm based on the covariance matrix for either feature extraction or classification is indeed in line with the rationale of MI

based BCIs, which are topography-orientated technology. Thereby, the performance of both CSP and LDA depends on the accurate estimation of SCMs. On the other hand, SCM is a symmetric positive definite (SPD) matrix that lies on a differential Riemannian manifold. Each SCM is one point in the manifold, whose Riemannian distance to a reference point can be used as a discriminative feature signal classified directly by a simple classifier called minimum distance to mean (MDM) [10]–[12]. The combination of SCM and MDM is another efficient method for creating a classification model in Riemannian space.

A major assumption in machine learning is that the training and testing data are drawn from the same feature space and the same distribution. In MI based BCIs, the assumption does not hold because large variability exists across subjects and across sessions of the same subject. As a result, the classification model needs to be rebuilt from scratch of each use. This problem severely limits the real-world application of a BCI. Thereby, one major problem in the development of a BCI is to reduce the training time or completely eliminate it. Transfer learning (TL) is a promising approach to achieve the goal. TL fuses the data from other subjects or previous sessions into the current learning process [13]–[18]. How to effectively transfer data from the previous subjects to a new subject, however, poses a huge challenge due to the great inter-subject variability.

A typical method for improving the TL in BCIs is to use the experimental samples of other subjects to regularize the model of a target subject in order to improve the performance of the model [19]–[21]. Kang *et al.* measured the similarity between subjects by KL-divergence to improve the covariance matrix by regularization [19]. The premise of the method is that the covariance matrices of source subjects are highly similar to those of the target subject. However, the assumption is violated due to the large individual discrepancy. To alleviate the problem, Lotte *et al.* used a subset of automatically selected subjects to formulate a weighted sum of covariance matrices [20]. Another method for improving the TL in BCIs is to align EEG data from different subjects [22]–[24]. Rodrigues *et al.* [22] proposed a method based on Procrustes analysis for matching the statistical distributions of two data sets using geometrical transformations (translation, scaling and rotation) over the data points. The method handles the statistical variability of EEG signals from different subjects. Zanini *et al.* [23] proposed a Riemannian space data alignment (RA) approach that aligns the covariance matrices of a subject with the reference matrix estimated by the EEG data of resting states, i.e. the transitional periods between two trials. The implicit assumption is that the covariance matrix shift relative to the reference matrix caused by different source configurations and electrode positions can be regarded as the covariance matrix moving in the same direction on the SPD manifold when the brain is involved in a specific task. The aligned covariance matrices are then transferred to the target subject for creating a classification model based on Riemannian geometry. The RA can improve the perfor-

mance of the MDM classifier by making use of subsidiary data from other subjects. Subsequently, He *et al.* [24] proposed a Euclidean space data alignment (EA) approach that aligns the EEG trials of a subject with the reference matrix estimated by the EEG data in task states without using the label information. The aligned trials are then transferred to the target subject for creating a classification model with CSP and LDA in Euclidean space. The EA further improved the classification performance of MI based BCIs compared to RA.

The above two alignment methods aim to make EEG data of source subjects more similar to those of the target subject, so that the classification model for a new subject can be built using EEG data from previous subjects. However, the two methods still have room for improvement due to the limitations of the calibration space and/or the reference matrix. After data alignment, the difference in data distribution between source subjects and a target subject is decreased, but not EEG data from all previous subjects are suitable for transferring. The reason is that huge difference exists in the statistical distribution of EEG data among subjects [25], [26] due to individual discrepancy of subjects in physiology and anatomy, the change of electrode positions as well as the difference in subjects' mental states. Thus, it is reasonable to select those source subjects relevant to a target subject, and then use their data as training data of the target subject for creating a classification model.

In the study, we propose a novel transfer learning algorithm for MI based BCIs that combines a new method for data alignment and a method for subject selection. The former is named hybrid Riemannian and Euclidean space data alignment (REA), which aligns EEG data with the reference matrix yielded by regularizing the two reference matrices estimated in Riemannian and Euclidean space (RA and EA) respectively, whereas the latter selects relevant source subjects with a sequential forward floating-point search algorithm [27], which was originally proposed by Pudil *et al.* for feature selection, modified for subject selection (SS) by Lotte and Guan [20] and adopted in this study. The proposed transfer learning algorithm was evaluated by systematic comparisons with existing algorithms on two MI data sets.

II. METHOD

Human limbs are associated with specific brain regions, for example, left hand, right hand and both feet correspond respectively to right lobe, left lobe and anterior central gyrus. Motor imagery results in two physiological phenomena known as event-related desynchronisation (ERD) and event-related synchronization (ERS), i.e., the power of the mu rhythm (8–12 Hz) will decrease in the specific lobe of his brain when a subject imagines a limb movement, whereas that of beta rhythm (14–18Hz) will rebound in the same lobe when the motor imagery is over. The topographical representation and band power changes of the mental tasks are the basis for creating a MI based BCI.

A. SPATIAL COVARIANCE MATRIX

SCM reflects the power distribution of EEG signals on single channels and across channels and thus is used as a significant descriptor to develop various algorithms for discriminating mental tasks in a BCI system. In general, a single MI trial consists of a task period and a resting period, and is denoted by a short segment of data (say 2 s) in the task period. Rhythmic signals are extracted via bandpass filtering, typically in 8–30 Hz because the frequency band includes mu and beta rhythms. Let $X \in R^{N_c \times N_s}$ be a single-trial bandpass filtered EEG signal, where N_c denotes the number of channels and N_s denotes the number of sampling points. The SCM of signal X is estimated by sample covariance matrix calculated as

$$P = \frac{1}{N_S - 1} X X^T \qquad (1)$$

where T represents transpose operation. The SCM is a symmetric positive definite (SPD) matrix that lies in the particular Riemannian manifold, a subset of Euclidean space. The Riemannian distance between two SCMs and the mean of multiple SCMs are often used for designing discriminative algorithms in machine learning field and can be calculated in either Euclidean or Riemannian space [28].

The Riemannian distance is defined as the shortest geodesic length between P_1 and P_2, which has a closed-form solution [29]–[32] calculated as follows

$$\delta_R (P_1, P_2) = \left\| \log \left(P_1^{-1} P_2 \right) \right\|_F = \sqrt{\sum_{n=1}^{N} \log^2 \lambda_n} \quad (2)$$

where λ_n is the eigenvalues of matrix $P_1^{-1} P_2$. Since P_1 and P_2 are SPD matrices, their Riemannian distance have an important property termed as congruence invariance [33]

$$\delta_R (P_1, P_2) = \delta_R \left(U P_1 U^T, U P_2 U^T \right) \qquad (3)$$

where U represents an invertible and orthogonal matrix. Because of the excellent property, the linear transformation of two SPD matrices will not change their relative distance in the Riemannian space.

The arithmetic (or Euclidean) mean of multiple SPD matrices minimizes the sum of the squared Euclidian distances:

$$\bar{P}_E (P_1, \cdots, P_N) = \operatorname*{arg\,min}_{P \in P(n)} \sum_{i=1}^{N} \delta_E^2 (P_i, P) = \frac{1}{N} \sum_{i=1}^{N} P_i \qquad (4)$$

where $P(n)$ belongs to the set of SPD matrices with dimensionality of $n \times n$. The geometric (or Riemannian) mean of multiple SPD matrices is defined as the matrix that minimizes the sum of the squared Riemannian distances [34]

$$\bar{P}_R (P_1, \cdots, P_N) = \operatorname*{arg\,min}_{P \in P(n)} \sum_{i=1}^{N} \delta_R^2 (P_i, P) \qquad (5)$$

The existence and unicity of the Riemannian mean were proved in [29], [35]. However, an explicit solution exists only for $N = 2$, where it coincides with the middle point of the geodesic connecting the two SPD matrices. For $N > 2$,

a solution can be found iteratively, and several algorithms following different approaches have been developed in [36].

B. DATA ALIGNMENT

As a signal transformation method, data alignment aligns single-trial EEG signals with a reference matrix to make them moving toward the reference matrix. Data alignment increases the similarity of EEG data between different subjects, improves the performance of TL in BCI [37]–[39], and hence makes cross-subject classification efficient. The choice of a reference matrix is crucial because it defines the center of all single-trial EEG signals from the same subject. Data alignment can be conducted in Riemannian space (RA) [23] or Euclidean space (EA) [24]. In the study, we propose a hybrid Riemannian and Euclidean space data alignment (REA) method.

1) DATA ALIGNMENT IN RIEMANNIAN SPACE

RA utilizes Riemannian mean of the covariance matrices from all EEG signals in resting periods as the reference matrix, in which the subject does not perform any mental tasks, to align the covariance matrices in the Riemannian manifold. The reference matrices of different subjects can be approximated as an identity matrix [23]. The aligned covariance matrix for ith trial X_i is computed as follows

$$P_i' = \bar{P}_R^{-1/2} P_i \bar{P}_R^{-1/2} \qquad (6)$$

where P_i is the covariance matrix of X_i and $\bar{P}_R$ is the reference matrix for a subject. Since classification in Riemannian space is based on the distance between a single-trial covariance matrix and the mean covariance matrix of each class, data alignment is performed for covariance matrices. Since all aligned EEG data from any one subject are positioned nearby identity matrix, the similarity of data distributions between the source subject and target subject increases substantially and thus classification of mental tasks becomes easy.

2) DATA ALIGNMENT IN EUCLIDEAN SPACE

EA employs Euclidean mean of covariance matrices from all EEG signals in task periods as a reference matrix, in which the subject conducts mental tasks of motor imagery. The reference matrix of a subject is strictly equivalent to an identity matrix [24]. Different from RA, EA aligns EEG trials instead of covariance matrices because in Euclidean space, classification is based on spatially filtered EEG trials. Like RA alignment, the purpose of EA alignment is to increase the similarity of data distributions between a source subject and the target subject so that classification of mental tasks becomes easy. The aligned EEG signal X_i' for ith trial X_i is computed as

$$X_i' = \bar{P}_E^{-1/2} X_i \qquad (7)$$

where $\bar{P}_E$ is the reference matrix for a subject.

3) DATA ALIGNMENT IN HYBRID RIEMANNIAN AND EUCLIDEAN SPACE

REA aims to overcome the limit of both RA and EA in alignment space and/or reference matrix so that the aligned data can be utilized for creating a classification model in both Riemannian and Euclidean space. The reference matrix of REA, $\bar{P}_{RE}$, is generated by regularizing the two reference matrices $\bar{P}_R$ and $\bar{P}_E$ yielded respectively in Riemannian and Euclidean space, and formulated as follows

$$\bar{P}_{RE} = \lambda \bar{P}_R + (1 - \lambda) \bar{P}_E \quad (8)$$

where $\lambda \in [0, 1]$ is a regularization parameter for controlling the weights of the two reference matrices. The optimal value of λ is determined by classification accuracy using exhaustive search in the range of $[0, 0.1, 0.2, \ldots, 1]$. With the reference matrix, the ith trial X_i and its covariance matrix P_i are respectively aligned as

$$X_i' = \bar{P}_{RE}^{-1/2} X_i, \, P_i' = \bar{P}_{RE}^{-1/2} P_i \bar{P}_{RE}^{-1/2} \quad (9)$$

where X_i' and P_i' are the aligned EEG signal and covariance matrix respectively.

C. SUBJECT SELECTION

The purpose of subject selection is to select those source subjects relevant to the target subject in order to ensure all transferred data useful for the target subject. The sequential forward floating-point search algorithm [27], proposed by Pudil *et al.* for feature selection, was modified for subject selection (Algorithm 1). This algorithm sequentially adds or deletes one subject from the current subset of source subjects at each loop as long as the resulting subset is better than the previously evaluated one at that level. The process proceeds until the classification accuracy of testing data from the target subject no longer increases. In the algorithm, the function $Acc = trainThenTest(trainingSet, testingSet)$ returns the accuracy yielded by training a classification model based on either CSP and LDA or SCM and MDM on training set (*trainingSet*) and testing it on test set (*testingSet*). The training set is initialized as the aligned data from all source subjects available, whereas the testing set is the aligned EEG data from the target subject.

D. CLASSIFICATION MODEL

The classification model of a BCI can be constructed in either Euclidean space or Riemannian space. In Euclidean space, the typical algorithm for feature extraction is common spatial pattern (CSP) and that for classification is linear discriminant analysis (LDA). The CSP aims to find a spatial filtering matrix $W \in R^{N_C \times N_C}$ that maximizes the variance of one class while minimizes the variance of the other class [5], [6]. By jointly diagonalizing the two-class signals, CSP enhances the difference in variance (or band power) between the two conditions. The spatial filtering matrix (also called projection matrix) W is obtained by maximizing or minimizing the

---

Algorithm 1 Selecting relevant transfer subjects

Input:

D_s: Aligned EEG training data from the source subjects.

D_T: Aligned EEG testing data from the target subjects.

Output:

D_{Tr}: A subset of relevant subjects whose data can be used to classify data D_T of the target subject

Initialization:

    stop when the certificate does not increase

    $Acc(D_k) \approx Acc(D_{k-1})$

step 1 (inclusion):

    $D^+ = \arg \max_{D_k \varepsilon D_s} Acc(D_k + D_s, D_T)$,

    $D_{k+1} = D_k + D^+, \, k = k + 1$

step 2 (conditional exclusion):

    $D^- = \arg \max_{D_k \varepsilon D_s} Acc(D_k - D_s, D_T)$

    If $Acc(D_k - D^-) > Acc(D_{k-1})$ then

        $D_{k-1} = D_k - D^-, \, k = k + 1$

        go to step 2

    else

        go to step 1

    $D_{\text{transfer}} = D_{k+1}$

---

following cost function

$$J(W) = \frac{W^T \bar{P}_1 W}{W^T \bar{P}_2 W} \quad (10)$$

where $\bar{P}_l, l = 1, 2$ is the arithmetic mean of SCMs from the EEG signals belonging to class l, which is computed as

$$\bar{P}_l = \frac{1}{N_t} \sum_{i=1}^{N_t} P_i^l = \frac{1}{N_t(N_s - 1)} \sum_{i=1}^{N_t} X_i^l (X_i^l)^T \quad (11)$$

where N_t and X_i^l are the total number of EEG trials and the ith EEG trial from class l respectively. The optimal projection matrix $\tilde{W}$ is composed of the eigenvectors of $\bar{P}_2^{-1} \bar{P}_1$, which corresponds to its m largest and m smallest eigenvalues ($m = 3$ in this study). The ith column of $\tilde{W}$, $\tilde{w}_i \in R^{N_C \times 1}$, is called a spatial filter. Once the projection matrix is obtained, the feature vector used for classification is extracted by $f = \log(\tilde{W}^T P \tilde{W})$, where P is a single-trial covariance matrix. The LDA classifier separates feature vectors using a linear hyperplane [7]. With LDA, the classification output of an input feature vector f is equal to $a^T f + b$, where the normal vector a and intercept b are calculated as follows

$$a = \hat{P}^{-1}(\bar{f}_1 - \bar{f}_2)^T, \quad b = -\frac{1}{2}(\bar{f}_1 + \bar{f}_2)a \quad (12)$$

where $\bar{f}_1$ and $\bar{f}_2$ are the mean feature vectors from each class, and $\hat{P}^{-1}$ is the composite feature vector of two classes.

In Riemannian space, Riemannian geometry provides a tool to manipulate SCMs. Specifically, the SCM of an EEG signal can be directly employed as a feature signal without the need for spatial filtering. The minimum distance to Riemannian mean (MDM) classifier is used for classifying the SCM.
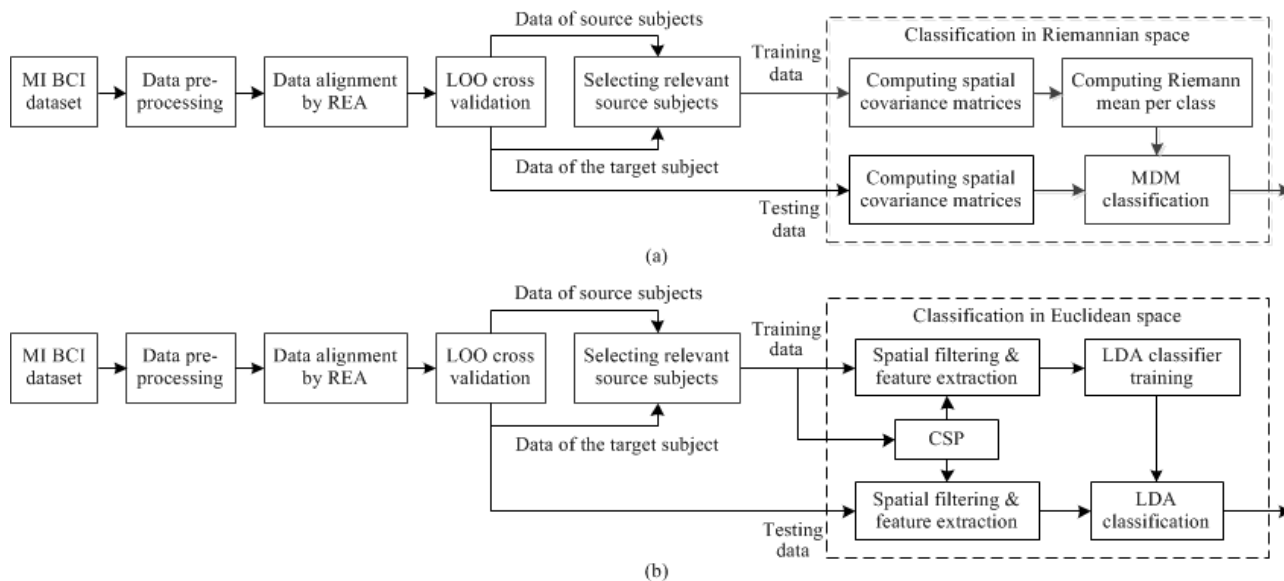
FIGURE 1. The processing flowchart for classifying EEG signals based on transfer learning in Riemannian space using SCM and MDM algorithm (a) and in Euclidean space using CSP and LDA (b). The subjects in a data set are divided into the source subjects and the target subject by leave-one subject-out (LOO) cross validation.

The single-trial EEG signals from each subject in a MI based BCI data set were first preprocessed by data segmentation and band pass filtering, and then aligned by REA. A leave-one subject-out (LOO) cross validation method is used to divide the subjects into source subjects and the target subject. After the selection of relevant source subjects, the aligned EEG trials from a chosen subset of source subjects are transferred to the target subject as the training data. A classification model based on transfer learning can be created in either Riemannian space or Euclidean space. The data processing flowchart for classifying MI EEG signals is illustrated in Fig. 1.

III. EXPERIMENTAL DATA

In this study, two MI based EEG data sets were used for evaluating the performance of the proposed algorithms. The first one is the data set 2a of BCI Competition IV [40], and the other is the second data set in the data base provided by the BNCI Horizon 2020 European Coordination project [41]. The two data sets were adopted because of the large number of subjects (9 in the first and 14 in the second). The main difference between them is that they have different number of channels, trials and mental tasks.

A. DATA SET 1

The data set contains EEG data from 9 subjects. The cue-based BCI paradigm consisted of four different mental tasks, namely MI of left hand, right hand, both feet and tongue. During the experiment, the subjects were sitting in an armchair in front of a computer screen. At the beginning of a trial, a fixation cross appeared on the screen and meanwhile a short acoustic warning tone was presented. After 2 s, an arrow pointing to either the left, right, down or up (corresponding

to one of the four classes) appeared and stayed on the screen for 1.25 s. This arrow cue prompted the subjects to perform the desired MI task. The mental imagination lasted until 6 s. No feedback was provided. A short break followed that lasted 1.5~2.5 s. The timing scheme of each trial is shown in Fig. 2(a).
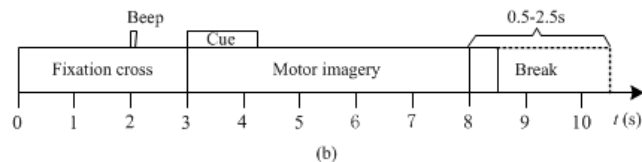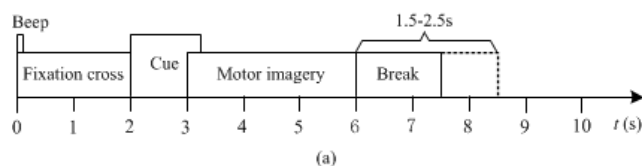

FIGURE 2. The timing scheme of a trial in (a) the data set 1 and (b) the data set 2.

The EEG data were recorded using 22 Ag/AgCl electrodes with the left and right mastoid serving as reference and ground respectively. The EEG signals were sampled at 250 Hz and bandpass filtered between 0.5 Hz and 100 Hz. An additional 50 Hz notch filter was used to suppress line noise. Each subject performed two sessions (session T for training and session E for evaluation) on different days. Each session includes six runs separated by short breaks. One run is composed of 48 trials, 12 trials per task, yielding a total of 288 trials per session. Only the session T and the two classes of EEG data from left and right hand were employed in the study.

B. DATA SET 2

The data set contains EEG data from 14 subjects, eight of whom are naïve. The cue-based BCI paradigm consisted of two different mental tasks, namely MI of right hand and feet. During the experiment, the subject fixated on a computer monitor 150 cm in front of her/him. Each trial is 8 s long and starts with the presentation of a fixation cross at the center of the monitor, followed by a short warning tone (beep) at 2 s. At 3 s, the fixation cross is overlaid with a cue arrow for 1.25 s, pointing to either the right or the down (corresponding to the two MI tasks). Depending on the direction of the arrow, the subject is asked to imagine the desired mental task, which sustains 5 s. A short break followed that lasted $0.5 \sim 2.5$ s. The timing scheme of each trial is shown in Fig. 2(b).

The EEG data were recorded using 15 Ag/AgCl electrodes with the left and right mastoid serving as reference and ground respectively. These electrodes were placed for obtaining three Laplacian derivations with center electrodes at positions C3, Cz and C4 and four additional electrodes around each center electrode. The EEG signals were sampled at 512 Hz. Each subject performed a single session, which consisted of eight runs, five of them for training and three with feedback for validation. The feedback was presented in form of a white bar graph, the length of which reflected the amount of correct classifications over the last second. One run was composed of 20 trials, 10 trials per task, yielding 50 trials per class for training and 30 trials per class for validation. Only the training runs were used in this study.

C. DATA PREPROCESSING

For the classification of mental tasks, each trial from the two data sets was band-pass filtered by a 5-th order Butterworth filter in the frequency band of 8–30 Hz. The data segments of 3 s were intercepted from 2.5 s to 5.5s and from 5.5 s to 8.5 s for the first and the second data set respectively. For the estimation of reference matrices, the data segments for EA and REA were the same as those for the classification of mental tasks, whereas those for RA were intercepted from 6.25 s to 7.25 s and from 8.5 s to 9.5 s for the first and the second data set respectively.

IV. RESULTS

In this section, the proposed classification models are analysized from five aspects, i.e. data visualization, the distribution of channel weights, classification performance, regularization parameter and the number of chosen source subjects. Classification models built with REA trials are compared with those built with the raw and RA or EA trials in the same space and under two conditions, i.e. with and without the selection of relevant source subject.

A. DATA VISUALIZATION

t-stochastic neighbor embedding (t-SNE) [42] was used as the visualization tool. It reduces the dimensionality of the EEG data to two or three so that the trials can be visualized in two

or three dimensions. The perplexity parameter was set at 50 in all these t-SNE applications. We display the difference in data distribution among the raw data and the aligned data by RA, EA and REA. It is noted that the regularization parameter λ in the REA is different for different alignment spaces. Thereby, the data distributions of the raw data, the RA data and the REA data are compared in the Riemannian space, whereas those of the raw data, the EA data and the REA data are compared in Euclidean space.

Fig. 3 and Fig. 4 show the data distributions from two representative subjects in Riemannian and Euclidean space respectively. In each subplot, the red dots denote the EEG trials from a target subject, whereas the blue dots stand for those from the source subjects. The two figures illustrate the effect of cross-subject shifts caused by data alignment. All the three alignment methods change the distributions of EEG trials compared with the raw data. For the raw data, red and blue dots are very dispersive. A classification model built
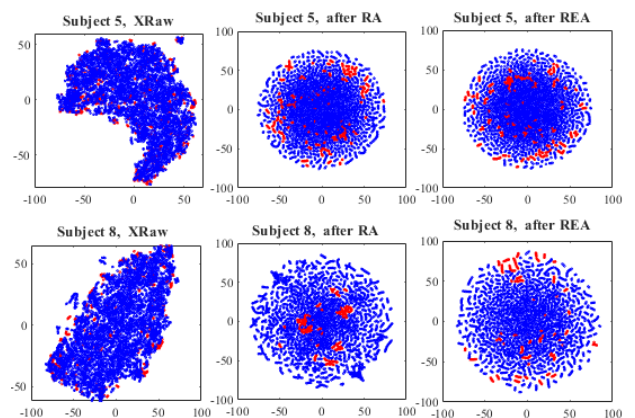

FIGURE 3. t-SNE visualization of the raw data, the RA data and the REA data from subject 5 in data set 1 (the first row) and subject 8 in data set 2 (the second row). The regularization parameter λ in REA was determined by a classification model based on SCM for feature extraction and MDM for classification.
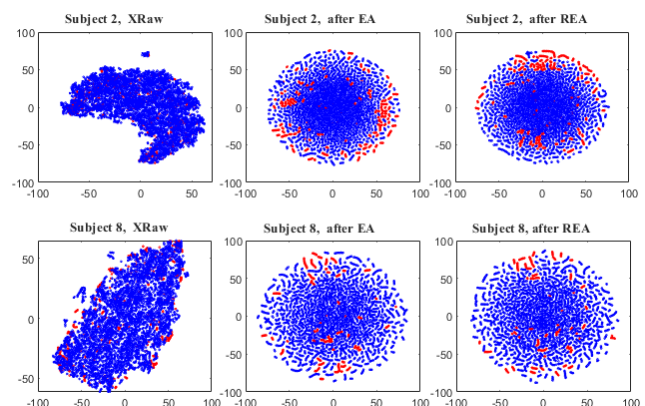

FIGURE 4. t-SNE visualization of the raw data, the EA data and the REA data from subject 2 in data set 1 (the first row) and subject 8 in data set 2 (the second row). The regularization parameter λ in REA was determined by a classification model based on CSP for feature extraction and LDA for classification.

with the blue trials may not classify the red trials very well. After RA, EA or REA, the training and testing trials overlap more with each other and the relative distance between them is reduced accordingly. The difference between RA/EA and REA is that the EA/RA trials are more concentrated than REA trials. The lower concentration of REA trials facilitates their classification.

B. DISTRIBUTION OF ELECTRODE WEIGHTS

We show the distribution of electrode weights from a subject in data set 1 because data set 2 does not provide the information of electrode labels. Fig. 5 shows the electrode weights for corresponding spatial filters evoked by MI of left hand and right hand for subject 4. The spatial filters were estimated by the CSP algorithms from three types of data, i.e. raw data, EA data and REA data, under two conditions, i.e. before and after the selection of source subjects. It is seen from the figure that for the raw data, the distributions of electrode weights of both left- and right-hand class roughly respect their representative regions, but strong weights from right-hand class scatter in left temporal lobe. After data alignment by EA and REA, strong weights of right-hand class scatter in two temporal lobes although the weight distributions of left-hand class are improved largely. After the selection of source subjects, the weight distribution of left-hand class is further improved for the EA data, but that of right-hand is not. The weight distributions of both classes for raw data and REA data are perfect from a neurophysiological point of view.
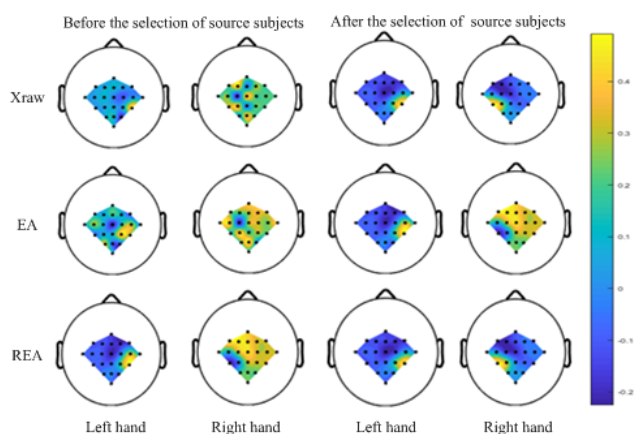

FIGURE 5. The electrode weights for corresponding spatial filters evoked by MI of left and right hand for subject 4 in data set 1. The spatial filters were estimated by the CSP algorithms with three types of data, i.e. raw data, EA data and REA data, and two conditions, i.e. before and after the selection of source subjects.

C. CLASSIFICATION ACCURACY

The classification model based on transfer learning can be created in either Euclidean or Riemannian space and thus there are two triads of algorithms for comparison in each space depending on whether the algorithm includes the selection of source subjects. Since the performance of a classification model is mainly determined by the quality of its

feature signals, these algorithms are thereby named according to their feature extraction methods. In Euclidean space, one triad is the basic CSP (CSP), the CSP with aligned data by EA (EA-CSP) and by REA (REA-CSP), and the other triad is the above three algorithms including subject selection (SS), named respectively as SS-CSP, EA-SS-CSP and REA-SS-CSP. In Riemannian space, a SCM is indeed served as a feature signal, and thus the two triads of algorithms are named around SCM. Accordingly, one triad is SCM, RA-SCM and REA-SCM, and the other is SS-SCM, RA-SS-SCM and REA-SS-SCM. The classification accuracies of the target subjects in data set 1 and 2 are illustrated in the Fig. 6 and Fig. 7 respectively.

From the two figures, it is observed that without the selection of source subjects (the first row of the two figures), all alignment-based algorithms outperform their original ones for most subjects. In terms of the averaged accuracy across subjects, the REA based algorithms consistently outperform the other two in each subplot. Specifically, for data set 1, REA-CSP outperforms CSP and EA-CSP by about 8% and 2% respectively; REA-SCM outperforms SCM and RA-SCM by about 17% and 3% respectively. For data set 2, REA-CSP outperforms CSP and EA-CSP by about 9% and 2% respectively; REA-SCM outperforms SCM and RA-SCM by about 13% and 14% respectively. These results confirm that as a method for data alignment, REA has superior performance compared to RA and EA and should be used to deal with individual differences among subjects.

With the selection of source subjects (the second raw of the two figures), the classification accuracies of all subjects are increased significantly. However, for different algorithms, the increase in accuracy is different. The worse the performance of an algorithm, the larger is the increase in accuracy. Specifically, REA-SS-CSP and EA-SS-CSP achieved approximately the same accuracy of 80% and 77%, and 3% and 4% higher accuracy than SS-CSP, for data set 1 and 2 respectively; REA-SS-SCM and RA-SS-SCM, however, behaved differently for the two data sets. For data set 1, REA-SS-SCM achieved 1% lower accuracy than RA-SS-SCM, but still 3% higher accuracy than SS-SCM, whereas for data set 2, REA-SS-SCM achieved 5% and 4% higher accuracy than RA-SS-SCM and SS-SCM respectively. These results verify that in MI based BCIs, subject selection is also very important for transfer learning.

To compare the difference between each of the four REA methods and the other two in each triad, the paired-sample t-tests were applied on their accuracy rates in Fig. 6 and 7 respectively. In all the t-tests, the significance level was set as $\alpha = 0.05$. The results of the statistical tests are shown in Table 1. It is observed from the table that REA-CSP and REA-SCM are significantly better than the other two in the same one triad with $p < 0.05$. Including the selection of relevant source subjects, REA-SS-CSP is significantly better than SS-CSP with $p < 0.05$ and $p < 0.01$ for data set 1 and 2 respectively, but there is no significant difference between REA-SS-CAP and EA-SS-CAP for the two data sets.
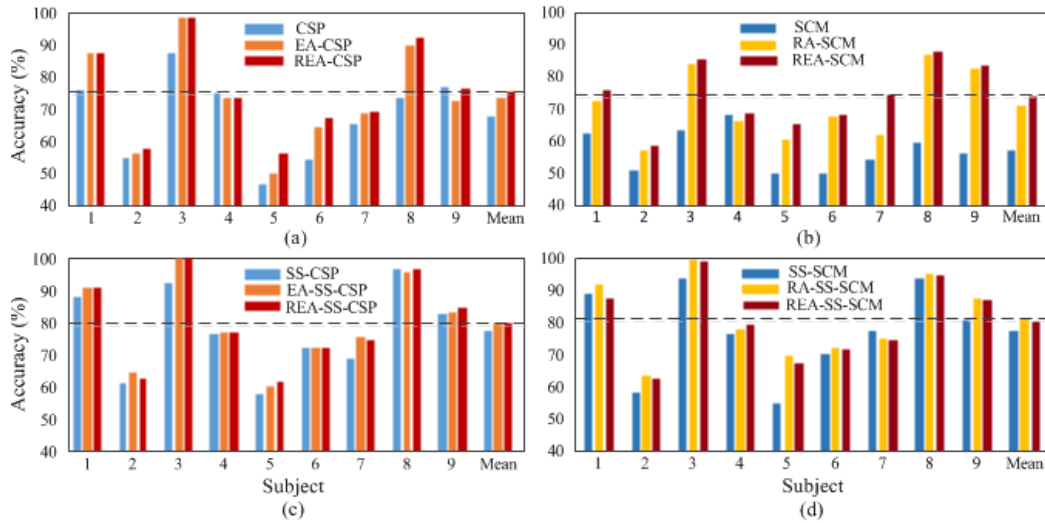
FIGURE 6. Classification accuracy rates achieved by the six CSP based algorithms (a, c) and the six SCM based algorithms (b, d) designed in Euclidean and Riemannian space respectively on Data set 1.

FIGURE 7. Classification accuracy rates achieved by the six CSP based algorithms (a, c) and the six SCM based algorithms (b, d) designed in Euclidean and Riemannian space respectively on Data set 2.
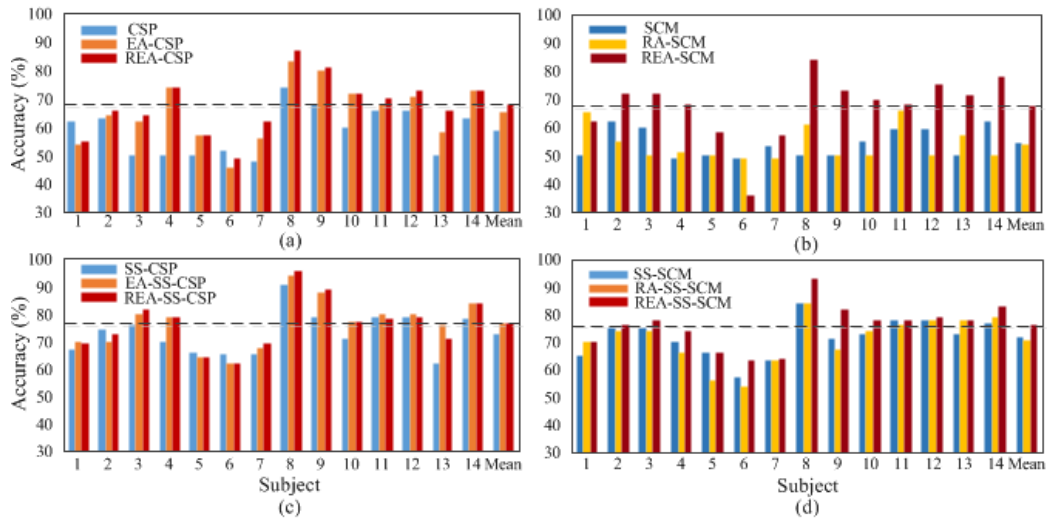
REA-SS-SCM is significantly better than SS-SCM and RA-SS-SCM with $p < 0.001$ and $p < 0.01$ respectively for the data set 2, but there is no significant difference among them for data set 1.

Fig. 8 depicts the averaged classification accuracy across all the six methods in the same raw of Fig. 6 and 7 for each source subject in the two data sets. In each subplot, the blue curve denotes the accuracy achieved without the selection of relevant source subjects, whereas the red curve denotes that obtained with the selection of relevant source subjects. It is clearly seen that the two red curves lie above the blue curve with a large interval between them for each subject in the two data sets. This is further proved that subject selection is effective and meaningful for transfer learning in MI based BCIs.

Fig. 9 shows the evolution of averaged accuracy rates across subjects of the six algorithms in each column with

TABLE 1. Results of paired t-tests of statistical significance between each of the four REA based methods and the other two methods in each triad of Fig. 6 and Fig. 7 for the two data sets.

	Data set 1		Data set 2	
	CSP	EA-CSP	CSP	EA-CSP
REA-CSP	p<0.01	p<0.05	p<0.001	p<0.01
	SCM	RA-SCM	SCM	RA-SCM
REA-SCM	p<0.001	p<0.05	p<0.001	p<0.001
	SS-CSP	EA-SS-CSP	SS-CSP	EA-SS-CSP
REA-SS-CSP	p<0.05	p=0.831	p<0.01	p=1.00
	SS-SCM	RA-SS-SCM	SS-SCM	RA-SS-SC M
REA-SS-SCM	p=0.071	p=0.076	p<0.001	p<0.01

the number of testing trials for the two data sets in each raw. The results in the first raw are derived from data set 1 with total trials of 144 per subject, and those in the second
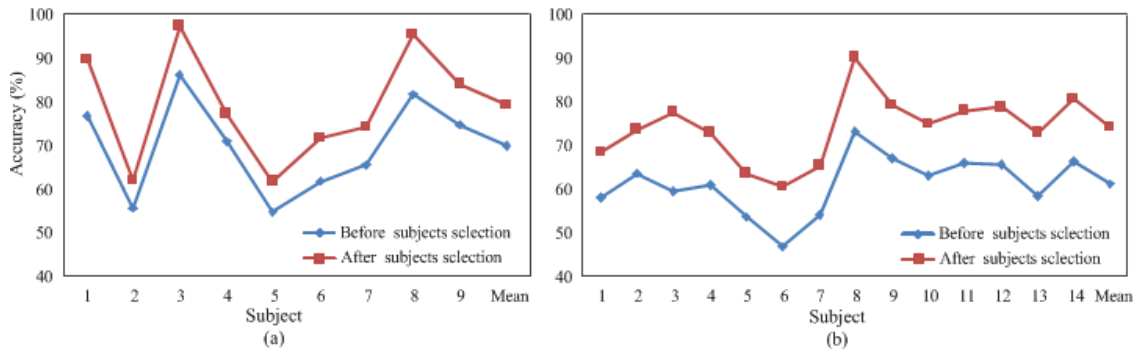
FIGURE 8. Averaged classification accuracy across the six methods of each subject before and after selecting relevant transfer subjects on (a) Data set 1; (b) Data set 2.
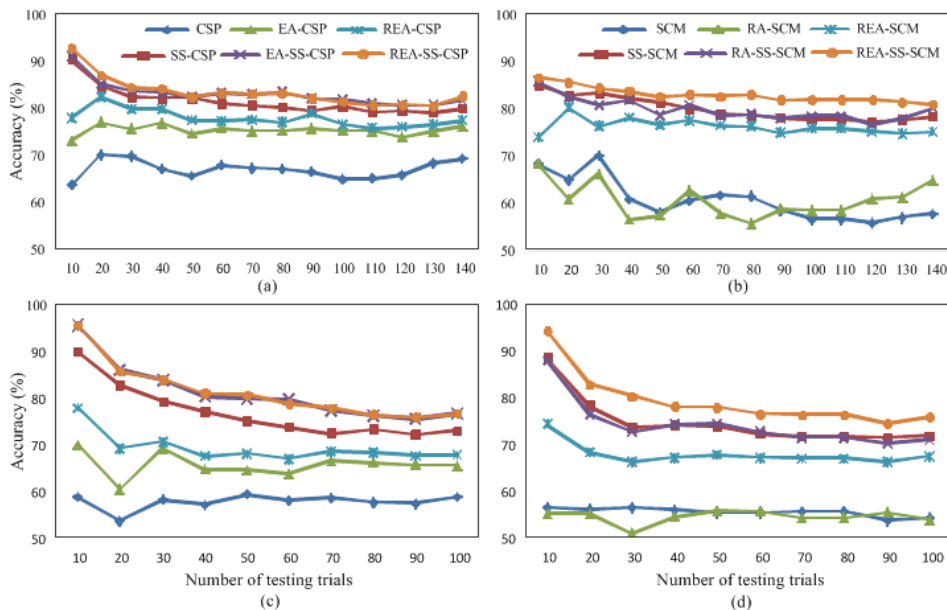

FIGURE 9. The evolution of averaged classification accuracies across subjects yielded by the six CSP based algorithms (a, c) and the six SCM based algorithms (b, d) with the number of testing trials on data set 1 (a, b) and data set 2 (c, d).

raw are derived from data set 2 with total trials of 100 per subject. In each subplot, the three accuracy curves with subject selection are consistently lying above those without subject selection. Looking carefully at those curves, it is revealed that within the initial dozens of testing trials, the lower three curves fluctuate up and down, whereas the upper three decline monotonically. The reason may be that when the target subjects had only a small amount of data, the reference matrix used for data alignment was not accurately estimated for all the six algorithms; The selection of source subjects alleviated the problem to some extent, but meanwhile brought overlearning. When the number of testing trials is greater than 40, all the six accuracy curves tend to be stable.

D. REGULARIZATION PARAMETER
Table 2 reports the regularization parameter λ in REA selected by each of the four methods, i.e. REA-SCM, REA-SS-SCM, REA-CSP and REA-SS-CSP, for each target

subject in the two data sets. The parameter λ for the former two methods is the same, so is it for the latter two. For data set 1, it is observed that if the classification model was built in Riemannian space (REA-SCM and REA-SS-SCM), most subjects selected Riemannian mean ($\lambda = 1$) as the reference matrix for data alignment; otherwise most subjects selected hybrid Riemannian and Euclidean mean ($\lambda \neq 0, 1$) as the reference matrix. For data set 2, the situation was different. If the classification model was built in Riemannian space, most subjects selected hybrid Riemannian and Euclidean mean ($\lambda \neq 0, 1$) as the reference matrix for data alignment; otherwise 7, 4 and 3 subjects selected hybrid Riemannian and Euclidean mean ($\lambda \neq 0, 1$), Riemannian mean ($\lambda = 1$) and Euclidean mean ($\lambda = 0$) as the reference matrix respectively. These results confirm that either RA or EA is not able to achieve the best effect of data alignment and REA should be used to deal with the problem of individual differences among subjects.

TABLE 2. The regularization parameter λ in REA selected by each of the four methods, i.e. REA-SCM, REA-SS-SCM, REA-CSP and REA-SS-CSP, for each target subject in the two data sets.

Method	Data set 1/Target subject								
	1	2	3	4	5	6	7	8	9
REA-SCM/REA-SS-SCM	1	1	1	0.8	0.4	1	1	0.2	1
REA-CSP/REA-SS-CSP	1	0.4	1	1	0.5	0.7	0.5	0.9	0.8

Method	Data set 2/Target subject													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
REA-SCM/REA-SS-SCM	1	0.6	0.8	0	0.7	0.8	0.1	1	1	0.6	1	0.5	0.3	1
REA-CSP/REA-SS-CSP	0	0	0.1	1	1	0.1	0	0.3	0.1	1	0.2	0.4	0.5	1

TABLE 3. The number of relevant source subjects selected by the six methods, i.e. SS-SCM, RA-SS-SCM, REA-SS-SCM, SS-CSP, EA-SS-CSP and REA-SS-CSP, for each target subject in the two data sets. M denotes Mean.

Method	Data set 1/Target subject									
	1	2	3	4	5	6	7	8	9	M
SS-SCM	3	3	2	2	2	3	5	3	3	2.9
RA-SS-SCM	2	2	3	4	4	3	3	2	5	3.1
REA-SS-SCM	1	2	3	2	6	3	5	2	2	2.9
SS-CSP	2	5	4	3	1	5	4	6	5	3.9
EA-SS-CSP	7	4	4	3	2	4	4	3	5	4.0
REA-SS-CSP	7	5	4	3	2	3	4	4	3	3.9

Method	Data set 2/Target subject														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	M
SS-SCM	3	6	2	6	3	4	2	3	6	3	6	4	5	4	4.1
RA-SS-SCM	6	6	4	8	3	3	3	3	5	6	5	3	6	5	**4.7**
REA-SS-SCM	6	5	3	3	7	1	4	3	4	6	6	1	6	7	**4.4**
SS-CSP	4	6	4	3	4	1	3	5	5	4	4	10	8	2	4.5
EA-SS-CSP	5	3	5	7	4	1	4	9	6	10	5	8	3	2	**5.1**
REA-SS-CSP	7	4	5	7	4	1	6	4	7	10	4	3	5	2	**4.9**

E. THE NUMBER OF CHOSEN SOURCE SUBJECTS

Table 3 reports the number of relevant source subjects selected by the six methods, i.e. SS-SCM, RA-SS-SCM, REA-SS-SCM, SS-CSP, EA-SS-CSP and REA-SS-CSP, for each target subject in the two data sets. It is revealed from the table that for the two data sets, the averaged number of relevant source subjects across target subjects selected by alignment-based algorithms is larger than that selected by original algorithms except for REA-SS-SCM and REA-SS-CSP, which selected the same number of source subjects as SS-SCM and SS-CSP respectively in data set 1. Comparing the two data sets, each algorithm in data set 2 selected more source subjects than its counterpart in data set 1. These results suggest that data alignment does increase the similarity between a source subject and the target subject, and a data set with more subjects is beneficial for transfer learning.

V. DISCUSSION AND CONCLUSION

Minimizing or suppressing training time in MI based BCI systems is a hard task. Transfer learning is a potential approach for decreasing training time that requires in-depth research. Data alignment and the selection of source subjects are two existing methods for improving transfer learning in

MI based BCIs. To the best of our knowledge, no research has combined the two methods for transfer learning. This paper proposes a novel transfer learning algorithm, which incorporates a new method for data aliment and a method for subject selection into the classification model. The hybrid Riemannian and Euclidean space data alignment (REA) method is used to reduce the variability across subjects, and hence makes transfer learning better applied to a BCI. Besides, the subject selection (SS) algorithm aims to select relevant transfer subjects for a target subject, in order to further decrease the difference between a chosen previous user and a new user.

Data alignment enhances the similarity between EEG signals of different subjects, but changes topographic representations of the spatial filters for different mental tasks. As shown in the first two columns of Fig. 5, this change is sometimes desired for a certain category, e.g. the first column, but sometimes undesired for another category, e.g. the second column, where both EA and REA deteriorate the topographic representation of the right-hand class compared to the raw data. The selection of relevant source subjects overcomes the problem to varying degrees, as shown in last column. From Fig. 9, it is observed that with only 40 trials, 20 trials per class,

the relatively high accuracy rates of about 80% and 75% can be achieved for a target subject in data set 1 and 2 respectively.

In MI based BCIs, a typical classification model created in Euclidean space is based on CSP, whereas that created in Riemannian is based on SCM. Both CSP and SCM rely on the estimation of sample covariance matrices, which are derived from a single-trial multi-channel EEG signal. When the number of channels is large, the dimension of covariance matrices is high. Working in high-dimensional space is computationally expensive or even impossible for very high dimensions. Moreover, when the dimension of SCMs is high, they may be badly conditioned with respect to inversion, i.e. their smallest eigenvalues are close to zero, harming the numerical stability of subsequent manipulations based on spectral functions of eigenvalues. Thus, it is necessary to reduce the dimension of SCMs. Channel selection is one of effective methods for dimensionality reduction. Recently, Jin *et al.* proposed two good methods for the purpose. One of them uses the sum of logarithmic amplitudes (SLA) and the first order spectral moment (FOSM) features extracted from bispectrum analysis to select EEG channels [43], whereas the other selects those channels containing more correlated information based on regularized feature optimization [44]. Both the two methods were applied to three MI data sets for BCI competition and the experimental results suggest that they achieved superior classification accuracy compared to existing methods.

Feature selection is another useful dimensionality reduction method. Recently, Jin *et al.* proposed an internal feature selection method [45] that selects discriminative features via suppressing outliers and discovering features with larger interclass distances. A fusion algorithm based on the Dempster–Shafer theory is used for internal feature selection. The method was evaluated on two MI based BCI data set and the experimental results show that the proposed method consumes less additional computational cost and results in a significant increase in the performance compared with other methods for feature selection.

RA aligns data with the reference matrix derived from resting-state EEG data, which might be affected by the healthy status of subjects or the environment. Thereby, when RA is used to preprocess EEG data from both source and target subjects, the effect of transfer learning might be compromised to some extent. Ment *et al.* first investigated the impacts of soft drinks such sugar, caffeine and regular coffee on resting state EEG and BCI performance [46]. The study shows that power in alpha and beta band after caffeine consumption were decreased substantially compared with control and sugar condition.

Two classification models were created with the data aligned by REA from chosen source subjects in Riemannian and Euclidean space respectively. Two MI data sets were employed to assess the classification performance of the two algorithms. The results show that REA significantly outperforms RA and EA, and the two algorithms combining REA and SS exhibits superior classification performance compared to existing algorithms. The proposed algorithms are helpful for improving transfer learning in BCIs and facilitate their applications in real world. This study analyzed the proposed algorithms offline. Future work will focus on the new method for the selection of relevant source subjects and the online implementation of the proposed algorithms.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtschller, and T. M. Vaughan, "Brain–computer interface for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: Control signals review," *Neurocomputing*, vol. 223, pp. 26–44, Feb. 2017.

[3] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[4] H. H. Ehrsson, S. Geyer, and E. Naito, "Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations," *J. Neurophysiol.*, vol. 90, no. 5, pp. 3304–3316, Nov. 2003.

[5] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[6] 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.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[8] A. P. Dawid, "Some matrix-variate distribution theory: Notational considerations and a Bayesian application," *Biometrika*, vol. 68, no. 1, pp. 265–274, 1981.

[9] S. T. Smith, "Covariance, subspace, and intrinsic Crame/spl acute/r-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.

[10] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.

[11] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, Dec. 2006.

[12] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, Oct. 2017.

[13] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, Sep. 2017.

[14] N. T. H. Gayraud, A. Rakotomamonjy, and M. Clerc, "Optimal transport applied to transfer learning for P300 detection," in *Proc. 7th Graz Conf.*, Graz, Austria, 2017, p. 6.

[15] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," *Ann. Inst. Stat. Math.*, vol. 60, no. 4, pp. 699–746, Dec. 2008.

[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[17] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 43, no. 3, pp. 155–174, 2017.

[18] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.

[19] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.

[20] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 614–617.

[21] J. Giles, K. K. Ang, L. S. Mihaylova, and M. Arvaneh, "A subject-to-subject transfer learning framework based on Jensen-Shannon divergence for improving brain-computer interface," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3087–3091.

[22] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2390–2401, Aug. 2019.

[23] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A Riemannian geometry framework with applications to brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.

[24] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 339–410, Feb. 2020.

[25] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*. Cambridge, MA, USA: MIT Press, 2009.

[26] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.

[27] P. Pudil, F. J. Ferri, and J. Novovicova, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. Int. Conf. Pattern Recognit., Signal Process.*, vol. 22, Oct. 1994, pp. 279–283.

[28] M. T. Harandi, M. Salzmann, and R. Hartley, "From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jul. 2014, pp. 17–32.

[29] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, Jan. 2005.

[30] M. Moakher, "A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.

[31] M. Moakher and P. G. Batchelor, "Symmetric positive-definite matrices: From geometry to applications and visualization," in *Visualization and Processing of Tensor Fields*. Berlin, Germany: Springer, 2006, pp. 285–298.

[32] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, Jan. 2006.

[33] W. Forstner and B. Moonen, "A metric for covariance matrix," in *Geodesy—The Challenge of the 3rd Millennium*, E. W. Grafarend, F. W. Krumm, and V. S. Schwarze, Eds. Berlin, Germany: Springer-Verlag, 2003, pp. 299–309.

[34] D. A. Bini and B. Iannazzo, "Computing the Karcher mean of symmetric positive definite matrices," *Linear Algebra Appl.*, vol. 438, no. 4, pp. 1700–1710, Feb. 2013.

[35] B. Afsari, "Riemannian L^P center of mass: Existence, uniqueness, and convexity," *Proc. Amer. Math. Soc.*, vol. 139, no. 2, pp. 655–673, 2011.

[36] B. Jeuris, R. Vandebril, and B. Vandereycken, "A survey and comparison of contemporary algorithms for computing the matrix geometric mean," *Electron. Trans. Numer. Anal.*, vol. 39, pp. 379–402, Apr. 2012.

[37] D. Wu, V. J. Lawhern, W. D. Hairston, and B. J. Lance, "Switching EEG headsets made easy: Reducing offline calibration effort using active weighted adaptation regularization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1125–1137, Nov. 2016.

[38] H. Cho, M. Ahn, K. Kim, and S. Chan Jun, "Increasing session-to-session transfer in a brain–computer interface with on-site background noise acquisition," *J. Neural Eng.*, vol. 12, no. 6, Dec. 2015, Art. no. 066009.

[39] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 12, no. 7, pp. 2058–2065.

[40] *BCI Competition IV, Data Sets 2A*. Accessed: Dec. 2008. [Online]. Available: http://www.bbci.de/competition/iv/

[41] *BNCI Horizon 2020 European Coordination Project, Data Sets 2*. Accessed: Apr. 2020. [Online]. Available: http://bnci-horizon-2020.eu/database/data-sets

[42] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[43] J. Jin, C. Liu, I. Daly, Y. Miao, S. Li, X. Wang, and A. Cichocki, "Bispectrum-based channel selection for motor imagery based brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2153–2163, Oct. 2020.

[44] 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 Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[45] J. Meng, J. H. Mundahl, T. D. Streitz, K. Maile, N. S. Gulachek, J. He, and B. He, "Effects of soft drinks on resting state EEG and brain–computer interface performance," *IEEE Access*, vol. 5, pp. 18756–18764, 2017.

[46] 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 Trans. Neural Netw. Learn. Syst.*, early access, Aug. 24, 2020, doi: 10.1109/TNNLS.2020.3015505.

YAN LI received the B.E. degree in communication engineering from Yichun University, China, in 2018. She is currently pursuing the M.E. degree in electronic and communication engineering. Her current research interest includes motor imagery (MI) based on brain–computer interface.

QINGGUO WEI received the B.E. degree in electronic engineering from Nanchang University, in 1984, the M.E. degree in communication and electronic engineering from the Beijing Institute of Technology, in 1991, and the Ph.D. degree in biomedical engineering from Tsinghua University, in 2007. He is currently a Professor with the Department of Electronic Information Engineering, Nanchang University. His current research interests include biomedical signal processing, brain–computer interface, and pattern recognition.

YUEBIN CHEN received the B.E. degree in biomedical engineering from Nanchang University, China. He is currently pursuing the M.E. degree in biomedical engineering. His current research interest includes steady-state visual evoked potential (SSVEP) based on brain–computer interface.

XICHEN ZHOU received the B.E. degree in software engineering from the Jiangxi University of Science and Technology, China. She is currently pursuing the M.E. degree in electronic and communication engineering. Her current research interest includes code modulated visual evoked potential (c-VEP) based on brain–computer interface.

• • •