



Selective Cross-Subject Transfer Learning Based on Riemannian Tangent Space for Motor Imagery Brain-Computer Interface

Yilu Xu¹, Xin Huang^{2*} and Quan Lan^{3*}

¹ School of Software, Jiangxi Agricultural University, Nanchang, China, ² Software College, Jiangxi Normal University, Nanchang, China, ³ Department of Neurology, First Affiliated Hospital of Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:

Yizhang Jiang,
Jiangnan University, China

Reviewed by:

Shipu Xu,
Agricultural Information Institute
of Science and Technology, Shanghai
Academy of Agricultural Sciences,
China

Yuanqing Xian,
Guangdong Ocean University, China

Wei Zhou,
Shenyang Aerospace University,
China

*Correspondence:

Xin Huang
xinhuang@jxnu.edu.cn
Quan Lan
xmdylanquan@163.com

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 18 September 2021

Accepted: 14 October 2021

Published: 03 November 2021

Citation:

Xu Y, Huang X and Lan Q (2021)
Selective Cross-Subject Transfer
Learning Based on Riemannian
Tangent Space for Motor Imagery
Brain-Computer Interface.
Front. Neurosci. 15:779231.
doi: 10.3389/fnins.2021.779231

A motor imagery (MI) brain-computer interface (BCI) plays an important role in the neurological rehabilitation training for stroke patients. Electroencephalogram (EEG)-based MI BCI has high temporal resolution, which is convenient for real-time BCI control. Therefore, we focus on EEG-based MI BCI in this paper. The identification of MI EEG signals is always quite challenging. Due to high inter-session/subject variability, each subject should spend long and tedious calibration time in collecting amounts of labeled samples for a subject-specific model. To cope with this problem, we present a supervised selective cross-subject transfer learning (sSCSTL) approach which simultaneously makes use of the labeled samples from target and source subjects based on Riemannian tangent space. Since the covariance matrices representing the multi-channel EEG signals belong to the smooth Riemannian manifold, we perform the Riemannian alignment to make the covariance matrices from different subjects close to each other. Then, all aligned covariance matrices are converted into the Riemannian tangent space features to train a classifier in the Euclidean space. To investigate the role of unlabeled samples, we further propose semi-supervised and unsupervised versions which utilize the total samples and unlabeled samples from target subject, respectively. Sequential forward floating search (SFFS) method is executed for source selection. All our proposed algorithms transfer the labeled samples from most suitable source subjects into the feature space of target subject. Experimental results on two publicly available MI datasets demonstrated that our algorithms outperformed several state-of-the-art algorithms using small number of the labeled samples from target subject, especially for good target subjects.

Keywords: transfer learning, cross-subject, source selection, Riemannian tangent space, motor imagery

INTRODUCTION

A motor imagery (MI) brain-computer interface (BCI) has drawn great attention for decades, since it can help a subject directly manipulate an electronic equipment using his brain activity evoked by imagined movements, without the participation of the traditional muscle-dependent pathway (He et al., 2015). It can not only help the stroke patients recover their neurological disorders, but also

give able-bodied people a novel way to control an external device (Wolpaw et al., 2002). Therefore, it plays an important role in rehabilitation engineering, military, and entertainment, etc.

Non-invasive electroencephalogram (EEG) is a popular recording modality in MI BCI due to its high safety and high temporal resolution, which is extremely crucial for the application of real-time BCI. However, MI EEG data analysis is quite challenging. The reason is that EEG signals are inherently weak, non-stationary, and easily contaminated by interference and noise (Sample et al., 2019). Moreover, compared with other traditional EEGs, such as event-related potentials (ERP) and steady-state visual evoked potentials (SSVEP), MI EEG signals have less obvious features, because they are evoked by spontaneous movement imagination without external stimulus, whereas ERP and SSVEP are evoked by some external stimulation (Blankertz et al., 2011; Yin et al., 2015). Consequently, MI EEG signals have higher inter-session/subject variability and fewer categories of BCI tasks than ERP and SSVEP. In MI BCI, each subject needs a tedious and annoying calibration time for a subject-specific classifier before performing real-time BCI tasks. A retraining session always increases user frustration.

Deep learning is a promising machine learning technique, which has been widely used in natural language processing and computer vision and so on Huang et al. (2019, 2020). However, it cannot be directly applied to the small training set scenario.

Transfer learning (TL) and semi-supervised learning (SSL) have been tried to reduce the need of abundant labeled samples. TL transfers the labeled samples from different source domains into the target domain. A domain denotes a subject, a session, a task, or a device. SSL simultaneously utilizes the labeled and unlabeled samples from the same subject. Both TL and SSL can shorten the calibration effort for the target subject by utilizing the samples as much as possible.

Here, we pay more attention to TL since it can use more samples than SSL. Generally, TL can be categorized into three groups, including instance TL (ITL), feature-representation transfer (FRT), and parameter TL (PTL) (Pan and Yang, 2010). Instance TL aims to transfer parts of the data in the source domains by reweighting. FRT tries to transfer a good feature representation for the target domain. PTL successively transfers and updates some parameters of the model under the assumption that the source and target domains share these parameters.

Among these groups, ITL approaches are most popular since they are easy to implement. Specifically, they can be divided into two categories.

The first category is filter based ITL, which aims to reweight the labeled sets from different domains based on the filters. In MI BCI, the common spatial patterns (CSP) approach is only effective when the labeled samples from a subject are abundant (Ramoser et al., 2000). Recently, regularized CSP (RCSP) approaches, belonging to filter based ITL, have been designed for the small training set scenario. In the framework of RCSP, the filtered EEG samples from the source and target subjects are separately reweighted based on their similarities which are always measured using different metrics, such as kullback-leibler (KL) divergence, Frobenius norm, cosine distance and so on (Kang et al., 2009; Cheng et al., 2017; Xu et al., 2019). A RCSP

based on dynamic time warping (DTW-RCSP) approach aligned the labeled samples from all source subjects to the average of a few target samples from the same class (Azab et al., 2020). In the module of feature extraction, RCSP approaches can generate more reliable CSP spatial filters for the target subject by effectively utilizing the labeled samples from the source subjects. However, calculating the optimal regularization terms might impose computational burden.

The second category is data alignment based ITL. On past decade, Riemannian alignment- (RA-) based ITL approaches have drawn a growing attention in EEG-based BCI since affine transformation can make the covariance matrices from different domains similar. Zanini et al. (2018) proposed a RA-based ITL approach to center the covariance matrices from each domain with respect to their reference matrix. Such reference matrix is the Riemannian mean of the covariance matrices of some resting trials in the corresponding domain. All re-centered covariance matrices from different source domains are concatenated altogether to train a minimum distance to mean (MDM) classifier based on Riemannian Gaussian distributions. Rodrigues et al. (2019) proposed a Riemannian procrustes analysis (RPA) approach using the steps of re-centering, stretching, and rotation for the covariance matrices from different domains to minimize the differences between the source and target domains as much as possible. Zhang and Wu (2020) presented a cross-subject manifold embedded knowledge transfer (MEKT) approach to boost zero-training for the target subject by combining the labeled samples from the source subjects with the unlabeled samples from the target subject. MEKT first performs RA to generate aligned covariance matrices for all subjects and then converts all aligned matrices into the tangent space feature vectors. Finally, MEKT finds optimal projection matrices for the tangent space feature vectors from the source and target subjects to reduce the joint probability distribution shift between subjects. Since most classifiers are designed for the Euclidean space instead of the Riemannian manifold, Euclidean alignment- (EA-) based approaches extend RA-based approaches in the Euclidean space using the Euclidean mean as the reference matrix (He and Wu, 2020). Both RA and EA can shorten the differences between the covariance matrices from the source and target subjects.

Inspired by MEKT, we focus on cross-subject TL based on Riemannian tangent space. In the calibration phase, labeled samples from target subject are successively processed by the modules of signal preprocessing, feature extraction, and classification to build a subject-specific classifier. To reduce the calibration time, the labeled samples from different source subjects can be transferred into the same modules for a more robust classifier.

Nevertheless, most cross-subject TL approaches utilize the labeled samples from all source subjects. Lotte and Guan (2011) designed a RCSP framework which evaluated the differences between the whole CSP feature set from all source subjects and the CSP feature set from the target subject, and then added two different weights for these two sets. Zhang and Wu (2020) found two optimal projection matrices for the whole tangent space vector set from all source subjects and the unlabeled tangent

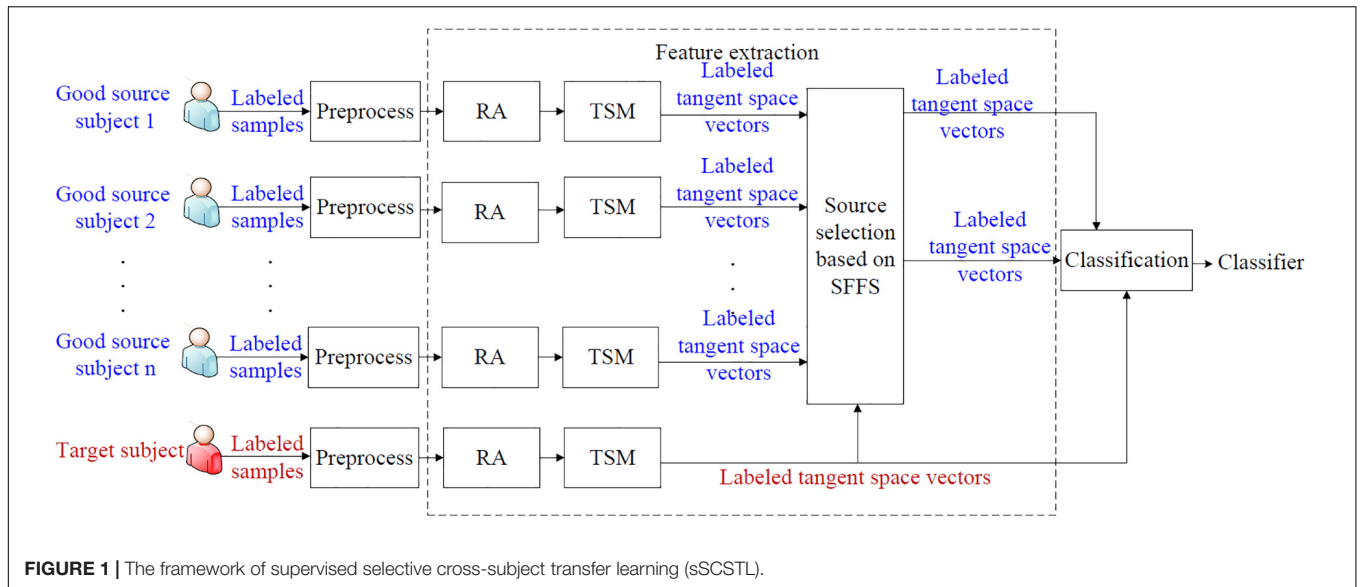


FIGURE 1 | The framework of supervised selective cross-subject transfer learning (sSCSTL).

space vector set from the target subject to lower their dimensions and differences.

To promote the positive transfer, the appropriate source subjects should be selected. Zanini et al. (2018) transferred all good source subjects (those with a precision higher than 0.75) for the target subject. To select suitable source subjects, Liang and Ma (2020) temporarily assigned the labeled samples from the target subject as the testing set, and the labeled samples from the candidate source subject as the training set. Then, the source subjects with higher classification accuracy were selected. Zhang and Wu (2020) presented a domain transferability estimation (DTE) approach based on MEKT, which selected the source subjects with the highest transferability.

In this paper, we first propose a supervised selective cross-subject TL (supervised SCSTL, sSCSTL) approach based on Riemannian tangent space for MI classification, which collects only a few labeled samples from target subject and transfers the labeled samples from most suitable source subjects. Moreover, we present an unsupervised SCSTL (uSCSTL) approach and a semi-supervised SCSTL (ssSCSTL) approach to investigate the role of the unlabeled samples from target subject. A sequential forward floating search (SFFS) method (Pudil et al., 1994) is used to iteratively select suitable source subjects for all SCSTL approaches.

The contribution of this paper lies in two aspects. Firstly, we iteratively select appropriate source subjects, instead of simply selecting them based on their classification performance, transferability and so on. Secondly, we consider SCSTL using different versions to exploit the labeled and unlabeled samples from target subject.

The remainder of this paper is structured as follows. In Section “Methods,” our SCSTL approaches are described in detail. In Section “Experimental Results,” the effectiveness of our approaches is validated by some experimental results. A discussion of results is presented in Section “Discussion.” Finally, our conclusions are drawn in Section “Conclusion.”

METHODS

We propose a supervised selective cross-subject transfer learning (sSCSTL) approach which utilizes the labeled samples from the selected good source subjects and the target subject simultaneously. In our opinion, the source subjects with good classification performance are more eligible to be selected than those with bad performance, since the labeled samples from the bad source subjects might mess up the feature distribution of the target subject. The framework of sSCSTL is shown in **Figure 1**.

In **Figure 1**, the labeled samples from each subject are band-pass filtered in the signal preprocessing module. During the feature extraction phase, the filtered labeled samples from each subject are first converted into the labeled covariance matrices. To preliminarily shorten the differences between subjects, the labeled covariance matrices from each subject are transformed into the aligned labeled covariance matrices by performing RA. To train a classifier in the Euclidean space, the aligned labeled matrices from each subject are converted into corresponding labeled tangent space vectors by executing tangent space mapping (TSM). To obtain best source selection, based on the SFFS method, the labeled tangent space vectors from different good source subjects are selected into the selected tangent space vector set by comparing with the labeled tangent space vector set from the target subject. Finally, the labeled tangent space vectors from the most suitable good source subjects and the target subject are fed into the classification module to obtain a subject-specific classifier.

Next, we introduce the sSCSTL approach in detail.

Riemannian Alignment

In MI BCI, supervised CSP is the most popular feature extraction algorithm for single user. Covariance matrices from different classes are used to learn optimal spatial filters which can maximize the variance differences between two classes. Such

spatial filters are used to extract low-dimensional features for the labeled and unlabeled samples. If the labeled samples are noisy, and/or few, covariance matrices might generate unreliable spatial filters. In this fashion, covariance matrices are processed in the Euclidean space. However, they belong to a smooth Riemannian manifold of symmetric positive definite (SPD) matrices, instead of Euclidean space. Thus, in this paper, we first handle covariance matrices in the differentiable Riemannian manifold.

All SPD matrices can form a Riemannian manifold. Here are the basic concepts of SPD matrix and its manifold. Let P_i and P_j be the i th SPD matrix and the j th SPD matrix, respectively, where $P_i, P_j \in \mathbb{R}^{N_c \times N_c}$ and N_c is the number of channels. Their manifold is $N_c \times (N_c + 1)/2$ dimensional. P_i and P_j can be regarded as the points of the manifold. The Riemannian distance $\delta(P_i, P_j)$ between P_i and P_j is the length of the minimum curve connecting them. It can be calculated by Moakher (2005).

$$\delta(P_i, P_j) = \|\log(P_i^{-1}P_j)\|_F = \left[\sum_{k=1}^{N_c} \log^2 \lambda_k \right]^{1/2} \quad (1)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\{\lambda_k\}_{k=1}^{N_c}$ are the eigenvalues of $P_i^{-1}P_j$. There exist important properties for the Riemannian distance as follows:

$$\delta(P_i, P_j) = \delta(P_j, P_i)$$

$$\delta(P_i, P_j) = \delta(P_i^{-1}, P_j^{-1})$$

$$\delta(P_i, P_j) = \delta(W^T P_i W, W^T P_j W), \forall W \in \text{Gl}(N_c)$$

The third property is crucial for the context of signal processing. It is named congruence invariance which means that the distance between the two SPD matrices is invariant after affine transformation using an invertible matrix. Suppose that $\text{Gl}(N_c)$ is the set of all $N_c \times N_c$ invertible matrices belonging to the space of square real matrices. The center of all SPD matrices is generally used for affine transformation, which can be calculated using Riemannian distance.

The Riemannian mean M_R , namely geometric mean, is the center point of the manifold which can be calculated using Riemannian distance as follows:

$$M_R = \underset{M}{\operatorname{argmin}} \sum_{k=1}^N \delta^2(P_k, M), \quad (2)$$

where N denotes the number of all SPD matrices on the manifold. There is no closed-form way to calculate the Riemannian mean. An iterative method can be used to effectively obtain the Riemannian mean (Fletcher and Joshi, 2004).

We can perform affine transformation using the center of all SPD matrices as the reference matrix. RA executes the following transformation using the Riemannian mean M_R as the reference matrix:

$$\hat{P}_i = M_R^{-1/2} P_i M_R^{-1/2}. \quad (3)$$

Likewise, EA extends RA using the Euclidean mean M_E as the reference matrix (He and Wu, 2020). The Euclidean mean M_E ,

namely arithmetic mean, is defined as the center point which can minimize the sum of squared Euclidean distances between other matrices and the center point as below:

$$M_E = \underset{M}{\operatorname{argmin}} \sum_{k=1}^N d^2(P_k, M) = \frac{1}{N} \sum_{k=1}^N P_k. \quad (4)$$

where $d(\cdot, \cdot)$ computes the Euclidean distance between the two SPD matrices. Since the Euclidean mean can be easily computed, it is widely used so far.

In this paper, we perform RA for all covariance matrices from each subject since RA is more suitable for the Riemannian manifold.

For RA-based ITL approaches, all aligned covariance matrices from each domain are centered at the identity matrix after affine transformation using each domain's Riemannian mean as the reference matrix. This property makes the aligned covariance matrices from different domains comparable, which can be testified by the following (Zhang and Wu, 2020).

$$\begin{aligned} & \mathcal{M}(M_R^{-1/2} P_1 M_R^{-1/2}, M_R^{-1/2} P_2 M_R^{-1/2}, \dots, M_R^{-1/2} P_N M_R^{-1/2}) \\ &= M_R^{-1/2} \mathcal{M}(P_1, P_2, \dots, P_N) M_R^{-1/2} \\ &= M_R^{-1/2} M_R M_R^{-1/2} = I \end{aligned} \quad (5)$$

where $\mathcal{M}(P_1, P_2, \dots, P_N)$ is the Riemannian mean operation, and I is an identity matrix.

Moreover, each aligned covariance matrix is nearly whitened since it is approximately an identity matrix. More details can be seen in Zhang and Wu (2020). Consequently, RA can preliminarily reduce the inter-domain differences.

Tangent Space Mapping

After RA, the aligned covariance matrices are usually input to an MDM classifier (Zanini et al., 2018; Rodrigues et al., 2019). However, most classifiers, such as linear discriminant analysis (LDA) and support vector machine (SVM), are designed for the Euclidean space. Thus, we perform tangent space mapping (TSM) for all aligned matrices from each subject.

As mentioned above, all SPD matrices lie in a differentiable Riemannian manifold. Then their derivatives at a matrix on the manifold form a tangent space. The tangent space has the same dimensions as the manifold. A Riemannian manifold and its tangent space at a point can be illustrated in **Figure 2**.

As depicted in **Figure 2**, P_i and P_j belong to the Riemannian manifold. On the tangent space at point M , A_i and A_j are the derivatives of P_i and P_j , respectively. A_i can be regarded as the logarithmic mapping of P_i at M as below:

$$A_i = \log_M(P_i) = M^{1/2} \log(M^{-1/2} P_i M^{-1/2}) M^{1/2}. \quad (6)$$

Conversely, the exponential mapping of A_i at M is the SPD matrix P_i defined as:

$$P_i = \exp_M(A_i) = M^{1/2} \exp(M^{-1/2} A_i M^{-1/2}) M^{1/2}. \quad (7)$$

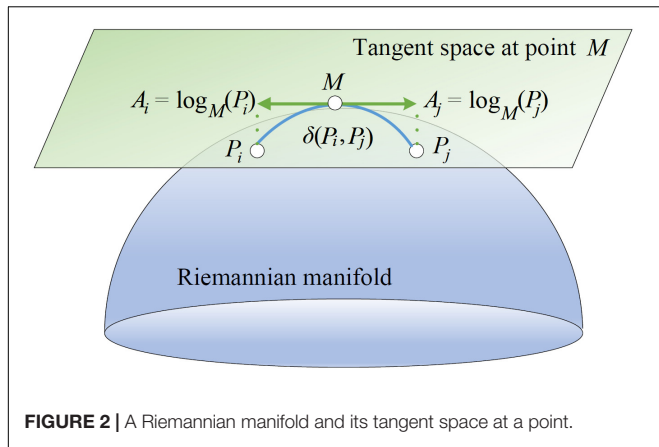


FIGURE 2 | A Riemannian manifold and its tangent space at a point.

Riemannian distance can be also defined as:

$$\begin{aligned} \delta(M, P_i) &= \|\log_M(P_i)\|_M = \|A_i\|_M \\ &= \|\text{vec}(M^{-1/2}A_iM^{-1/2})\|_2 = \|\text{vec}(\log(M^{-1/2}P_iM^{-1/2}))\|_2 \\ &= \|\text{vec}(\hat{A}_i)\|_2 = \|a_i\|_2, \end{aligned} \tag{8}$$

where $\|A_i\|_M$ is the norm of A_i on the tangent space at point M , and $\text{vec}(A_i)$ vectorizes a symmetry matrix \hat{A}_i . Let \hat{A}_i and a_i be $\log(M^{-1/2}P_iM^{-1/2})$ and $\text{vec}(\hat{A}_i)$, respectively. Without loss of generality, we apply a $\sqrt{2}$ coefficient for out-of-diagonal elements of \hat{A}_i and transform the upper triangular part of modified \hat{A}_i into a $N_c \times (N_c + 1)/2$ column vector a_i as below:

$$\begin{aligned} a_i &= \text{vec}(\hat{A}_i) = [\hat{A}_{i1,1}; \sqrt{2}\hat{A}_{i1,2}; \hat{A}_{i2,2}; \sqrt{2}\hat{A}_{i1,3}; \sqrt{2}\hat{A}_{i2,3}; \hat{A}_{i3,3} \\ &; \dots; \hat{A}_{iN_c,N_c}] \end{aligned} \tag{9}$$

where $\hat{A}_{ij,k} \in \hat{A}_i$. Then, a_i is the tangent space vector of the SPD matrix P_i at point M .

Furthermore, there is an approximation in terms of distance between the manifold and its tangent space as below (Tuzel et al., 2008):

$$\delta(P_i, P_j) \approx \|a_i - a_j\|_2 \tag{10}$$

where P_i and P_j are locally distributed into the manifold. Their tangent space vectors at point M are a_i and a_j , respectively. Such approximation stands only when M is the Riemannian mean of the manifold.

Therefore, Riemannian tangent space is Euclidean and locally homomorphic to the Riemannian manifold (Barachant et al., 2013). In the TSM phase of our proposed sSCSTL, all aligned covariance matrices $\{M_R^{-1/2}P_iM_R^{-1/2}\}_{i=1}^N$ from each domain are converted into corresponding tangent space vectors $\{a_i = \text{vec}(\log(M_R^{-1/2}P_iM_R^{-1/2}))\}_{i=1}^N$ using M_R as the reference matrix.

Source Selection Based on Sequential Forward Floating Search

As mentioned above, due to high inter-subject variability and expensive computational burden, it is unsuitable to utilize the features from all source subjects. In this paper, we only transfer

the labeled tangent space vectors from most suitable good source subjects.

We define the transferability for the selected tangent space vector set Sel to predict its usefulness for the tangent space vector set \mathcal{D}_T from the target subject as in Zhang and Wu (2020):

$$\text{transferability}(Sel, \mathcal{D}_T) = \frac{\|S_b^{Sel}\|_1}{\|S_b^{Sel, \mathcal{D}_T}\|_1}, \tag{11}$$

where S_b^{Sel} is the between-class scatter matrix of Sel . Similarly, S_b^{Sel, \mathcal{D}_T} is the scatter matrix between Sel and \mathcal{D}_T . Then, $\|S_b^{Sel}\|_1$ is used for the discriminability of between-class of Sel , and $\|S_b^{Sel, \mathcal{D}_T}\|_1$ is used to evaluate the differences between the selected source subjects and the target subject.

Instead of simply integrating the labeled tangent space vector sets from all source subjects, the labeled tangent space vector set from the good source subject is iteratively selected to add into or remove from the selected tangent space vector set to maximize its transferability. Our proposed source selection procedure for sSCSTL uses the framework of SFFS to obtain convergence of selection. More details are shown in Algorithm 1.

Algorithm 1: The source selection procedure for sSCSTL

Input: N_{TL} labeled tangent space vectors from the target subject $\mathcal{D}_{TL} = \{a_i^{(TL)}, y_i^{(TL)}\}_{i=1}^{N_{TL}}$, N_{S_k} labeled tangent space vectors from the k th good source subject $\mathcal{D}_{S_k} = \{a_i^{(S_k)}, y_i^{(S_k)}\}_{i=1}^{N_{S_k}}$, $k = \{1, 2, \dots, N_{GS}\}$, where N_{GS} is the number of good source subjects. Let $y_i^{(TL)}$ and $y_i^{(S_k)}$ be the labels of tangent space vectors from target and source subjects, respectively.

Output: The finally selected tangent space vector set Sel ;

Initialize: The initially selected tangent space vector set $Sel_0 = \emptyset$ and its transferability $Trans_0 = 0$, the initially remaining tangent space vector set $Rem_0 = \{\mathcal{D}_{S_k}\}_{k=1}^{N_{GS}}$, $n = 1$;

Repeat

Step 1: The labeled tangent space vector set from the most suitable good source subject is selected and added into Sel_{n-1}

$$\mathcal{D}_{S_{good}} = \underset{\mathcal{D}_{S_k} \in Rem_{n-1}}{\text{argmax}} (\text{transferability}(Sel_{n-1} + \mathcal{D}_{S_k}, \mathcal{D}_{TL}));$$

$$Sel_n = Sel_{n-1} + \mathcal{D}_{S_{good}}; Rem_n = Rem_{n-1} - \mathcal{D}_{S_{good}};$$

$$Trans_n = \text{transferability}(Sel_n, \mathcal{D}_{TL});$$

$$n = n + 1;$$

Step 2: The labeled tangent space vector set from the most unsuitable good source subject is selected and removed from Sel_n

if $n > 2$ **then**

$$\mathcal{D}_{S_{bad}} = \underset{\mathcal{D}_{S_k} \in Sel_n}{\text{argmax}} (\text{transferability}(Sel_n - \mathcal{D}_{S_k}, \mathcal{D}_{TL}));$$

if $\text{transferability}(Sel_n - \mathcal{D}_{S_{bad}}, \mathcal{D}_{TL}) > Trans_{n-1}$ **then**

$$Sel_{n-1} = Sel_n - \mathcal{D}_{S_{bad}}; Rem_{n-1} = Rem_n + \mathcal{D}_{S_{bad}};$$

```

Transn-1 = transferability(Seln-1, DTL);
n = n - 1;
go to Step 2;

else go to Step 1;
end
end
Until n = NGS

Sel = Selarg maxk(Transk).
    
```

For sSCSTL, labeled covariance matrices from each subject are converted into labeled tangent space vectors using their Riemannian mean as the reference matrix.

In **Algorithm 1**, during each iteration, *Sel*_{*n*} and *Trans*_{*n*} separately denote the currently selected tangent space vector set from *n* good source subjects and its transferability, whereas *Rem*_{*n*} denotes the currently remaining tangent space vector set from (*N*_{GS} - *n*) good source subjects. After all iterations, the selected tangent space vector set with the highest transferability is chosen as the finally selected tangent space vector set *Sel*.

Since $\|S_b^{Sel, \mathcal{D}_T}\|_1$ in Equation (11) is evaluated without using the labels of *Sel* and *D*_T, we modify the source selection procedure in **Algorithm 1** for an unsupervised SCSTL (uSCSTL) approach and a semi-supervised SCSTL (ssSCSTL) approach to investigate the role of unlabeled samples from the target subject.

For uSCSTL, unlabeled covariance matrices from target subject are converted into corresponding tangent space vectors *D*_{TU} using their Riemannian mean as the reference matrix. Then, in each iteration, the currently selected tangent space vector set *Sel*_{*n*} is compared with *D*_{TU} to obtain its transferability (*transferability*(*Sel*_{*n*}, *D*_{TU})).

For ssSCSTL, all covariance matrices from target subject are projected into corresponding tangent space vectors (*D*_T = *D*_{TL} + *D*_{TU}) using their Riemannian mean as the reference point. Then, more tangent space vectors from target subject can be used to yield more convincing transferability (*transferability*(*Sel*_{*n*}, *D*_T)).

Classification

The supervised LDA classifier has been widely used because of simplicity and rapid computational speed, which is crucial for real time MI BCI. Nevertheless, it is of poor quality for high dimensional features, such as the tangent space vectors. Therefore, we use a supervised shrinkage LDA (sLDA) classifier (Peck and Van Ness, 1982) which not only inherits the advantages of LDA but also behaves well in high dimensions.

Our goal of sSCSTL, uSCSTL, and ssSCSTL is to extract discriminative features for the classifier. In our classification module, no unlabeled tangent space vectors from the target subject are utilized for the classifier. Therefore, for sSCSTL and ssSCSTL, the labeled tangent space vectors from the selected good source subjects and target subject are fed into the sLDA classifier after the best source selection based on SFFS, whereas for uSCSTL, only the labeled tangent space vectors from the selected good source subjects are used for sLDA.

EXPERIMENTAL RESULTS

Datasets

In this paper, the MI dataset from 52 healthy subjects (Cho et al., 2017) and the BCI competition IV dataset 1 (Tangermann et al., 2012) were used to evaluate the effectiveness of our proposed methods in MI classification.

- (1) MI dataset from 52 healthy subjects (MI1): 64-channel MI EEG signals from 52 healthy subjects were recorded at a sampling rate of 512 Hz. Each subject executed the assigned MI task when a random instruction (“left hand” or “right hand”) appeared on the screen for 3 s. Only a single session was performed for each subject, which consisted of five or six runs. In each run, the MI experiments per class were repeated 20 times. Consequently, there were a total of 200 or 240 trials per subject. This dataset was much suitable for cross-subject TL due to the existence of many subjects.
- (2) BCI competition IV dataset 1 (MI2): this dataset was comprised of the calibration data and the evaluation data. However, we only chose the calibration data for our experiments since the evaluation data contained the periods in which the subject had no control intention. For the calibration data, 59-channel EEG signals from seven healthy subjects (a, b, c, d, e, f, and g) were collected at a sampling rate of 100 Hz. Each subject was shown a visual cue for 4 s and performed the cued MI task (“left hand” or “foot”). In total, for each subject, 100 trials per class were gathered. All EEG signals were band-pass filtered between 0.05 Hz and 200 Hz.

Electroencephalogram Data Preprocessing

For dataset MI1, the BBCI toolbox (Blankertz et al., 2016) was used for preprocessing. All EEG signals from each subject were preprocessed by common average reference, spectrally filtered by a third order Butterworth filter with cutoff frequencies of 8 and 30 Hz, and temporally segmented from 0.5 to 2.5 s after the instruction onsets.

For dataset MI2, all EEG recordings from each subject were band-pass filtered between 8 and 30 Hz using a fiftieth order finite impulse response filter. Then, the filtered EEG recordings were extracted from the time interval between 0.5 and 3.5 s after the visual cue onsets.

As mentioned above, the dimensionality of tangent space vectors closely relates to the number of channels. High-dimensional tangent space vectors may lead to curse of dimensionality. Therefore, for datasets MI1 and MI2, 27 and 29 channels were separately selected from the sensorimotor areas. In **Figure 3**, the selected channels used for datasets MI1 and MI2 are marked in green.

Experimental Design

First, we selected the good subjects as the source subjects based on their high classification accuracies. In our experiments, the labels of their samples were assumed to be known already. All their samples were used for cross-subject TL. If one of the good

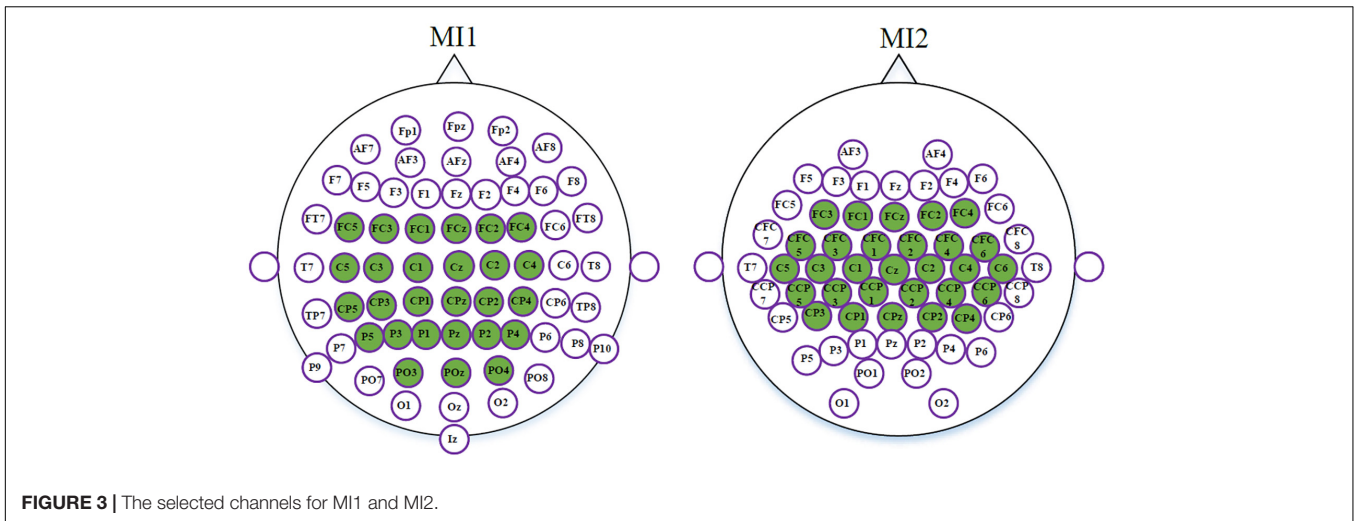


FIGURE 3 | The selected channels for MI1 and MI2.

subjects was the target subject, the remaining good subjects were his source subjects. All samples from each target subject were randomly partitioned into the labeled set for training and the unlabeled set for testing. To avoid randomness involved, this process was repeated ten times and the average accuracies of the unlabeled set were reported. The ratio of number of labeled samples to all samples (R_l) ranged from 10 to 40% with step 10%. Correspondingly, the ratio of number of unlabeled samples to all samples (R_u) varied from 90 to 60%.

Baseline Algorithms

We compared our proposed algorithms (sSCSTL, ssSCSTL, and uSCSTL) with the following baseline algorithms. They can be grouped into three categories according to their feature spaces:

- (1) Euclidean space algorithms:
 - (a) CSP (Ramoser et al., 2000): it was a traditional feature extraction algorithm for MI. We used LDA as its classifier.
 - (b) CSP with generic learning regularization (GLRCSP) (Lu et al., 2009): it learned generic learning regularized CSP filters to realize cross-subject TL. LDA was its classifier.
 - (c) Composite CSP (CCSP) (Kang et al., 2009): it weighed covariance matrices from different subjects using the KL divergence measure based on the framework of RCSP. LDA was also its classifier.
- (2) Riemannian space algorithm: all covariance matrices belong to the Riemannian manifold. We designed a covariance matrix -based feature extraction algorithm, named COV, which directly input all labeled covariance matrices into the MDM classifier without RA.
- (3) Riemannian Tangent space algorithms:
 - (a) Tangent space (TS) feature extraction algorithm (Barachant et al., 2012): it successively performed RA and TSM for all covariance matrices.

Then, the tangent space vectors were fed into the LDA classifier.

- (b) Manifold embedded knowledge transfer (MEKT) (Zhang and Wu, 2020): it combined the labeled tangent space vectors from the source subjects with the unlabeled tangent space vectors from the target subject. sLDA was used as its classifier.
- (c) Domain transferability estimation (DTE) (Zhang and Wu, 2020): it identified one most suitable source subject for the target subject based on MEKT. sLDA was also used for MI classification.

We set the hyper-parameters of all baseline algorithms instructed by their publications. Our proposed algorithms and seven baselines can also be further divided into two groups. The first group consists of CSP, GLRCSP, CCSP, COV, TS, sSCSTL, and ssSCSTL. They are all supervised or semi-supervised feature extraction algorithms. The second group includes MEKT, DTE, and uSCSTL. They are unsupervised feature extraction algorithms.

Classification Accuracies

Classification Performance on Dataset MI1

For a small percentage of subjects, their classification accuracies hardly reach the benchmark of 70%. This phenomenon has been named BCI illiteracy. It was suggested that at least 40 labeled samples per class can reduce BCI illiteracy (Blankertz et al., 2007). Therefore, for dataset MI1, we chose the good subjects whose accuracies were over 80% on average when 40 or 48 labeled samples per class ($R_l = 40\%$) were randomly selected for 10 times and trained by CSP and LDA.

In our experiments, we focused on the good and bad subjects to discuss how they influenced the BCI performance. Furthermore, we cared about the supervised, semi-supervised, and unsupervised algorithms to investigate the role of the labeled and unlabeled samples.

The classification performances of supervised and semi-supervised algorithms for good subjects on dataset MI1 are

given in **Table 1** using their 40 or 48 labeled samples per class ($R_l = 40\%$). Then, the performances of unsupervised algorithms for the same good subjects are shown in **Table 2** using their remaining unlabeled samples ($R_u = 60\%$). Meanwhile, the means and standard deviations (Std) of good subjects for two groups of algorithms are listed in **Tables 1, 2**, respectively.

As shown in **Table 1**, there were only seven good subjects selected from 52 subjects on dataset MI1 according to the performance of CSP. For all Euclidean space-based algorithms, there were no obvious differences among them. CSP slightly outperformed GLRCSP and lagged behind CCSP a little. Without RA, the Riemannian space-based algorithm COV performed the worst. For Riemannian tangent space-based algorithms, the mean of TS was lower than that of CSP. sSCSTL showed its superiority among all supervised algorithms. A paired *t*-test showed that the result of sSCSTL was statistically higher than that of CSP ($p = 0.0067$), GLRCSP ($p = 0.0055$), CCSP ($p = 0.0064$), COV ($p = 0.0052$), and TS ($p = 0.0077$). ssSCSTL stood out itself with the help of unlabeled samples. In **Table 2**, all unsupervised algorithms utilized the labeled samples from the selected source subjects and added the remaining unlabeled samples from the good target subject. MEKT surpassed most of supervised algorithms in **Table 1**, except for sSCSTL. Although DTE had the same framework of MEKT, it performed worse than MEKT. The possible reason was that there was only one suitable source subject selected. For uSCSTL, the average classification accuracies of six good subjects were up to 90%.

To further evaluate the effect of the labeled and unlabeled samples from the target subject, we plotted the means of accuracies of all good subjects on dataset MI1 for different algorithms with varying R_l or R_u in **Figure 4**.

As depicted in **Figure 4**, all supervised and semi-supervised algorithms provided good classification performances with varying R_l from 10 to 40%. Except for COV, most algorithms outperformed CSP when few labeled samples from target subject

were available. Moreover, the classification performances of most algorithms improved obviously along with the increasement of R_l . However, sSCSTL and ssSCSTL steadily showed compelling validity. In addition, the curves of unsupervised algorithms improved slowly with the increasement of R_u , suggesting that the labeled samples were more useful for the performance than the unlabeled samples. Nevertheless, even using 60% unlabeled samples from the target subject, the average classification accuracy was over 95% for uSCSTL.

Then, since there were 45 bad subjects on dataset MI1, we did not provide their average classification accuracies one by one. Instead, in **Tables 3, 4**, we separately investigated how many bad subjects were trapped in BCI illiteracy with varying R_l and R_u . Here, those whose average classification accuracies were less than 70% were regarded as unqualified subjects.

As shown in **Table 3**, Euclidean space-based algorithms sharply reduced the number of unqualified subjects when their labeled samples increased. It implied that they performed well when abundant labeled samples from target subject were available. In contrast, other algorithms reduced BCI illiteracy slowly with the increasement of R_l , suggesting that their performances did not excessively rely on the number of labeled samples. sSCSTL reduced more unqualified subjects than CCSP when R_l was 10%. In **Table 4**, the number of unqualified subjects barely changed with the increasement of R_u for all unsupervised algorithms. uSCSTL also reduced more unqualified subjects than MEKT when R_u was 60%.

Like **Figure 4**, the means of accuracies of all bad subjects on dataset MI1 for different algorithms with varying R_l or R_u are plotted in **Figure 5**.

As illustrated in **Figures 4, 5**, the advantages of sSCSTL and ssSCSTL for all bad subjects were less apparent than those for good ones. In **Figure 5**, there were few differences between sSCSTL and ssSCSTL. Possible explanation was that low class-discriminability of unlabeled samples from bad target subjects influenced the performance improvements of ssSCSTL. In addition, CCSP and GLRCSP still outperformed CSP, TS, and COV. Additionally, the average classification accuracies provided by uSCSTL were slightly lower than those of MEKT with varying R_u . The performance of DTE showed the importance of using source subjects as much as possible.

Classification Performance on Dataset MI2

On dataset MI2, five good subjects (a, d, e, f, and g) were chosen out of the seven subjects depending on the same rule as mentioned above. Then subjects b and c were regarded as bad subjects. The classification performances of supervised and semi-supervised algorithms for all subjects on dataset MI2 are shown in **Table 5** using their 40 labeled samples per class ($R_l = 40\%$). And the performances of unsupervised algorithms for the same subjects are given in **Table 6** using their remaining unlabeled samples ($R_u = 60\%$).

As shown in **Table 5**, among all Euclidean space-based algorithms, GLRCSP and CCSP outperformed CSP on average. COV still performed the worst due to without RA. For Riemannian tangent space-based algorithms, the average accuracy of TS was slightly lower than that of CSP. sSCSTL

TABLE 1 | Average classification accuracies of supervised and semi-supervised algorithms for good subjects on dataset MI1 ($R_l = 40\%$).

	3	4	14	23	41	43	48	Mean (Std)
CSP	89.50	82.50	96.92	81.33	83.67	96.92	82.33	87.60 (6.90)
GLRCSP	89.58	82.92	96.17	80.17	85.83	95.92	80.58	87.31 (6.78)
CCSP	90.08	84.33	96.92	80.75	85.50	96.08	82.08	87.96 (6.54)
COV	86.42	77.33	96.92	73.08	73.92	93.17	75.58	82.35 (9.78)
TS	90.17	77.92	96.25	80.83	80.50	96.08	82.83	86.37 (7.70)
sSCSTL	92.00	95.75	100.00	94.25	97.67	99.83	97.75	96.75 (2.94)
ssSCSTL	91.33	99.42	100.00	99.42	99.42	100.00	100.00	98.51 (3.18)

TABLE 2 | Average classification accuracies of unsupervised algorithms for good subjects on dataset MI1 ($R_u = 60\%$).

	3	4	14	23	41	43	48	Mean (Std)
MEKT	90.42	84.17	95.25	82.17	86.92	94.83	82.75	88.07 (5.51)
DTE	89.42	80.75	95.00	78.58	82.42	92.33	81.92	85.77 (6.38)
uSCSTL	88.83	97.17	99.50	93.83	97.50	100.00	99.17	96.57 (3.99)

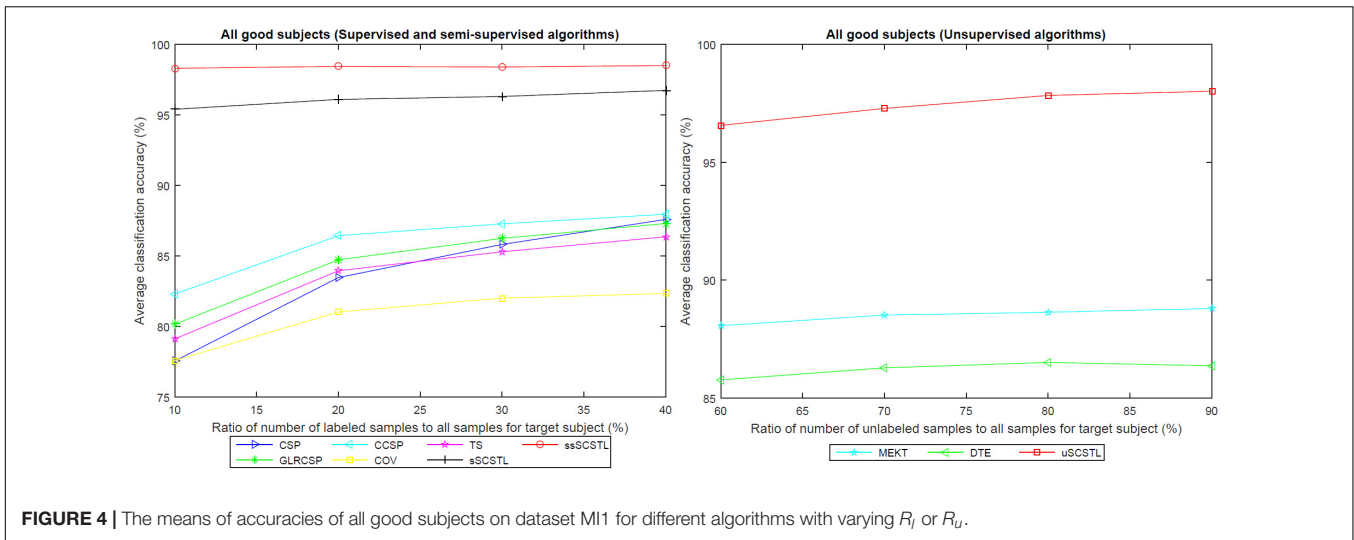


FIGURE 4 | The means of accuracies of all good subjects on dataset MI1 for different algorithms with varying R_l or R_u .

and ssSCSTL stood out themselves on average. However, for bad subjects, GLRCSP and CCSP were comparably superior to other algorithms. For most good subjects (d, e, f, and g), sSCSTL and ssSCSTL exhibited their superiorities. Paired *t*-test results between sSCSTL and other supervised algorithms showed that the performance improvements of sSCSTL over others were statistically significant, such as sSCSTL vs. CSP ($p = 0.1795$), sSCSTL vs. GLRCSP ($p = 0.2901$), sSCSTL vs. CCSP ($p = 0.3753$), sSCSTL vs. COV ($p = 0.0260$), and sSCSTL vs. TS ($p = 0.1631$). As reported in **Table 6**, although uSCSTL showed its compelling validity on average, MEKT performed better than uSCSTL for bad subjects.

To further evaluate the role of labeled and unlabeled samples, for different algorithms, we plotted the average classification

accuracies of all subjects on dataset MI2, and their means, with varying R_l and R_u in **Figures 6, 7**, respectively.

In the last subfigure of **Figure 6**, ssSCSTL outperformed other algorithms on average with the help of unlabeled samples from target subject. The gap between sSCSTL and ssSCSTL became small with the increase of labeled samples from target subject. For Euclidean space-based algorithms, CCSP and GLRCSP still performed better than CSP on average. Riemannian tangent space-based algorithm TS showed higher performance than Riemannian space-based algorithm COV.

Moreover, as depicted in other subfigures of **Figure 6**, sSCSTL and ssSCSTL showed amazing accuracies for most good subjects (d, e, f, and g), suggesting that the available samples from a good subject can provide positive cross-subject TL. Exceptionally, a good subject (a) benefited little from TL algorithms. Additionally, sSCSTL and ssSCSTL performed more unstably than CCSP and GLRCSP for bad subjects (b and c) in total.

As shown in **Figure 7**, the average classification accuracies of uSCSTL were inferior to those of MEKT and DTE for the good subject (a) and bad subjects (b and c). However, uSCSTL stood out itself for other subjects. Therefore, in the last subfigure of **Figure 7**, uSCSTL performed better than other unsupervised algorithms overall.

Data Visualization

To appreciate the effect of cross-subject TL, in **Figure 8**, the t-SNE method was applied to visualize the feature distributions of the following Riemannian space-based and Riemannian tangent space-based algorithms, including COV, TS, sSCSTL, ssSCSTL, and uSCSTL (Van Der Maaten and Hinton, 2008). In our case, each feature was regarded as a point in a two-dimensional space. For the dataset MI1, the first 40 samples per class from good target subject 23 were labeled for training, the rest of samples were unlabeled for testing. Additionally, all samples from other good subjects were the labeled source samples.

In **Figure 8**, for COV and TS, the discrimination of two labeled target classes marked in red were obscure, resulting in

TABLE 3 | The number of unqualified subjects whose accuracies were less than 70% on dataset MI1 for supervised and semi-supervised algorithms with varying R_l .

	10%	20%	30%	40%	Mean
CSP	45	42	37	32	39
GLRCSP	44	39	36	31	37.5
CCSP	42	37	33	31	35.75
COV	43	43	43	41	42.5
TS	43	42	40	38	40.75
sSCSTL	35	34	33	34	34
ssSCSTL	34	34	34	31	33.25

TABLE 4 | The number of unqualified subjects whose accuracies were less than 70% on dataset MI1 for unsupervised algorithms with varying R_u .

	60%	70%	80%	90%	Mean
MEKT	45	45	45	45	45
DTE	45	45	45	45	45
uSCSTL	38	37	38	37	37.5

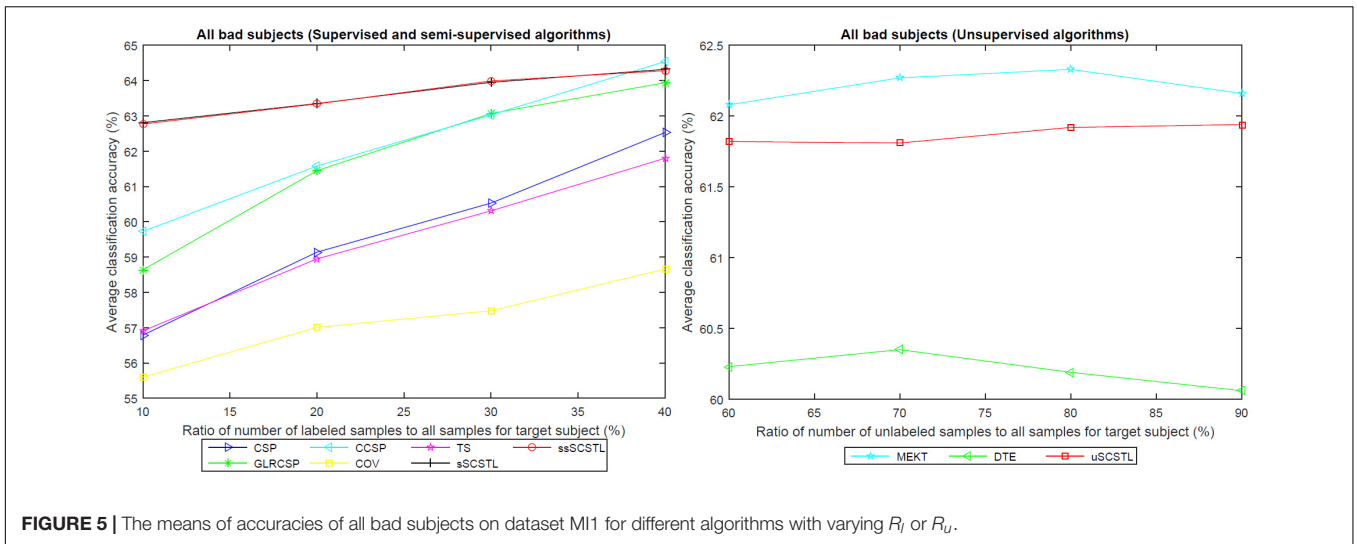


FIGURE 5 | The means of accuracies of all bad subjects on dataset MI1 for different algorithms with varying R_l or R_u .

the difficulty of classification for the unlabeled target samples marked in blue. However, after RA and TSM, the labeled target samples with same class for TS were closer than those for COV. ssSCSTL and uSCSTL discriminated the two-class labeled samples from target and source subjects well, after source selection using the labeled and unlabeled target samples, respectively. Moreover, the classification performances of our proposed algorithms might increase since many unlabeled target samples overlapped with labeled source samples which had the same class. Although, ssSCSTL could not separate the two labeled classes well, it performed well in most cases.

Computation Time Comparison

Figure 9 shows average computation time of different algorithms on datasets MI1 and MI2. For supervised and semi-supervised algorithms, 40% of total samples from target subjects are used

for training. For unsupervised algorithms, 60% of total samples from target subjects are unlabeled and used for cross-subject TL. Our experimental results were conducted in Matlab 2015a on a laptop with intel i5 CPU@1.60 GHz, 8 GB memory, running 64-bit Windows 10 Home Edition.

As shown in Figure 9, GLRCSP and CCSP spend much longer computation time than other algorithms when many good source subjects are available (seven for MI1 and five for MI2). Although CSP, COV, and TS save the running time, SCSTL and ssSCSTL spend acceptable running time. In addition, among all unsupervised algorithms, it takes the lowest time for DTE since DTE only selects one most appropriate source subject. The computation time of uSCSTL is also acceptable, although uSCSTL approximately spends twice running time of MEKT.

DISCUSSION

In this section, we discuss the experimental results from the following aspects.

- (1) In terms of cross-subject TL

There are three non-TL algorithms (CSP, COV, and TS) in our paper. As shown in Figures 4-6, the average classification accuracies of TL algorithms are superior to those of non-TL algorithms in total. It proves the validity of cross-subject TL methods when few labeled samples from target subject are available, thus effectively shortening the calibration time for target subject. As depicted in Figure 9, compared with non-TL algorithms, most cross-subject TL algorithms spend more but acceptable running time except for GLRCSP and CCSP. However, the cross-subject TL algorithms in our paper are all offline algorithms, in which all samples from target and source subjects are ready in advance. Therefore, they are not suitable for real-time BCI applications.

- (2) In terms of different feature spaces

TABLE 5 | Average classification accuracies of supervised and semi-supervised algorithms for all subjects on dataset MI2 ($R_l = 40\%$).

	a	b	c	d	e	f	g	Mean (Std)
CSP	87.67	69.33	68.75	87.17	92.50	89.50	91.17	83.73 (10.20)
GLRCSP	86.17	73.33	72.08	84.08	93.75	89.92	89.92	84.18 (8.42)
CCSP	87.67	74.50	71.17	86.75	93.00	90.75	90.92	84.96 (8.60)
COV	81.42	65.50	67.75	76.00	74.50	80.00	76.58	74.54 (5.93)
TS	88.33	65.50	70.83	82.42	89.83	89.33	92.33	82.65 (10.46)
ssSCSTL	79.33	69.92	68.08	99.92	100.00	98.17	100.00	87.92 (14.90)
ssSCSTL	83.25	67.50	71.50	100.00	100.00	100.00	100.00	88.89 (14.64)

TABLE 6 | Average classification accuracies of unsupervised algorithms for all subjects on dataset MI2 ($R_u = 60\%$).

	a	b	c	d	e	f	g	Mean (Std)
MEKT	78.33	73.50	73.00	70.00	90.33	80.83	91.92	79.70 (8.59)
DTE	71.00	72.25	71.25	81.08	86.33	68.42	79.33	75.67 (6.61)
uSCSTL	63.00	65.58	62.67	100.00	100.00	99.75	99.92	84.42 (19.35)

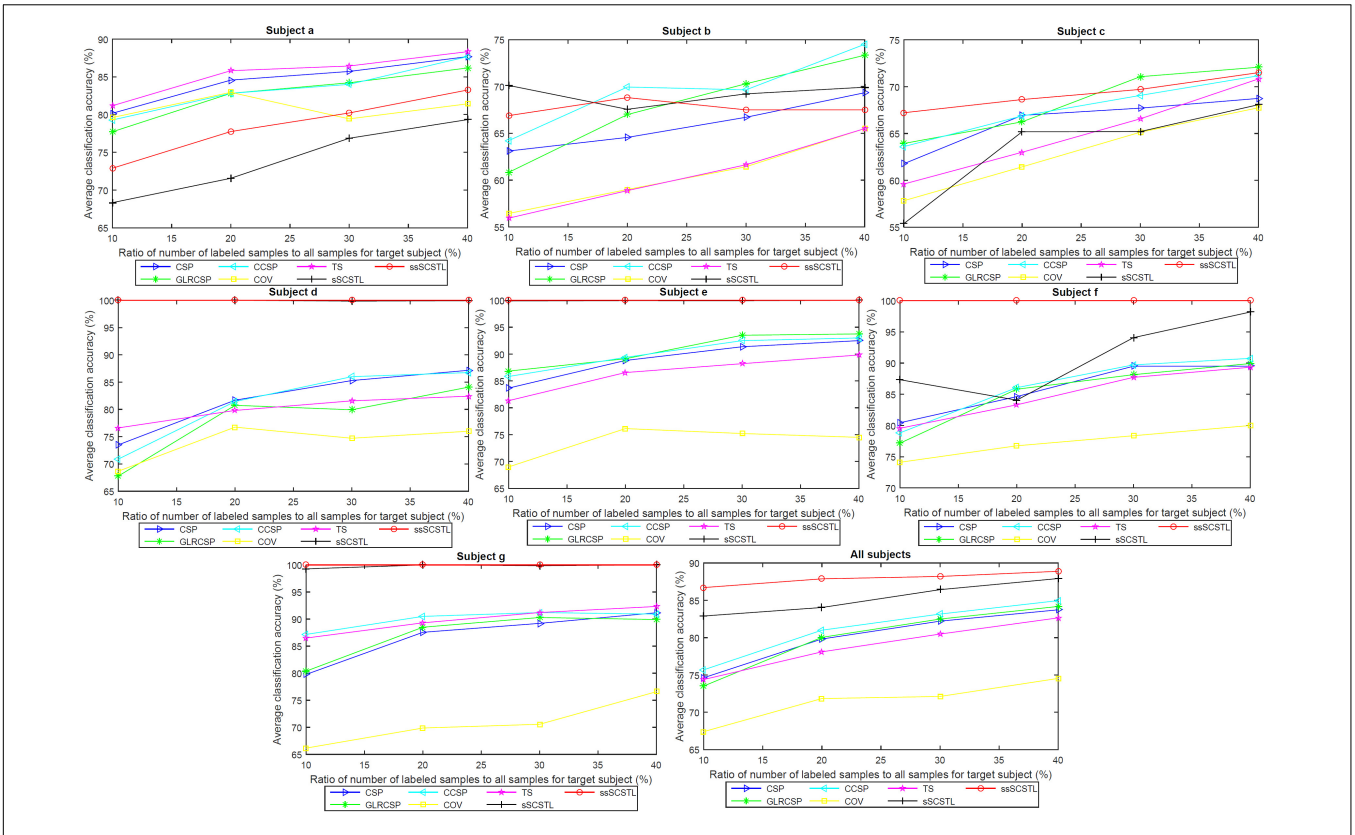


FIGURE 6 | Average classification accuracies of all subjects on dataset M12, and their means, with varying R_l , for supervised and semi-supervised algorithms.

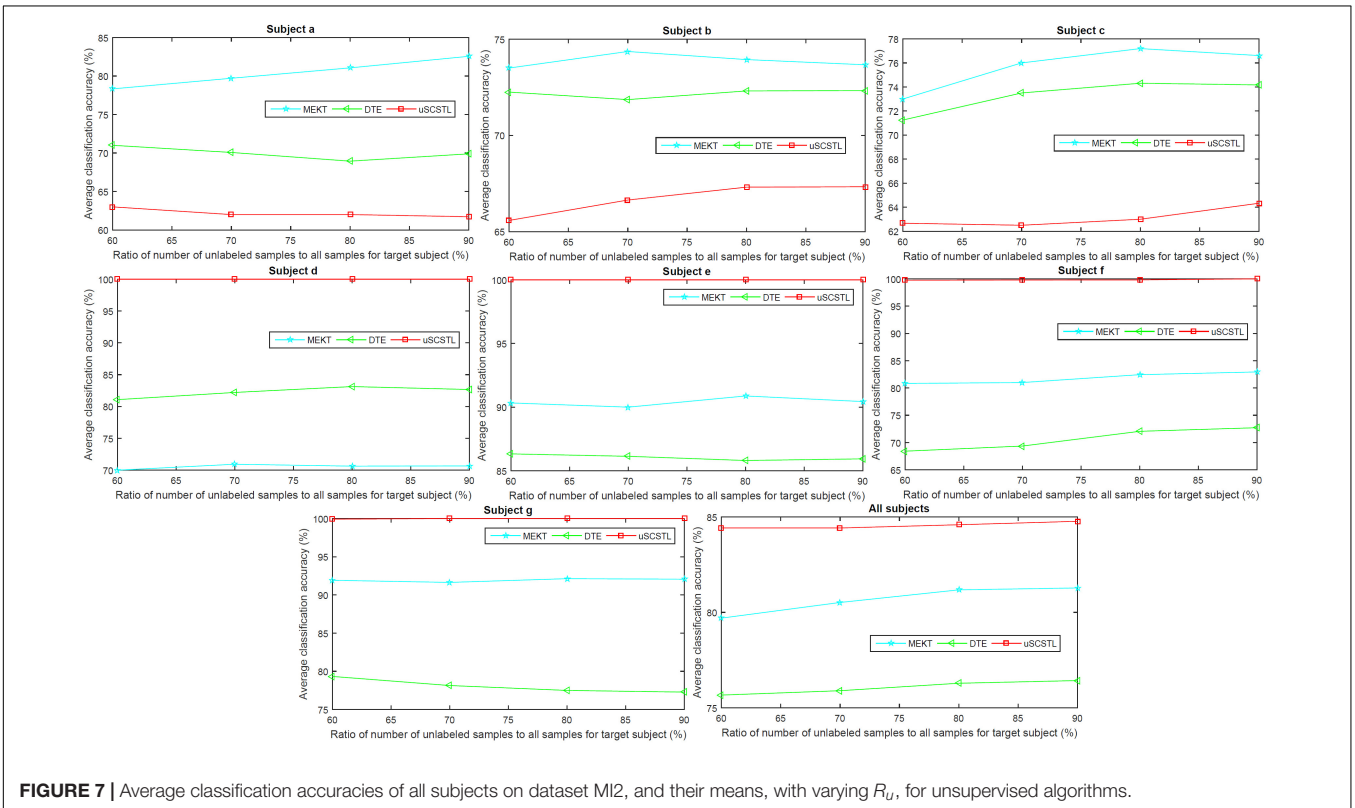


FIGURE 7 | Average classification accuracies of all subjects on dataset M12, and their means, with varying R_u , for unsupervised algorithms.

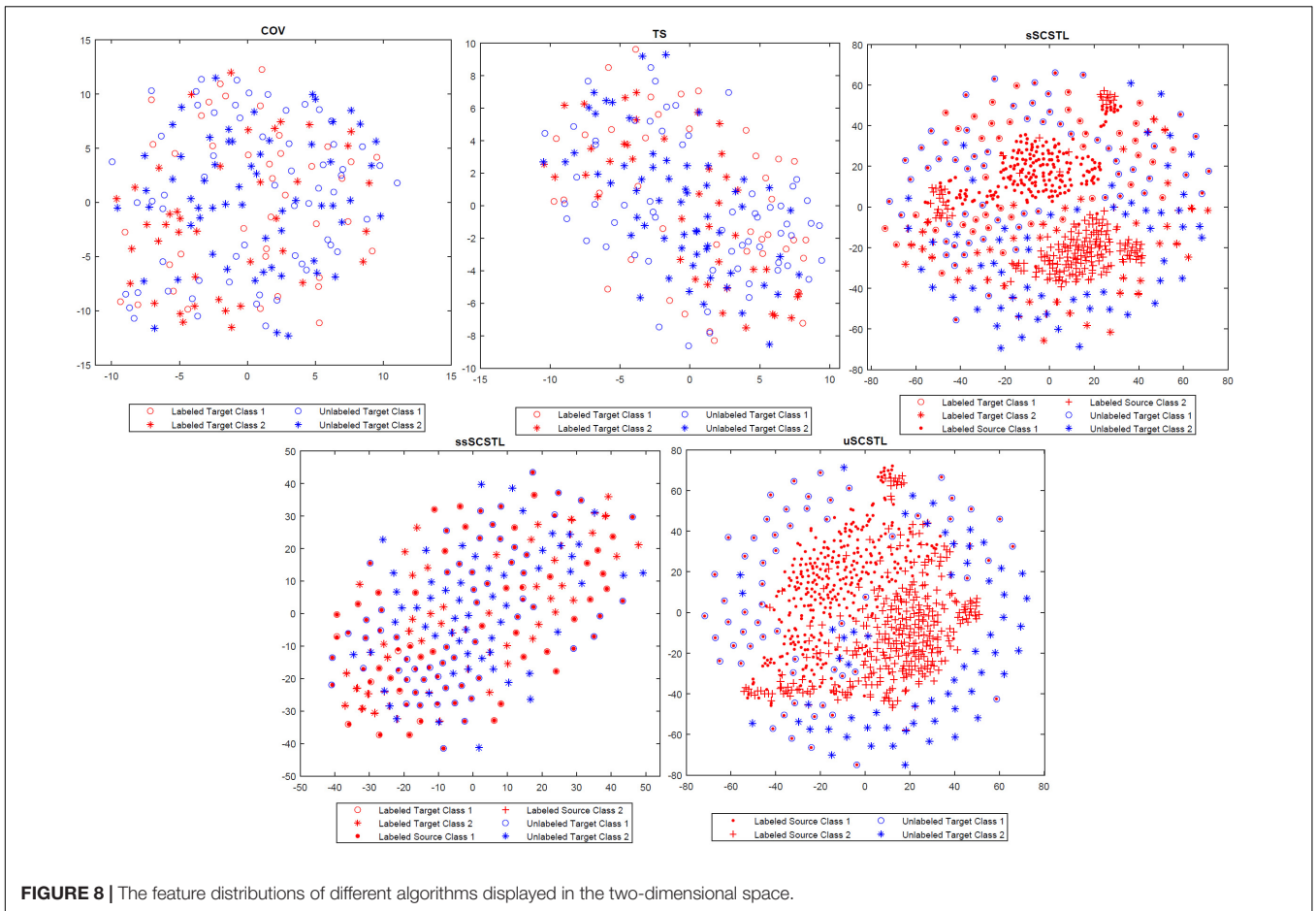


FIGURE 8 | The feature distributions of different algorithms displayed in the two-dimensional space.

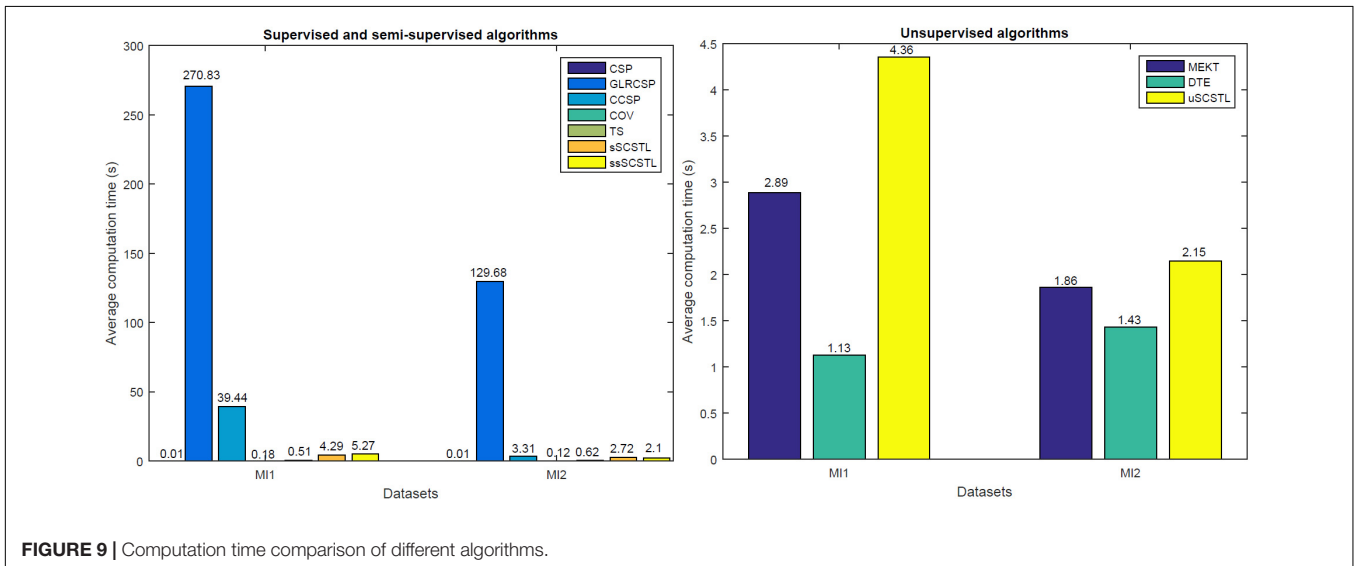


FIGURE 9 | Computation time comparison of different algorithms.

The baseline algorithms and our proposed algorithms are divided into three categories according to their feature spaces, including Euclidean space-based, Riemannian space-based, and Riemannian tangent space-based algorithms. For Euclidean space-based cross-subject TL algorithms, GLRCSP and

CCSP spend much time in computing optimal regularization parameters by means of cross-validation to weight the labeled samples from different subjects. For Riemannian tangent space-based algorithms, the hyper-parameters of MEKT and DTE are set by experience. Our proposed algorithms are parameter

free. As shown in **Figure 8**, compared with Riemannian space-based algorithm COV, our proposed algorithms can make the samples from same class close to each other. In a word, our proposed algorithms can balance the classification performance and computation time well.

(3) In terms of different subjects

All subjects on MI1 or MI2 are separated into good and bad subjects according to their classification performances. To avoid negative transfer, most suitable source subjects are selected from good subjects for our proposed algorithms. As shown in **Figures 4-7**, our proposed algorithms exhibit obvious improvements over other algorithms for good target subjects. However, our proposed algorithms perform unsteadily for bad target subjects. The reason is that the classification performances of our proposed algorithms depend on the discriminability of the two-class target samples. For good target subjects, high discriminability of labeled or unlabeled target samples can help our proposed algorithms shorten the distances of samples with same class and select most appropriate source subjects. Nevertheless, for bad target subjects, low discriminability of labeled or unlabeled target samples influences the aggregation of the samples with same class and the source selection.

(4) In terms of labeled and unlabeled samples

To investigate the role of labeled and unlabeled samples, we propose sSCSTL, ssSCSTL, and uSCSTL. The labeled, total, or unlabeled samples from target subject are used to compute the Riemannian mean and source selection. As given in **Tables 1, 5**, with the help of unlabeled samples from target subject, ssSCSTL performs slightly better than sSCSTL. Possible explanation is that adding unlabeled samples from target subject can not only make all samples from target subject closer, but also be good for better source selection. As shown in **Tables 2, 6**, our proposed uSCSTL displays better classification performance than MEKT and DTE for good subjects. It suggests that unlabeled samples from good target subject which have inherent high discriminability can boost the performance improvement even if labeled samples from target subject are scarce.

(5) In terms of different classifiers

The baseline algorithms and our proposed algorithms use different supervised classifiers. Even for unsupervised TL algorithms, MEKT, DTE, and uSCSTL only use the unlabeled samples from target subject in the feature extraction phase. Except for TS, all Riemannian tangent space-based algorithms use sLDA as classifier. From the experimental results mentioned above, sLDA is much suitable for high-dimensional features. All Euclidean space-based algorithms and TS use LDA as classifier. As shown in **Tables 1, 5**, CSP outperforms TS in most cases. Possible explanation is that LDA may not cope with high-dimensional features of TS well. Although the MDM classifier is much suitable for the classification of covariance matrices, the COV algorithm does not perform ideally due to without RA.

CONCLUSION

In this paper, we propose selective cross-subject transfer learning in supervised, semi-supervised, and unsupervised versions based on Riemannian tangent space. They perform RA using the available samples from target subject and the labeled samples from source subjects to preliminarily reduce the inter-subject differences. Then, all aligned covariance matrices from different subjects are converted into corresponding tangent space vectors for classification in Euclidean space. To realize positive transfer and relieve computational cost, the available tangent space vectors from target subjects are used to choose most suitable good source subjects based on SFFS. Experimental results show that our proposed algorithms are superior to several state-of-the-art algorithms, especially for good target subjects, when they have few or no labeled samples. Therefore, our proposed algorithms provide a new idea of solution for shortening calibration time in MI BCI.

However, our proposed ssSCSTL and uSCSTL algorithms are designed offline since the unlabeled samples from target subject are obtained a *prior*, instead of on-the-fly. Thus, we will further investigate how to adapt ssSCSTL and uSCSTL to the real-time BCI applications. Besides, ssSCSTL and uSCSTL only utilize the unlabeled samples from target subject in the feature extraction phase. Future work will be dedicated to extending ssSCSTL and uSCSTL in the classification phase to make full use of the unlabeled samples from target subject.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Dataset MI1: <http://dx.doi.org/10.5524/100295>, and Dataset MI2: http://www.bbci.de/competition/iv/desc_1.html.

AUTHOR CONTRIBUTIONS

YX wrote the manuscript, performed the research, and analyzed the experimental data. XH and QL designed, reviewed, and edited the manuscript. All authors read and approved the submitted manuscript.

FUNDING

This work was partially funded by the National Natural Science Foundation of China (Grant No. 62166020) and the Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ200318).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ronghua Hu and the referees for their constructive recommendations, which help to improve the manuscript substantially.

REFERENCES

- Azab, A., Ahmadi, H., Mihaylova, L., and Arvaneh, M. (2020). Dynamic time warping-based transfer learning for improving common spatial patterns in brain-computer interface. *J. Neural Eng.* 17:016061. doi: 10.1088/1741-2552/ab64a0
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* 59, 920–928. doi: 10.1109/TBME.2011.2172210
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039
- Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., et al. (2016). The Berlin brain-computer interface: progress beyond communication and control. *Front. Neurosci.* 10:530. doi: 10.3389/fnins.2016.00530
- Blankertz, B., Dornhege, G., and Krauledat, M. (2007). The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components a tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048
- Cheng, M., Lu, Z., and Wang, H. (2017). Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cogn. Neurodyn.* 11, 173–181. doi: 10.1007/s11571-016-9417-x
- Cho, H., Ahn, M., Ahn, S., Kwon, M., and Jun, S. C. (2017). EEG datasets for motor imagery brain computer interface. *GigaScience* 6, 1–8. doi: 10.1093/gigascience/gix034
- Fletcher, P. T., and Joshi, S. (2004). “Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors,” in *Proceedings of the 2004 Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, Prague, 87–98. doi: 10.1007/978-3-540-27816-0_8
- He, B., Baxter, B., Edelman, B. J., Cline, C. C., and Ye, W. W. (2015). Noninvasive brain-computer interfaces based on sensorimotor rhythms. *Proc. IEEE* 103, 907–925. doi: 10.1109/JPROC.2015.2407272
- He, H., and Wu, D. (2020). Transfer learning for brain-computer interfaces: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914
- Huang, C., Huang, X., Fang, Y., Xu, J., Qu, Y., Zhai, P., et al. (2020). Sample imbalance disease classification model based on association rule feature selection. *Pattern Recogn. Lett.* 133, 280–286. doi: 10.1016/j.patrec.2020.03.016
- Huang, X., Yan, F., Xu, W., and Li, M. (2019). Multi-attention and incorporating background Information model for chest X-ray image report generation. *IEEE Access* 7, 154808–154817. doi: 10.1109/access.2019.2947134
- Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Process. Lett.* 16, 683–686. doi: 10.1109/LSP.2009.2022557
- Liang, Y., and Ma, Y. (2020). Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning. *Biomed. Signal Proces.* 62:102101. doi: 10.1016/j.bspc.2020.102101
- Lotte, F., and Guan, C. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, 355–362. doi: 10.1109/TBME.2010.2082539
- Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. N. (2009). “Regularized common spatial patterns with generic learning for EEG signal classification,” in *Proceedings of the 2009 IEEE Engineering in Medicine and Biology Society (Minneapolis, MN: IEEE)*, 6599–6602. doi: 10.1109/iembs.2009.5332554
- Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 26, 735–747. doi: 10.1137/s0895479803436937
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Peck, R., and Van Ness, J. (1982). The use of shrinkage estimators in linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 4, 530–537. doi: 10.1109/tpami.1982.4767298
- Pudil, P., Ferri, F. J., and Kittler, J. (1994). “Floating search methods for feature selection with nonmonotonic criterion functions,” in *Proceedings of the 12th International Conference on Pattern Recognition (Jerusalem, Israel: IEEE)*, 279–283. doi: 10.1109/ICPR.1994.576920
- Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946
- Rodriguez, P. L. C., Jutten, C., and Congedo, M. (2019). Riemannian procrustes analysis: transfer learning for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 66, 2390–2401. doi: 10.1109/TBME.2018.2889705
- Sample, M., Aunos, M., Blain-Moraes, S., Bublitz, C., Chandler, J. A., Falk, T. H., et al. (2019). Brain-computer interfaces and personhood: interdisciplinary deliberations on neural technology. *J. Neural Eng.* 16:063001. doi: 10.1088/1741-2552/ab39cd
- Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055
- Tuzel, O., Porikli, F., and Meer, P. (2008). Pedestrian detection via classification on Riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 1713–1727. doi: 10.1109/TPAMI.2008.75
- Van Der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3
- Xu, Y., Wei, Q., Zhang, H., Hu, R., Liu, J., Hua, J., et al. (2019). Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. *J. Circuits Syst. Comp.* 28, 1950123. doi: 10.1142/S0218126619501238
- Yin, E., Zeyl, T., Saab, R., Chau, T., Hu, D., and Zhou, Z. (2015). A hybrid brain-computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 693–701. doi: 10.1109/TNSRE.2015.2403270
- Zanini, P., Congedo, M., Jutten, C., Said, S., and Berthoumieu, Y. (2018). Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 65, 1107–1116. doi: 10.1109/TBME.2017.2742541
- Zhang, W., and Wu, D. (2020). Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1117–1127. doi: 10.1109/TNSRE.2020.2985996

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Xu, Huang and Lan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.