Manifold Embedded Domain Adaptation for Motor Imagery EEG Decoding

Qin Jiang*, Yi Zhang*, Wei Wang, Qian Huang

Abstract—**Training a robust classifier for a motor imagery-based brain-computer interface (MI-BCI) system typically requires a substantial amount of time to collect calibration data, which can be a burdensome task for participants. To enhance the classifier performance while reducing the effort during the training phase, this paper proposes a domain adaptation algorithm based on manifold embedding (eSPDA) by combining the domain adaptation approach with a dimensionality reduction framework derived from the common spatial patterns (CSP). Specifically, the CSP spatial filtering theory is construed as the dimensionality reduction in Riemannian manifold with maximum intra-class variance or maximum inter-class distance. Based on the principle of maximum inter-class distance, the labeled source data is embedded into a more discriminative submanifold, where the principal characteristics of the unlabeled target subject are preserved by the rule of maximum intra-class variance. Meanwhile, the joint distribution alignment is integrated into the framework to minimize the distribution divergences across subjects. The results on two datasets demonstrate that eSPDA outperforms several state-of-the-art domain adaptation methods, with the average accuracies 70.35% and 80.67% on BCI Competition IV dataset IIa and BCI Competition IV dataset IVa, respectively. This research indicates that eSPDA has potentials to reduce the labeling effort, resulting in calibration time and effort savings.**

Index Terms—**small training sets, common spatial patterns, domain adaptation, Riemannian manifold, dimensionality reduction**

I. INTRODUCTION

Motor imagery-based brain-computer interfaces (MI-BCI) is technology that enable individuals to control devices or interact with the environment solely through their imagination. [1]. Generally, motor imagery is associated with changes in neural oscillation known as event-related desynchronization and event-related synchronization [2], and the intended motion can be predicted by decoding these rhythmic activities [3]. With a well-designed algorithm, these motor intentions can be converted into recognizable computer commands to control a wheelchair [4] or an artificial limb [5]. However, one of the challenges in MI-BCI systems is the time-consuming collection of sufficient training samples, coupled with the mental burden on the subjects, which may hinder the widespread adoption and practicality of MI-BCI applications. Since each motor imagery-based sample typically consists of a task prompt, task execution, and rest period, a complete sample takes around 9 seconds. Furthermore, the collection process requires the subject to be remain fully engaged, silent, and avoid conscious eye blinks and physical movements, which may be mentally burdensome. Therefore, the development of powerful decoding algorithms to reduce the calibration burden is a prominent area of research in the MI-BCI field.

The common spatial patterns (CSP) algorithm is widely used as an efficient feature extraction method in decoding MI-associated EEG signals [6],[7]. As a supervised method, CSP extracts the discriminant information of EEG recordings by assuming that the sample covariance matrices differ significantly between different classes. However, the CSP filter is prone to overfitting with a small training set [8],[9]. To improve the generalization ability, various regularized CSPs have been developed by incorporate a-priori information to the estimation of the inter-class covariance matrix [10]. The regularization matrix can be either an identity matrix [11], a diagonal matrix [12], or a generic covariance matrix. The generic covariance matrix is constructed as a weighted sum of covariance matrices using data from other subjects, where the weights can be determined based on metric like Kullback-Leibler divergence [13], information entropy [14], or cosine similarity between subjects [15].

Although regularized CSPs have undergone some advancement in alleviating the noise sensitivity and overfitting of spatial filters, the effects are not necessarily satisfying when only limited training data are available. This is because most existing regularized CSPs are finely parameterized, and the parameters are data-dependent and subject-specific, which makes it difficult to search for the optimal parameters in small labeled training sets, and even more challenging for unlabeled training data [16]. For this problem, domain adaptation methods offer a potential solution by leveraging knowledge from multiple subjects and transferring it to new subjects with limited labeled training samples [17]. These methods aim to find a shared feature

Q. Jiang is a researcher of Chongqing College of Electronic Engineering, Chongqing, 401331 China (corresponding author phone: 023-65926593; fax:023-65927000; e-mail: namy_jiang@hotmail.com).

Y. Zhang is a professor of School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065 China (corresponding author phone: 023-62480054; fax: 023-62480054; e-mail: zhangyi@cqupt.edu.cn).

W. Wang is a PhD candidate of Chongqing University of Posts and Telecommunications, Chongqing, 400065 China (e-mail: d190201021@stu.cqupt.edu.cn).

Q. Huang is a researcher of Chongqing College of Electronic Engineering, Chongqing, 401331 China (e-mail: huangqian@cqcet.edu.cn).

representation of different subjects through subspace learning, allowing a classifier trained on labeled source data to be applied to the target data [18].

Since the covariance matrices of EEG recordings, being symmetric positive definite (SPD), can be used as a feature descriptor and analyzed by Riemannian metrics. Covariance matrices with geometric information provide a novel approach for BCI data analysis [19],[20],[21]. Barachant et al. [22],[23] were among the first to introduce Riemannian geometry into the classification of EEG covariance matrices and proposed two widely used classifiers, the minimum distance to the Riemannian mean (MDRM) algorithm and the support vector machine classifier in tangent space (TSVM). Capitalizing on the geometric properties of SPD matrices, several domain adaptation techniques have been proposed to enable the time-series data across sessions or subjects comparable. Zanini et al. [24] proposed a Riemannian alignment method that normalizes the covariance matrices of sessions or subjects with respect to the reference matrix in the resting state. Yair et al. [25] provided a domain adaptation method using parallel transport on the SPD manifold, which projects the symmetric matrices onto a shared tangent space to alleviate the domain shifts. Rodrigues et al. [26] proposed a Riemannian Procrustes analysis that estimates the statistical characteristics of datasets with the geometric means of the SPD matrices and matches statistical distributions through simple geometric transformations like translation, scaling, and rotation. Cai et al. [27] developed a manifold embedding transfer learning method that leveraged the geometric properties of the Riemannian manifold and joint distribution adaptation technique to learn a shared feature presenting across domain. Zhang et al. [28] introduced a manifold embedded knowledge transfer (MEKT) framework, which centers covariance matrices of subjects with respect to their Riemannian mean and then performs domain adaptation by minimizing distribution divergence between domains while preserving geometric structure. Xu et al. [29] presented a feature selection method for tangent space, where tangent vectors of aligned covariance matrices are extracted as features and selected through the sequential forward floating search method. However, many existing domain adaptation algorithms rely heavily on exponential/logarithmic maps for transitioning back and forth to the tangent space, leading to high computational costs. Besides, the tangent space is only a first-order approximation of the Riemannian manifold and does not scale well with the number of training samples. Moreover, the high dimensionality of tangent vectors can potentially lead to overfitting, as the dimension of tangent vector expands rapidly with an increase in the number of recording electrodes, causing the dimension of tangent vector becomes too high compared to the number of available training samples [30],[31].

The paper presents a domain adaptation method based on manifold embedding (eSPDA) to address the issue of a shortage of labeled target training data. The proposed method aims to learn a more discriminative submanifold across the target and other subjects, where the projected features from the target subject have a similar distribution to those of other subjects, thus enables the classifier trained on auxiliary data from other subjects to be applied to the target data. To this end,

the covariance matrix of each EEG trial is treated as a feature descriptor in Riemannian manifold, and then a subspace learning framework combining the CSP principle and the joint distribution alignment technique is used to project the raw covariance matrix into a more discriminative submanifold. In concretely, the CSP filtering is interpreted as dimensionality reduction that maximizes inter-class distance and intra-class variance in the Riemannian manifold. The labeled source data are embedded into a more discriminative submanifold space based on the maximization of interclass distance while the principal characteristics of the unlabeled target data are preserved based on the maximization of intra-class variance. Additionally, the joint distribution alignment method is incorporated to minimize the distribution divergence between the target and source data.

The rest of the article is as follows. Section 2 introduces the related work based on distribution alignment and CSP theory; Section 3 describes our proposed framework; Section 4 provides a detailed description of the experimental design and the results on two datasets. Section 5 concludes the study.

II. BACKGROUND

In this paper, S_+^n is the space spanned by the $n \times n$ SPD matrices, I_n denotes an $n \times n$ identity matrix. $E_i \in \mathbb{R}^{n \times t}$ denotes a trial of EEG signal recorded with t time samples and n channels. C_i represents a covariance matrix in Euclidean space, and X_i denotes the covariance matrix descriptor in the Riemannian manifold. x_i represents a vectorized feature in the Euclidean space. $\|X\|_F = \sqrt{Tr(X^T X)}$ represents the Frobenius norm, where $tr(.)$ is the sum of the diagonal elements, and $(.)^T$ denotes the transpose operator. $\{(X_i^s, y_i)\}_{i=1}^{Ns}$ denotes Ns labeled samples from other subjects (referred to as 'sources', denoted as D_s), where $y_i \in \{1,...,l\}$ denotes the tag. $\{X_i^t\}_1^{Nt}$ is a collection of unlabeled data from target subject (denoted as D_t). It assumes that feature space and label space between domains are same, while the conditional and marginal probability distributions are different however due to domain shifts, i.e. $\{X_i^s\} = \{X_i^t\}$, $y_s = y_t$, $Q_s(y_s | X_i^s) \neq Q_t(y_t | X_i^t)$, $P_s(X_i^s) \neq P_t(X_i^t)$.

A. Domain Adaptation based on Distribution Alignment

The aim of domain adaptation is to find an underlying feature subspace across domains where the data distributions across domains are similar, ensuring the classifier trained on the source domain to be effectively applied to the target. Borgwardt et al. [32] proposed a parameterless metric, maximum mean discrepancy (MMD) to estimate the discrepancy via the empirical mean between domains:

$$dist(D_s, D_t) = \left\| \frac{1}{Ns} \sum_{i=1}^{Ns} \phi(x_i^s) - \frac{1}{Nt} \sum_{i=1}^{Nt} \phi(x_i^t) \right\|_{\mathrm{H}}^2 \quad (1)$$

where H refers to the reproducing kernel Hilbert space (RKHS), and $\phi(.)$ is a nonlinear map that embeds the features onto the RKHS.

Equation (1) calculates the distribution difference between domains without labels. Capitalizing on the concept of MMD, Pan et al. [33] formulated a feature transfer component analysis (TCA) algorithm, aiming to find a transformation matrix to minimize the discrepancy between the projected features across domains. Long et al. [34] introduced the joint distribution adaptation (JDA) method, an extension of TCA algorithm that aims to minimize the difference in both conditional and marginal probability distribution across domains in the subspace. Based on the JDA framework, the balanced distribution adaptation (BDA) sets a distribution weight factor to adaptively weigh the importance of the difference between conditional and marginal distributions [35]. JDA incorporates both the conditional and marginal distributions in its adaptation process where the labels from both source and target domains are considered to align the conditional distributions, while TCA is an unsupervised domain adaptation method that focuses on aligning the marginal distributions between the source and target domains. Moreover, JDA iteratively trains a weak classifier on the projected source data and updates the feature transformation and classifier iteratively to minimize the divergence between the conditional distributions.

In JDA, the difference in marginal distribution in the subspace is expressed as:

$$Dist(P(x^s), P(x^t)) = \left\| \frac{1}{Ns} \sum_{i=1}^{Ns} A^T x_i^s - \frac{1}{Nt} \sum_{i=1}^{Nt} A^T x_i^t \right\|_F^2 \quad (2)$$

The discrepancy in conditional probability distribution is expressed as:

$$Dist(Q(x^s), Q(x^t)) = \left\| \frac{1}{Ns^{(c)}} \sum_{X_i^s \in Ds^{(c)}} A^T x_i^s - \frac{1}{Nt^{(c)}} \sum_{X_i^t \in Dt^{(c)}} A^T x_i^t \right\|_F^2 \quad (3)$$

where $A \in \mathbb{R}^{n \times m}$ is an orthogonal transformation matrix.

Based on joint distribution adaptation theory, Zhang et al. [36] proposed a discriminative joint probability maximum mean distribution adaptation (DMJP-MMD) method, which simultaneously minimized the joint probability distribution discrepancy in the same class between different domains for transferability and maximizes the joint probability distribution discrepancy between different classes of different domains for discriminability. Wang et al.[37] proposed domain adaptation with manifold embedded distribution alignment (MEDA), where the source and target domains are treated as two points on a Grassmann manifold, their features are then projected onto a common manifold using geodesic flow kernel (GFK)[38], and a shared classifier is then learned by minimizing the joint distribution discrepancies across domains.

In this paper, a domain adaptation method based on SPD manifold embedding (eSPDA) is proposed to address the challenge of domain adaptation when dealing with symmetric positive definite (SPD) matrices. Unlike the aforementioned algorithms that assume features are in vector form, eSPDA exploits the geometric information of SPD matrices.

B. Common Spatial Patterns (CSP)

$E_{(c,i)} \in \mathbb{R}^{n \times t}$ denotes an EEG record of class $c \in \{1, 2\}$, and the covariance matrix $C_{(c,i)}$ is computed as:

$$C_{(c,i)} = \frac{E_{(c,i)}.E^T_{(c,i)}}{tr(E_{(c,i)}.E^T_{(c,i)})} \quad (4)$$

The CSP filters aim at maximizing the variance of EEG recordings from one class while minimizing their variance from the other class. This optimization objective can be formalized as:

$$J(\omega) = \arg\max_{\omega} \frac{\omega^T \Sigma_1 \omega}{\omega^T \Sigma_1 \omega + \omega^T \Sigma_2 \omega} \quad (5)$$

where Σ_1 and Σ_2 denote the means of the inter-class covariance matrices and ω represents a spatial projection shared by the two classes.

The optimization of (5) amounts to $\arg\max \omega^T \Sigma_1 \omega$ $s.t$ $\omega^T \Sigma_1 \omega + \omega^T \Sigma_2 \omega = I_m$, i.e., it solves a projection matrix ω that maximizes the intraclass variance under the restriction of $\omega^T \Sigma_1 \omega + \omega^T \Sigma_1 \omega = I_m$. Formula (5) indicates that CSP filters aim to find a subspace where the variance within each class is maximized, and the variance between classes is maximally separated. After filtering by ω, the most distinguished information in the raw EEG signal is preserved. In the context of the covariance matrix, it is observed that a high-dimensional covariance matrix is embedded into a more discriminant low-dimensional manifold, which implies that the CSP filtering can be interpreted as a discriminant dimensionality reduction on Riemannian manifolds.

III. DOMAIN ADAPTATION BASED ON SPD MANIFOLD EMBEDDING

The eSPDA is designed to learn a latent submanifold underlying the domains where the classifier trained by the labeled source data is adapted to the unlabeled target data. To this end, the subspace needs to satisfy two requirements: (1) the distribution discrepancy between domains is minimized, and (2) the discriminant information is maximumly preserved. To satisfy these requirements, the eSPDA method follows a two-step process, as illustrated in Fig. 1 In the first beginning, an unsupervised Riemannian alignment is used to align the Riemannian means of the source and target datasets to ensure that the principal component directions of the two domains are similar, thus reducing the marginal discrepancy between the domains. Subsequently, a subspace learning is performed to embed the high-dimensional SPD manifold into a more discriminative submanifold that preserves the maximum discriminant information while simultaneously reducing the conditional difference between the source and target domains.

A. Riemannian Alignment

The correlation alignment (CORAL) [39] mitigates the domain shift by aligning the second-order statistics of features. Inspired by CORAL, a Riemannian alignment method (RA) is formulated to align the Riemannian means of source and target features.

In RA, the first step is to decorrelate the source domain by centering it to an identity matrix, i.e., removing the correlations of the source domain by:

$$X_i^{s'} = (\bar{X}^s)^{-1/2} X_i^s (\bar{X}^s)^{-1/2} \quad (6)$$

Then, the source is recorrelated with correlations of the

target domain.

$$X_i^{st} = (\bar{X}^t)^{1/2} X_i^{s'} (\bar{X}^t)^{1/2} \qquad (7)$$

where $\bar{X}^s$ and $\bar{X}^t$ denote the Riemannian mean of the source and target features, respectively, and X_i^{st} represents the source matrix after re-correlation. Equation (7) reconstructs the source matrices by the target Riemannian mean, i.e., regenerating the source SPD matrices using all eigenvectors and eigenvalues of the target, which results in similar principal axes and the same Riemannian mean between regenerated source and target. RA is conducive to alleviating the domain shift caused by electrode installation position, environmental factors, and individual differences.

B. Dimensionality Reduction of Riemannian Manifold

The feature extraction of CSP is expressed as a general dimensionality reduction on the SPD manifold, $g : S_+^n \to S_+^m$:

$$g(X) = \omega^T X \omega \qquad (8)$$

where X denotes any point on the SPD manifold ($0 \prec X \in S_+^n$). To guarantee that the result of $g(X)$ lies in the submanifold of S_+^m , the transformation matrix $\omega \in R^{n \times m}$ ($m < n$) must be full rank, i.e. subject to $\omega^T \omega = I_m$.

An effective subspace should preserve the principal characteristics of target data and avoid projecting the features into irrelevant dimensions. For unlabeled matrix data, the most discriminative information can be expressed with its variance:

$$\max_{\omega} \frac{1}{Nt} \sum_{i=1}^{Nt} \left\| g(X_i^t) - g(\bar{X}^t) \right\|_F^2$$
$$\Rightarrow \max_{\omega} tr(\omega^T F_1(\omega)\omega) \qquad (9)$$
$$s.t. \qquad \omega^T \omega = I_m$$

where $F_1(\omega) = \frac{1}{Nt} \sum_{i=1}^{Nt} (X_i^t - \bar{X}^t)\omega\omega^T (X_i^t - \bar{X}^t)^T$. $\bar{X}^t$ is the Riemannian mean of target matrices.

In (9), the variance of projected matrix-based data is evaluated by considering the maximum Riemannian distance to their Riemannian mean, and the optimal project matrix strives to preserve the maximum dispersion on the submanifold, rather than the maximum variance of the EEG records.

Additionally, the transformation matrix should maximally preserve the discriminative information of the labeled source data in the submanifold. In CSP, the optimal projection is determined by maximizing the ratio of inter-class variance, i.e., the difference between the filtered means of different classes (i.e., $\omega^T \Sigma_1 \omega$ and $\omega^T \Sigma_2 \omega$). In eSPDA, discriminative information is evaluated by maximizing the distance between the inter-class Riemannian means:

$$\max_{\omega_o} \left\| g(\bar{X}_1^{st}) - g(\bar{X}_2^{st}) \right\|_F^2$$
$$\Rightarrow \max_{\omega} tr(\omega^T F_2(\omega)\omega) \qquad (10)$$
$$s.t. \qquad \omega^T \omega = I_m$$

where $F_2(\omega) = (\bar{X}_1^{st} - \bar{X}_2^{st})\omega\omega^T (\bar{X}_1^{st} - \bar{X}_2^{st})^T$, $\bar{X}_c^{st}$ represents the Riemannian mean of the reconstructed c-class source matrices.

C. Joint Distribution Alignment

Similar to (2), the marginal distribution difference (MMD_{marg}) is estimated by the MMD on the submanifold:

$$MMD_{marg} = \left\| \frac{1}{Ns} \sum_{i=1}^{Ns} g(\tilde{X}_i^{st}) - \frac{1}{Nt} \sum_{i=1}^{Nt} g(\tilde{X}_i^t) \right\|_F^2 \qquad (11)$$

By substituting (8) into (11), (11) is further simplified as:

$$MMD_{marg} = \left\| \frac{1}{Ns} \sum_{i=1}^{Ns} g(\tilde{X}_i^{st}) - \frac{1}{Nt} \sum_{i=1}^{Nt} g(\tilde{X}_i^t) \right\|_F^2 \qquad (12)$$
$$= tr(\omega^T F_3(\omega)\omega)$$

with $F_3(\omega) = \sum_{i=1}^{Ns} \sum_{j=1}^{Nt} (\frac{1}{Ns} \tilde{X}_i^{st} - \frac{1}{Nt} \tilde{X}_j^t)\omega\omega^T (\frac{1}{Ns} \tilde{X}_i^{st} - \frac{1}{Nt} \tilde{X}_j^t)^T$

According to (3), the conditional distribution difference (MMD_{cond}) in the embedding manifold can be estimated as:

$$MMD_{cond} = \sum_{c=1}^{l} \left\| \frac{1}{Ns^{(c)}} \sum_{X_i^{st} \in Ds^{(c)}} g(\tilde{X}_i^{st}) - \frac{1}{Nt^{(c)}} \sum_{X_i^t \in Dt^{(c)}} g(\tilde{X}_i^t) \right\|_F^2 \qquad (13)$$
$$= \sum_{c=1}^{l} tr(\omega^T F_4^{(c)}(\omega)\omega)$$

$$F_4^{(c)}(\omega) = \sum_{i=1}^{Ns^{(c)}} \sum_{j=1}^{Nt^{(c)}} (\frac{1}{Ns^{(c)}} \tilde{X}_i^{st} - \frac{1}{Nt^{(c)}} \tilde{X}_j^t)\omega\omega^T (\frac{1}{Ns^{(c)}} \tilde{X}_i^{st} - \frac{1}{Nt^{(c)}} \tilde{X}_j^t)^T$$

Therefore, the joint distribution alignment is obtained by:

$$\min \mu MMD_{marg} + (1 - \mu)MMD_{cond} + \lambda \|W\|_F^2 \qquad (14)$$

Combining all of the optimization objectives (9), (10) and (14), the overall objective function of the proposed eSPDA method is formulated as:

$$\omega^* = \max_{W \in \mathbb{R}^{n \times m}} \frac{\alpha tr(\omega^T F_1(\omega)\omega) + \beta tr(\omega^T F_2(\omega)\omega)}{\mu tr(\omega^T F_3(\omega)\omega) + (1-\mu) \sum_{c=1}^{l} tr(\omega^T F_4^{(c)}(\omega)\omega) + \lambda \|\omega\|_F^2}$$
$$s.t. \quad \omega^T \omega = I_m$$
$$(15)$$

where α and β are the trade-off parameters to balance the importance of the source and target; μ is the balance factor, which is used to weigh the importance of conditional and marginal distribution; and λ is the regularization parameter to guarantee a well-defined optimization problem. Due to the distribution difference across domains and the uncertainty of the target conditional distribution, the conditional and marginal distribution between domains have different effects on the domain alignment. Therefore, a balance factor μ is designed to quantitatively estimate the weight between the conditional and marginal distribution. Here, μ is estimated by the inter-class and intraclass distances between domains.

The inter-class distance between the domains is defined as $d_m = \left\| \bar{X}^s - \bar{X}^t \right\|_F^2$, and the intra-class distance is expressed as $d_m^{(c)} = \left\| \bar{X}_s^{(c)} - \bar{X}_t^{(c)} \right\|_F^2$, where $\bar{X}_s^{(c)}$ and $\bar{X}_t^{(c)}$ denote the c-class Riemann mean of the source and target, respectively. Eventually, the balance factor is estimated as $\mu = 1 - 2d_m / (2d_m + \sum d_m^{(c)})$, and μ is re-estimated at each iteration.

Equation (15) involves the pseudo-labels of the target, which are initially predicted by a classifier trained on the labeled source data and then updated by the learned

projection matrix with the labelling quality iteratively improved until convergence. It is worth noting that existing standard machine learnings are primarily designed for Euclidean space and may not perform well when directly applied to manifold-valued data. To address this issue, several studies have explored extending kernel methods to Riemannian manifolds by defining appropriate distance metrics and kernel functions [40],[41]. This paper defines a kernel SVM classifier based on the log-Euclidean metric (LEM) Gaussian RBF kernel [40], with the Euclidean distance replaced by the log-Euclidean geodesic distance on the manifold. Thus, a given target sample in the submanifold ($X \in S_+^m$) is determined by:

$$
\begin{aligned}
f(X) &= \omega^T \phi(X) + b \\
&= \sum_{i=1}^{Ns} \alpha_i y_i \kappa(X_i, X) + b
\end{aligned}
\tag{16}
$$

where ω is the weight of the feature vectors, b is the bias, α is the dual coefficient, and κ is a positive definite kernel defined on the Riemannian manifold.

Training with a LEM-based kernel SVM is equivalent to solving the standard kernel SVM problem with a kernel matrix generated by κ. This means that any existing SVM software package can be utilized for training and classification task.

D. Optimization

Since $F_n(\omega)$ is an explicit function of ω, equation (15) usually does not have a closed-loop solution. An iterative generalized eigendecomposition optimization solution is presented to approximate or optimize the ω value.

In the k-th iteration, $F_1(\omega)$, $F_2(\omega)$, $F_3(\omega)$ and $F_4(\omega)$ are calculated by using the projection matrix of the k-1-th iteration, and the projection matrix of the k-th iteration is expressed as:

$$
\omega_k = \arg\max_{W \in \mathbb{R}^{n \times m}} \frac{tr(\omega^T L_n \omega)}{tr(\omega^T L_d \omega)} \qquad s.t. \quad \omega^T \omega = I_m
$$

$$
\text{with} \quad L_d = \mu F_3(\omega_{k-1}) + (1-\mu)\sum_{c=1}^{l} F_4^{(c)}(\omega_{k-1}) + \lambda I \tag{17}
$$

$$
L_n = \alpha F_1(\omega_{k-1}) + \beta F_2(\omega_{k-1})
$$

The optimization problem of (17) is transformed into a generalized eigen decomposition of (L_n, L_d), and ω_k is composed of eigenvectors corresponding to m largest eigenvalues. The iteration is repeated until a relatively stable classifier is obtained, and the execution process is described in Algorithm 1.

Algorithm 1: Iterative optimization of eSPDA

Input: labeled training set from the sources $\{(X_i^s, y_i)\}_{i=1}^{Ns}$, unlabeled target data $\{X_i^t\}_1^{Nt}$, iterations M, convergence threshold ε, parameters α, β, λ.

Output: transformation matrix ω^*

1. Initialize $W_0 \leftarrow I_{n \times m}$
2. **for** k=1,2,…M
3.   Calculate the low-dimensional SPD matrices of the source and target domains by (8).
4.   Train the TSVM, obtain the pseudo-labels of the target, and then estimate μ.
5.   **Repeat:**
6.   Calculate $F_1(\omega)$, $F_2(\omega)$, $F_3(\omega)$ and $F_4(\omega)$ terms of (9) (10) (12) (13).
7.   Obtain the projection matrix ω_k by generalized eigendecomposition of (17).
8.   **if** $|\omega_k - \omega_{k-1}| < \varepsilon$
9.   **Break;**
10.   **end if**
11.   **Until convergence**
12.   **end for**
11.   Obtain transformation matrix ω^*

IV. EXPERIMENTS AND RESULTS

A. Dataset Description

A series of experiments were performed on two public datasets to validate the effectiveness of the proposed eSPDA.

BCI Competition III Dataset IVa contains two-class EEG signals, a total of 280 cue-based trials recorded by 118 channels at a 100Hz sampling rate. **BCI Competition IV Dataset IIa** is composed of the four-class cued motor imagery data from nine subjects (A1−A9) recorded by 22 EEG channels with a 250Hz sampling rate. Please refer to [42] for more information about these two datasets.

Fig. 2 shows the cue timing schedule and event-related desynchronization (ERD) brain topography of DatasetIVa and DatasetIIa, respectively. It figures out that each task induces a unique ERD distribution, and the energy distribution exhibits significant individual differences, indicating the presence of domain drift. In the perspective of electrode-based feature representation, each electrode is considered as a dimension of the feature, and the corresponding energy value represents the feature value. In this context, the CSP spatial filter is interpreted as a mechanism to reduce the influence of irrelevant electrodes and enhance the discriminative power of relevant electrodes by emphasizing the electrodes carrying more discriminative information and de-emphasizing irrelevant electrodes.

In our experiments, the two datasets were first filtered by a six-order 8-30Hz bandpass filter and then segmented into trials. For DatasetIIa, the trials were captured from 3.5s to 5.5s, and DatasetIVa was captured using a window of 3s after a cue onset of 0.5s, yielding trial size of 22×500 and 118×300, respectively.

B. Experimental Setting

This paper verified the merits of the proposed eSPDA on above datasets, and compared it with other state-of-art domain adaptation algorithms. Table 1 gives the descriptions of the competing methods.

Feature extract: eSPDA extracted reduced covariance matrices as features, while the features of the competing methods were represented by tangent vectors with respect to the Riemannian means. Since the tangent vector size corresponding to DatasetIVa was 1×7021, which was excessively high considering the limited number of samples, principal component analysis (PCA) was employed to reduce the dimensionality to 200 in the experiment.

Classifier: All domain adaptation algorithms used SVM as the basic classifier.

Settings for the sources: Preliminary experiments on DatasetIIa revealed that subjects A3, A7, A8 and A9 performed better in motor imagery tasks than the other subjects. In DatasetIVa, subject AL was the best, AV was the least skilled, and AA, AW and AY performed well. Therefore, the transfer scenarios for DatasetIIa were set as A8->A1, A8->A2… A3->A8, A8->A9, and AY->AA, AY->AL… AW->AY for DatasetIVa. All experiments were conducted using the calibration data.

Measurement: The performance of DatasetII and DatasetIV was evaluated by classification accuracy (Acc), positive precision (Pre), positive recall (Rec), and F-measure (F1). F-measure is defined as the mean of the precision and recall indicators, providing a balanced measure of both indicators.

Accuracy measures the overall correctness of the classification results, defined as:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (18)$$

where TP and TN represent instances that are correctly classified as positive and negative, respectively. FP and FN are misclassified as positive and negative instances, respectively.

Precision measures the ratio of correctly classified true positives instances out of all samples predicted as positive:

$$Pre = \frac{TP}{TP + FP} \quad (19)$$

Recall measures the proportion of correctly classified true positive samples out of all actual positive samples:

$$Rec = \frac{TP}{TP + FN} \quad (20)$$

TABLE 1.
DESCRIPTIONS OF COMPETING ALGORITHMS AND PARAMETERS.

Methods	Descriptions	Parameters
TCA	Minimizing the marginal probability distribution difference in RKHS [33]	none
JDA	Minimizing the joint distribution difference of marginal and conditional probability in RKHS [34]	$\lambda = 0.1$
BDA	The principle of subspace learning is similar to JDA, but BDA can adaptively adjust the weight of marginal distribution and conditional distribution [35]	$\lambda = 0.1$
MEDA	Aligning the source and target features in Grassmann manifold and then learning a classifier to minimizing the dynamic distribution discrepancy and preserving the geometrical property of nearest points in manifold [37]	$\lambda = 10$ $\eta = 0.1$, $\rho = 1$
MEKT	Whitening the covariance matrices of source and target in Riemannian manifold, and learning two subspaces to reduce the domain divergences [28]	$\alpha = 0.01$ $\beta = 0.1$, $\rho = 20$

C. Results

1) Visualization of Riemannian Alignment Method

This section visualized the process and effect of the proposed Riemannian alignment algorithm on two sets of generated SPD matrices in S_+^2. The steps of the synthetic datasets are as follows:

Generate a base SPD matrix $X_i = \frac{1}{2}\begin{bmatrix} 1 & -\sin(\phi_i) \\ -\sin(\phi_i) & 1 \end{bmatrix}$ $i = 1, \ldots, 100$, where X_i is governed by a one-dimensional variable ϕ_i and ϕ_i is uniformly drawn from $[-\pi/2, 0]$.

Generate the target subset $T_i = R * X_i * R^T$, where R is randomly chosen.

Generate the source subset $S_i = R_s * X_i * R_s^T$, where $R_s = 1.5\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} * R$.

For further explanations of the parameters involved in the datasets, please refer to [25].

Fig. 3 depicts the two subsets marked by ϕ_i (left), and domain category (right), where the black points indicate the boundaries of the cone manifold, the center point indicates the identity matrix, and the red line, an extension of the identity matrix indicates the center of the cone. Fig. 3(a) visualizes the raw 2×2 SPD matrices of the target and source in $\mathbb{R}^3$, which highlights the significant differences between the two subsets in terms of scope, directional distribution, and structure. Fig. 3(b) presents the source subset after de-correlation, where the Riemannian mean of the de-correlation source is approximate to the identity matrix. Fig. 3(c) depicts the source subset reconstructed with the eigenvectors of the target. It shous that the Riemannian mean and the principal component directions of the source are nearly close to those of the target subset, which indicates that the proposed Riemannian alignment method effectively eliminates the discrepancy caused by R and R_s while preserving the inherent structure determined by ϕ_i. It is worth noting that the Riemannian alignment completely preserves the intrinsic distribution characteristics of the target.

2) Visualization of the Feature Transfer

Fig. 4 shows the feature distribution of TCA, JDA, BCD, MEDA, MEKT, and eSPDA in the scenario of A8->A3. As shown in Fig. 4(a), the raw features differ in both conditional and marginal distributions between the A8 and A3. TCA successfully aligns the marginal distributions but fails to align the conditional distribution, leading to a significant distribution discrepancy (Fig. 4(b)), which indicates that only considering the marginal distribution between domains is not enough to reduce the domain shift. JDA and BDA improve the alignment effect of both conditional and marginal distributions compared to TCA (Fig. 4(c) and (d)). However, an apparent distribution discrepancy can still be observed in these methods, indicating that they do not completely eliminate the domain shift. MEDA further reduces the divergence in both conditional and marginal distribution, but there is still a large distance between the intraclass features, suggesting that MEDA ignores variations within class (Fig. 4(e)). Both MEKT and eSPDA minimize the distribution differences and reduce the intraclass distance. However, MEKT results in identically and uniformly distributed features in both the source and target domains, which destroys the structure of the target domain (Fig. 4(f)). By contrast, eSPDA reconstructs the source covariance matrices using the Riemannian mean of the target, preserving the distribution

characteristics of the target and minimizing the domain shift as well (Fig. 4(g)).

3) Visualization of Projected Matrix

The projection matrix of eSPDA is similar to the spatial filters of CSP, with each column corresponding to a principal axis. The elements of each column are the weights of the electrodes, which reflects the importance of electrode in the mental task. Fig. 5 describes the topographical maps of the largest spatial filter corresponding to CSP and our method in typical transfer scenarios. It is observed that the distribution of eSPDA is similar to that of the CSP, suggesting that eSPDA achieves a similar effect as the CSP method. In eSPDA, despite the absence of target labels, the discriminative features of the target data are still captured by leveraging the labeled source data and the inherent structure of the target data.

TABLE 2.
CLASSIFICATION ACCURACY OF DIFFERENT DOMAIN ADAPTATION ALGORITHMS ON DATASETIVA.

	TCA	JDA	BDA	MEDA	MEKT	Ours
AY->AA	63.65	70.45	71.87	74.13	72.07	**80.5**
AY->AL	87.57	91.63	93.71	93.71	91.63	**95.71**
AY->AV	60.3	63.36	**67.64**	63.91	63.46	67.32
AY->AW	62.2	68.22	71.69	**76.07**	70.27	74.41
AW->AY	72.22	69.39	77.78	84.72	74.31	**85.42**
Average	69.19	72.61	76.54	78.51	74.35	**80.67**

TABLE 3.
CLASSIFICATION ACCURACY OF DIFFERENT DOMAIN ADAPTATION ALGORITHMS ON DATASETIIA. *P*-VALUES ARE DERIVED BY WILCOXON SIGNED-RANK TEST BETWEEN RESULTS OF ESPDA AND OTHER METHODS.

	TCA	JDA	BDA	MEDA	MEKT	Ours
A8->A1	62.49	68.13	69.37	**73.7**	69.64	73.21
A8->A2	50.4	56.32	54.04	60.81	56.76	**63.33**
A8->A3	65.42	68.63	70.32	77.06	75.34	**80.12**
A8->A4	51.48	52.73	51.48	56.76	53.33	**60.81**
A8->A5	51.9	55.17	52.39	**57.04**	52.73	56.76
A8->A6	49.96	51.64	53.7	57.28	53.32	**58.73**
A8->A7	59.09	61.33	63.63	67.7	65.33	**69.59**
A3->A8	64.58	61.11	73.61	78.86	73.56	**88.89**
A8->A9	66.2	67.17	65.33	78.97	70.1	**81.28**
Average	57.95	60.25	61.54	67.58	63.35	**70.35**
p-value	**.0054**	**.0298**	**.0304**	.4953	**.0382**	--

4) Classification Effect

Tables 2 to 5 report the classification results of the TCA, JDA, BDA, MEDA, MEKT, and our algorithms on the two datasets. Overall, eSPDA performed best, with an average of 70.35% on DatasetIIa and 80.67% on DatasetIVa. MEDA had an average accuracy of 67.58 and 78.51% on the two datasets, second only to eSPDA, and even higher than eSPDA in scenarios A8->A1, A8->A5, AY->AV and AY->AW. However, as a feature matching algorithm, MEDA has the problem of being unable to eliminate the inherent features biases, and there is interference caused by redundant

information in the features during the domain adaptation. In contrast, eSPDA is a transfer approach based on feature representation that learns a latent feature subspace across domains while extracting the most discriminative features, reducing information redundancy between features. Wilcoxon's signed-rank test (one-sided) was performed to verify the significant improvement of eSPDA over other methods on DatasetIIa. The results show that eSPDA significantly outperformed the other algorithms, except MEDA ($p < 0.05$).

TABLE 4. PRECISION, RECALL AND FI VALUES FOR POSITIVE INSTANCES OF DATASETIVA.

		TCA	JDA	BDA	MEDA	MEKT	Ours
P r e	AY->AA	63.38	70.50	70.75	78.15	79.61	**76.07**
	AY->AL	86.21	93.94	94.59	98.41	89.73	**93.79**
	AY->AV	61.29	65.29	**69.60**	66.10	64.12	65.58
	AY->AW	62.14	67.83	**73.44**	76.47	66.09	69.54
	AW->AY	73.48	69.57	70.16	92.38	71.52	**85.61**
R e c	AY->AA	64.29	70.00	74.29	66.43	58.57	**88.57**
	AY->AL	89.29	88.57	89.74	88.57	93.57	**97.14**
	AY->AV	54.29	56.43	62.14	55.71	60.00	**72.14**
	AY->AW	62.14	69.29	67.14	74.29	82.14	**86.43**
	AW->AY	69.29	68.57	**95.71**	73.48	80.71	85.00
F1	AY->AA	63.83	70.25	72.47	71.81	67.49	**81.85**
	AY->AL	87.72	91.18	92.11	93.23	91.61	**95.44**
	AY->AV	57.58	60.54	65.66	60.47	61.99	**68.71**
	AY->AW	62.14	68.55	70.15	75.36	73.25	**77.07**
	AW->AY	71.32	69.06	80.97	81.86	75.84	**85.30**

5) Convergence and Time Complexity

This section explored the convergence and computational costs of TCA, JDA, BDA, MEDA, MEKT, and our methods. The classification accuracy and training costs of 15 iterations were recorded under the scenarios of AW->AY and A3->A8. As depicted in Fig. 6, eSPDA achieves convergence within six iterations, suggesting that by iteratively adjusting the target pseudo-labels, the difference in eSPDA conditional distribution gradually decreases, leading to an improvement in classifier performance. TCA does not require iterations, and accuracy tends to be low when the raw conditional distributions of the source and target differ significantly. Nevertheless, TCA is considered a simple and fast method that serves as an effective preprocessing technique to alleviate domain shifts in conjunction with other transfer methods.

Table 6 summarizes the time complexity of different domain adaptation algorithms and the training time under the A3->A8 and AW->AY scenarios. In table 6, n and m respectively denote the dimensionality of the covariance matrix of an EEG trial and the tangent feature space, N represents the total number of target and source training samples, q is the dimension of the learned subspace, T indicates the number of iterations, and M indicates the number of iterations to obtain the optimal transformation matrix in the eSPDA algorithm. It is observed that eSPDA has the largest computational complexity and longest training time among the considered algorithms. The reasons are

twofold: (1) eSPDA involves more matrix multiplication and decomposition operations $O(T_6(M(n^3 + q_6 n^2))$ during optimization; (2) in each iteration, it is necessary to re-estimate the Riemann mean of the target, which involves the transformation of flattening the covariance matrix into a tangent vector $O(T_6 N t_{(c)} n^3)$. The computational burden of other algorithms is primarily attributed to the generation of tangent vectors ($O(N n^3)$), and the higher the dimension of covariance matrix, the greater the computational cost of tangent vector features. Although eSPDA method may not be competitive in term of training time, it has advantages in maintaining the geometric structure of SPD matrix-value features. In the testing phase, the covariance matrix of the EEG signal can be directly used as input without the need for tangent vector transformation, greatly improving the testing speed. Therefore, considering the performance of the model comprehensively, the eSPDA method has significant advantages in application scenarios without considering training time.

TABLE 5. PRECISION, RECALL AND FI VALUES FOR POSITIVE INSTANCES OF DATASETIIA.

		TCA	JDA	BDA	MEDA	MEKT	Ours
P r e	A8->A1	59.36	74.07	68.21	**72.67**	67.07	70.63
	A8->A2	50.40	54.89	53.90	58.86	55.69	**62.84**
	A8->A3	68.03	68.28	70.14	74.21	73.38	**76.22**
	A8->A4	51.67	53.93	51.35	56.46	53.02	**61.48**
	A8->A5	52.81	56.60	52.24	55.38	51.85	**58.88**
	A8->A6	49.64	51.92	53.29	**57.14**	52.41	56.41
	A8->A7	59.70	61.43	**65.35**	64.88	60.28	65.22
	A3->A8	73.03	60.40	74.10	78.62	76.80	**87.84**
	A8->A9	68.25	65.61	62.64	77.12	70.50	**80.82**
R e c	A8->A1	77.08	55.56	71.53	75.69	76.39	**78.47**
	A8->A2	43.75	70.14	52.78	**71.53**	64.58	64.58
	A8->A3	57.64	68.75	70.14	81.94	78.47	**86.81**
	A8->A4	43.06	33.33	52.78	57.64	54.86	**57.64**
	A8->A5	32.64	41.67	48.61	**71.53**	68.06	43.75
	A8->A6	47.22	37.50	56.25	55.56	68.06	**76.39**
	A8->A7	55.56	59.72	57.64	75.69	**89.58**	83.33
	A3->A8	45.14	62.50	71.53	79.17	66.67	**90.28**
	A8->A9	59.72	71.53	75.69	81.94	68.06	**81.94**
F1	A8->A1	67.07	63.49	69.83	74.15	71.43	**74.34**
	A8->A2	46.84	61.59	53.33	**64.58**	59.81	63.70
	A8->A3	62.41	68.51	70.14	77.89	75.84	**81.17**
	A8->A4	46.97	41.20	52.05	57.04	53.92	**59.50**
	A8->A5	40.34	48.00	50.36	**62.42**	58.86	50.20
	A8->A6	48.40	43.55	54.73	56.34	59.21	**64.90**
	A8->A7	57.55	60.56	61.25	69.87	72.07	**73.17**
	A3->A8	55.79	61.43	72.79	78.89	71.38	**89.04**
	A8->A9	63.70	68.44	68.55	79.46	69.26	**81.38**
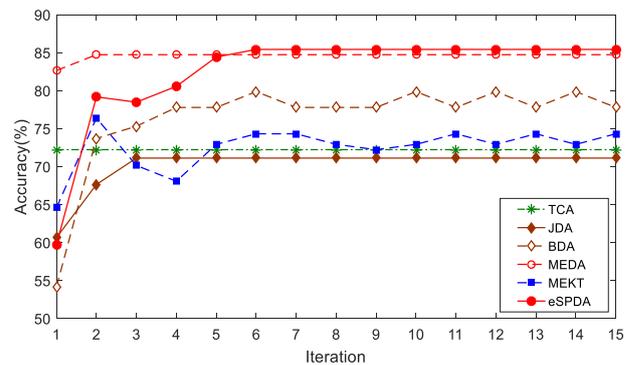
(a) Convergence in AW->AY scenario
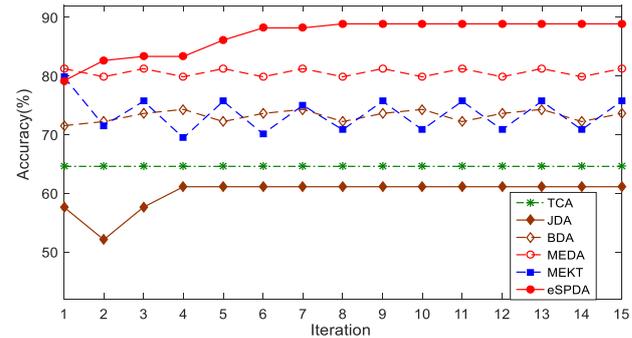
(b) Convergence in A3->A8 scenario
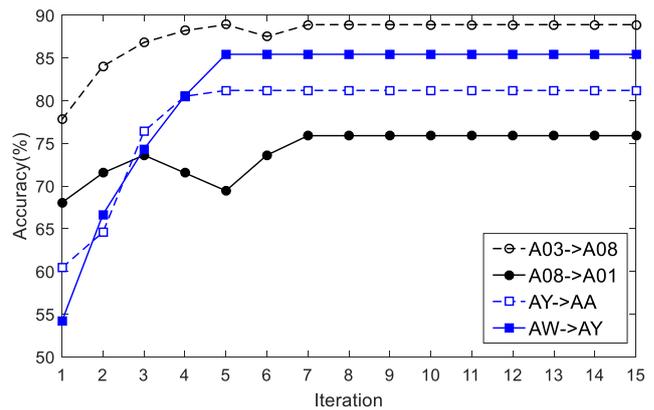
(c) Convergence of eSPDA

Fig. 6. Convergence analysis of TCA, JDA, BDA MEDA, MEKT, and eSPDA.

TABLE 6.
COMPARISON OF COMPUTATION COMPLEXITY AND TRAINING TIME IN AW -> AY AND A3 -> A8 SCENARIOS.

Methods	Computational Complexity	Training Time (s)	
		A3->A8	AW->AY
TCA	$O(N n^3 + q_1 N^2)$	0.54	7.23
JDA	$O(N n^3 + T_2(q_2 m^2 + N^2 + mN))$	1.4	7.92
BDA	$O(N n^3 + T_3(q_3 m^2 + N^2 + mN))$	1.29	7.82
MEDA	$O(N n^3 + q_4 N^2 + T_4(N^2 + q_4 N^2))$	1.36	8.01
MEKT	$O(N n^3 + q_5 N^2 + T_5(N^2 + q_5 N^2))$	1.99	8.44
Ours	$O(T_6(N t_{(c)} n^3 + M(n^3 + q_6 n^2) + N^2))$	3.05	25.04

6) Balance Factor

This section investigated the influence of the balance factor in the eSPDA algorithm in scenarios A8->A1, A3->A8, AY->AA and AW->AY. Fig. 7(a) illustrates how the accuracy varied with different values of balance factor μ. It is observed that the optimal μ value varied across the

different scenarios, which validates the necessity and effectiveness of dynamically adjusting the weight of conditional and marginal distribution differences. It is also found that the optimal μ is not unique, instead, the best choice is $\mu \in [0.4 \quad 1]$, indicating that eSPDA can achieve satisfactory performance with a wide range of values of μ. According to the estimated μ value in each iteration in Fig. 7(b), it is observed that the μ gradually converges to the range of $(0.8 \quad 1)$, meanwhile, the classification accuracy gradually converges. This indicates that eSPDA assigns greater weight to the conditional distribution. This result can be explained by the proposed Riemannian alignment method aligning the principal components of the target and the source, thereby making subspace learning more constrained by differences in conditional distributions.

V. CONCLUSIONS

This paper introduces a domain adaptation algorithm based on manifold embedding (eSPDA) for cross-subject feature transfer. In eSPDA, the rationale of CSP is formulated as a dimensionality reduction that maximizes the inter-class distance and intraclass variance in a Riemannian manifold. Two dimensionality reduction frameworks in Riemannian manifold are defined: one for extracting discriminative information from labeled source data by maximizing the inter-class distance, and another for preserving the principal characteristics of the unlabeled target data by maximizing the intra-class variance. The domain adaptation techniques are then integrated to align the distribution of source and target to ensure the classifier trained on the labeled source domain be suitable for the target. Extensive experiments on two public BCI datasets demonstrated the effectiveness of eSPDA in cross-subject transfer. The results suggests that eSPDA overcomes the need for a large-scale target training set and reduces the calibration time of the MI-BCI system. It is worth noting that this paper focuses on the classification of target data in a transductive setting, where the unlabeled data is available. Future work will explore the effectiveness of eSPDA in the inductive transfer setting, where the learned subspace will be adjusted using the target data to make it more suitable for unseen target data.
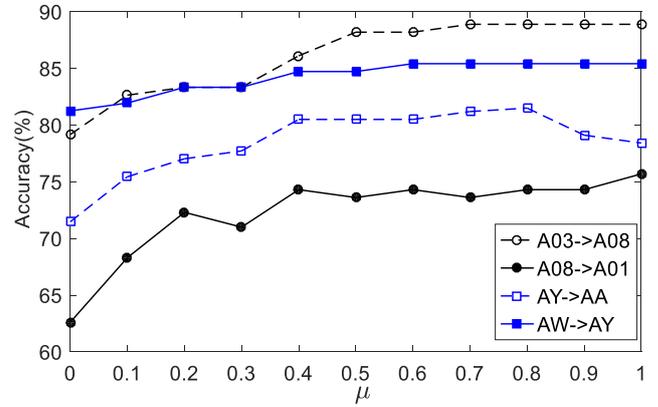
(a) classification accuracy w.r.t. μ
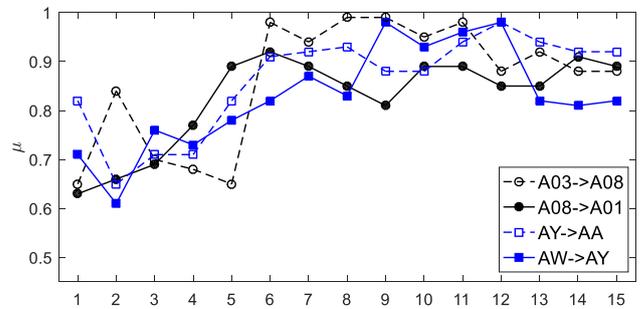
(b) μ w.r.t. iterations

Fig. 7. Parameter sensitivity of eSPDA in different scenarios.
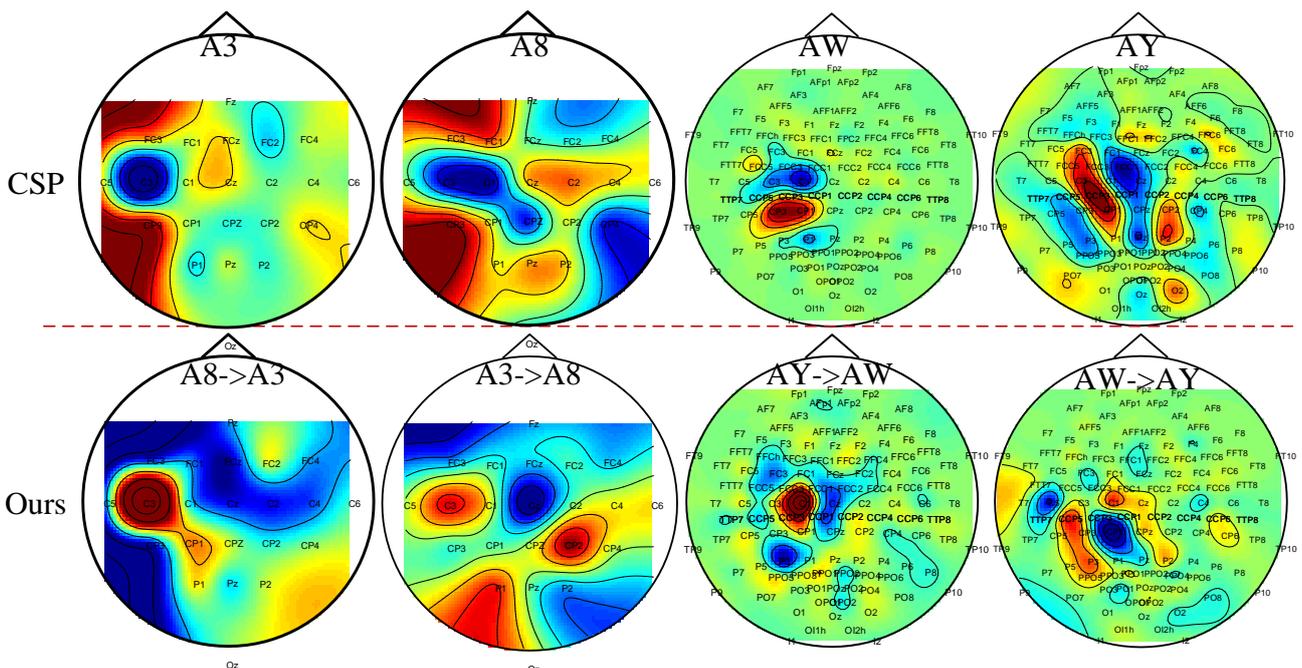
Fig. 5. Topographic map of the CSP and eSPDA
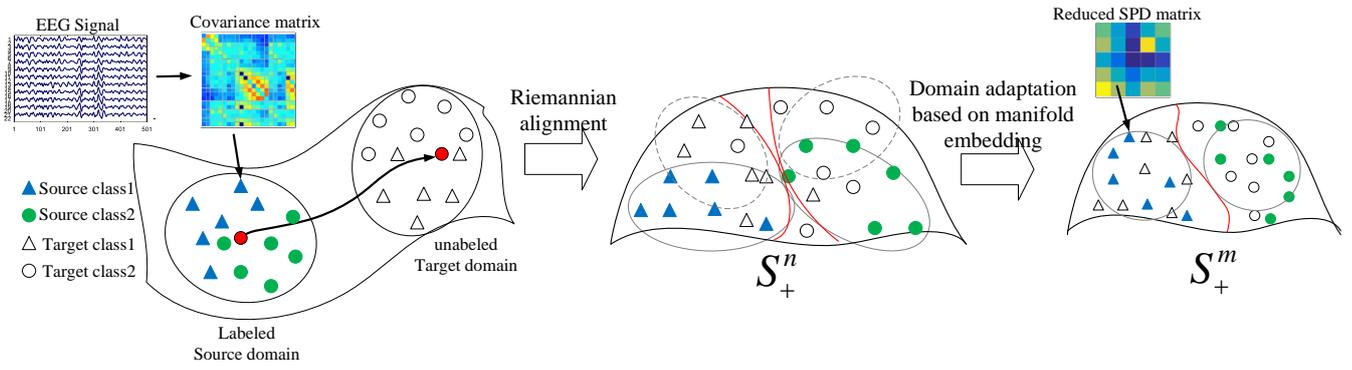
Fig. 1 Schematic diagram of eSPDA.
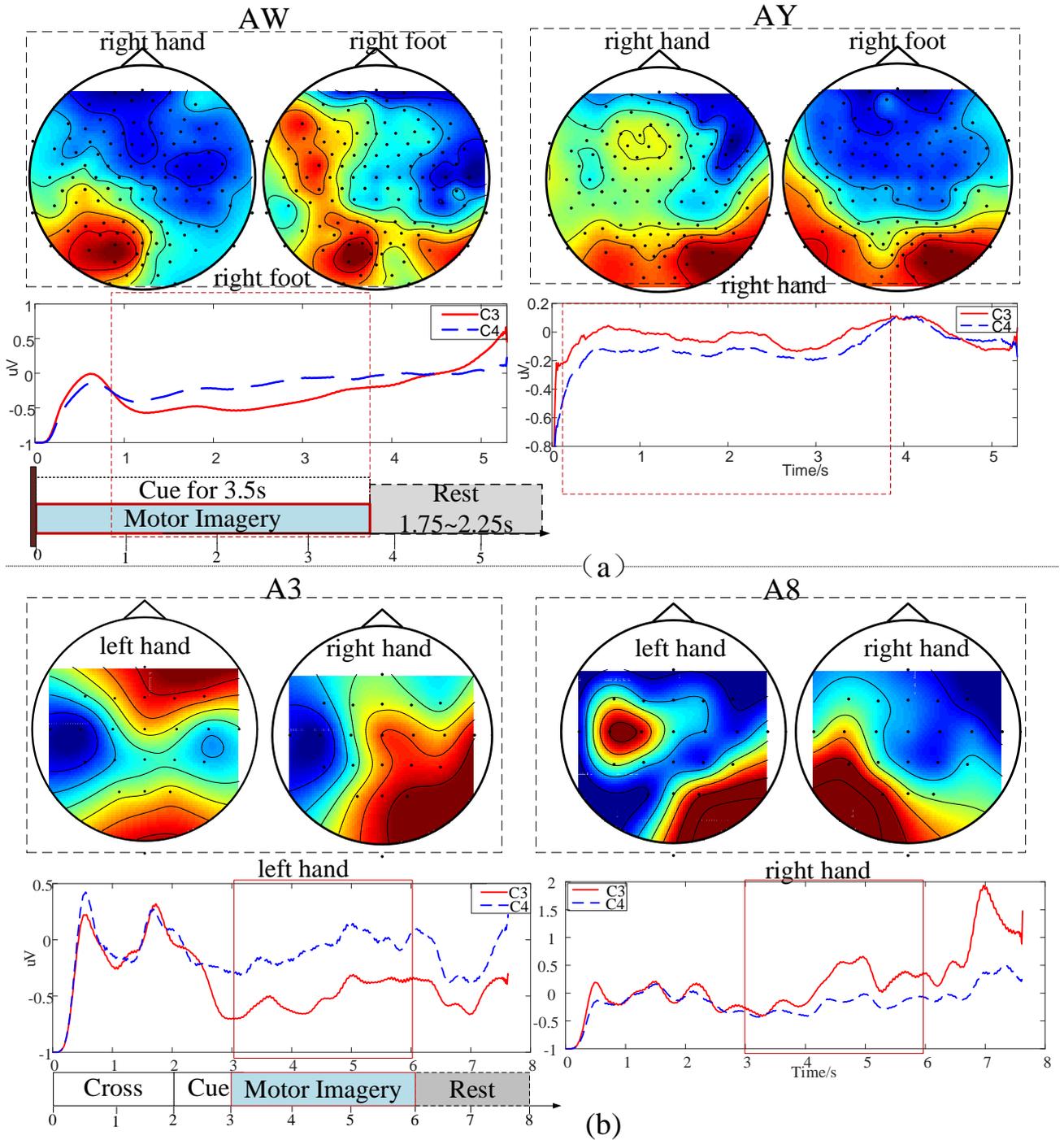
Fig. 2. ERD distributions and the timing schedules of the datasets, (a) ERD distribution and timing schedule of DatasetIVa; (b) ERD distribution and timing schedule of DatasetIIa.
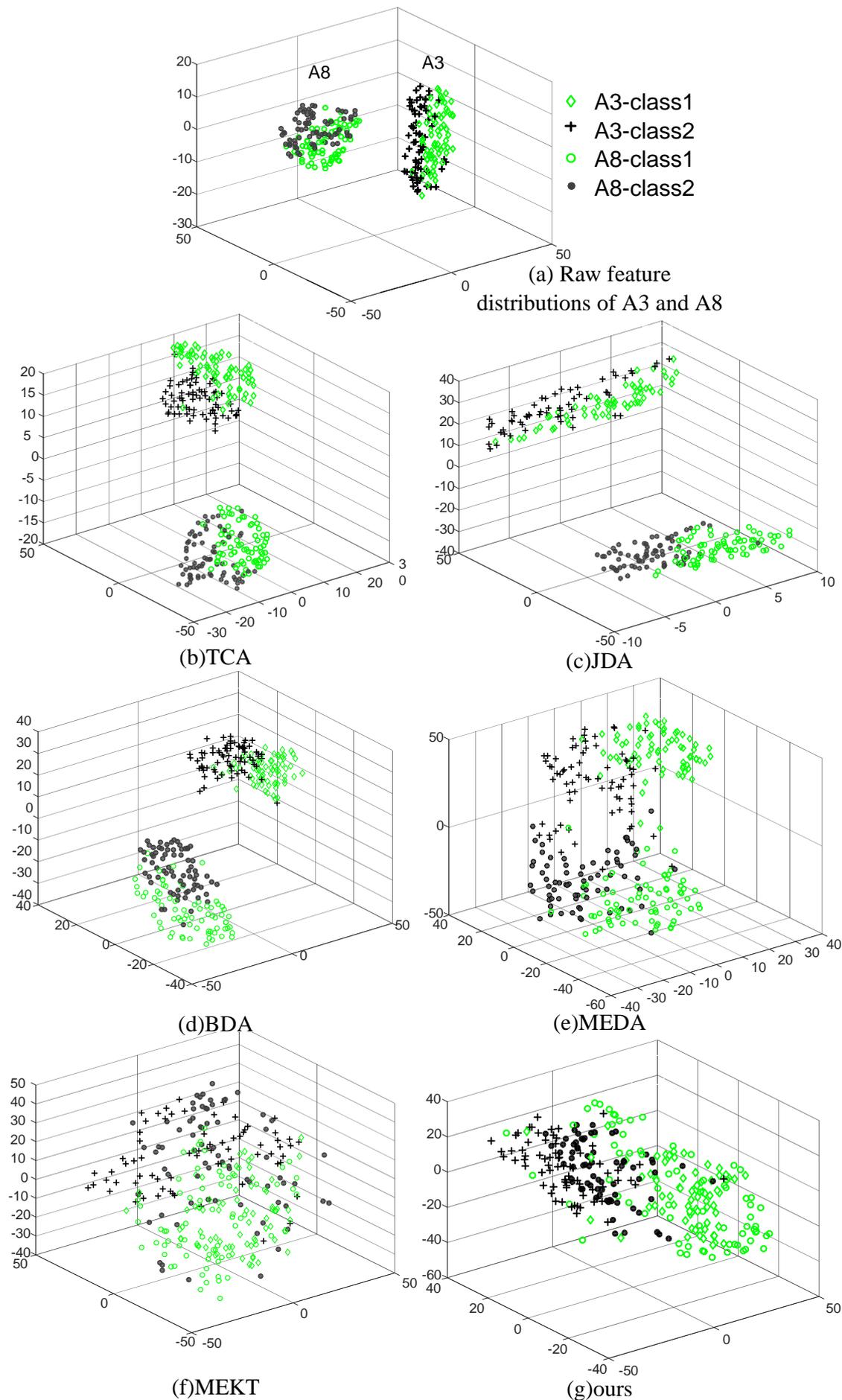
Fig. 4. The visualization of transferring source data to classify the unlabeled target data by TCA, JDA, BDA, MEDA, MEKT and KMDA under A8->A3 scenario.
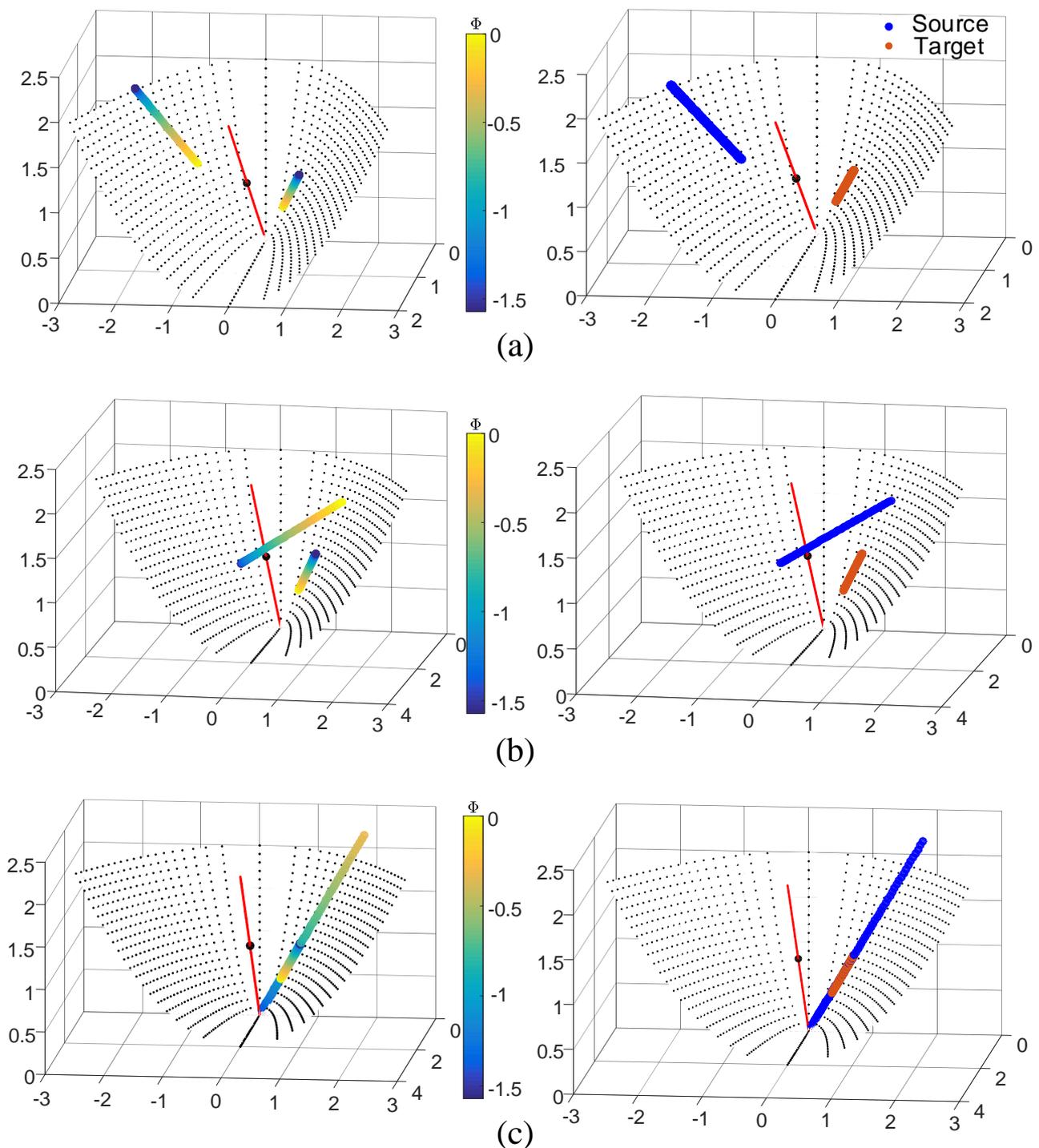
Fig. 3. Visualization of Riemannian alignment steps on synthetic data, (a) raw distribution, (b) whitening the source subset, (c) reconstructing the source subset.

REFERENCES

[1] Said Abenna, Mohammed Nahid, Abderrahim Bajit, "Motor imagery-based brain-computer interface: improving the EEG classification using Delta rhythm and LightGBM algorithm," *Biomedical Signal Processing and Control*, Vol.71, Part A, pp.103102, 2022.

[2] Zapała D., Zabielska-Mendyk E., Augustynowicz P., et al.,"The effects of handedness on sensorimotor rhythm desynchronization and motor-imagery BCI control," *Scientific reports*, vol. 10, no. 1, pp 1-11, 2020.

[3] Zhangfang Hu, Lingxiao He, and Haoze Wu, "A Multi-feature Fusion Transformer Neural Network for Motor Imagery EEG Signal Classification," *Engineering Letters,* Vol. 31, no.4, pp1822-1831, 2023.

[4] Yuliana S., Riyadi M.," Implementation of brain computer interface (BCI) as a smart wheelchair motion command," *Journal of Biomedical Science and Bioengineering,* vol. 2, no. 1, 2937097, 2022.

[5] Yuriy M., Murat K., Erkan O., et al.,"Developing a 3-to 6-state EEG-based brain-computer interface for a virtual robotic manipulator control," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 4, pp 977-987, 2019.

[6] Dipti Pawar, and Sudhir Dhage, "Feature Extraction Methods for Electroencephalography based Brain-Computer Interface: A Review," *IAENG International Journal of Computer Science*, vol. 47, no.3, pp501-515, 2020.

[7] Lotte F., Guan C.," Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering*, vol. 58, no. 2, pp 355-362, 2010.

[8] Wang B., Wong C. M., Kang Z., et al.," Common spatial pattern reformulated for regularizations in brain–computer interfaces," *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp 5008-5020, 2020.

[9] Park S. H., Lee D., Lee S. G.," Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp 498-505, 2017.

[10] Park Y., Chung W.," Frequency-optimized local region common spatial pattern approach for motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp 1378-1388, 2019.

[11] Zhang J., Wang X., Xu B., et al.," An overview of methods of left and right foot motor imagery based on Tikhonov regularisation common spatial pattern," *Medical & Biological Engineering & Computing*, pp 1-10, 2023.

[12] Mishuhina V., Jiang X.," Complex common spatial patterns on time-frequency decomposed EEG for brain-computer interface," *Pattern Recognition*, vol. 115, pp 107918, 2021.

[13] Kang H., Nam Y., Choi S.,"Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp 683-686, 2009.

[14] Wan Z., Yang R., Huang M., et al.," A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1-14, 2021.

[15] Yilu X, Qingguo W, Hua Z, et al.,"Transfer Learning Based on Regularized Common Spatial Patterns Using Cosine Similarities of Spatial Filters for Motor-Imagery BCI,"*Journal of Circuits, Systems and Computers*, vol. 28, no. 7, 1950123, 2019.

[16] Jiao Y., Zhang Y., Chen X., et al.,"Sparse group representation model for motor imagery EEG classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp 631-641, 2018.

[17] Zhang L., Gao X.,"Transfer adaptation learning: A decade survey," IEEE *Transactions on Neural Networks and Learning Systems*, 2022.doi:10.1109/TNNLS.2022.3183326.

[18] Wu D, Jiang X, Peng R.," Transfer learning for motor imagery-based brain-computer interfaces: A tutorial," *Neural Networks*, vol. 153, pp 235-253, 2022.

[19] Peng H. Lei C., Zheng, S., et al.,"Automatic epileptic seizure detection via Stein kernel-based sparse representation," *Computers in Biology and Medicine*, vol. 132, 104338, 2021.

[20] Harandi M., Salzmann M, Hartley R.,"Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp 48-62, 2017.

[21] Jin J., Xiao R., Daly I., et al.,"Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory," *IEEE transactions on neural networks and learning systems*, vol.32, no. 11, pp 4814-4825, 2020.

[22] Barachant A., Bonnet S., Congedo M., et al.,"Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol.59, no. 4, pp 920-928, 2011.

[23] Congedo M., Barachant A., Bhatia R.,"Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp 155-174, 2017.

[24] Zanini P., Congedo M., Jutten C., et al.,"Transfer learning: A Riemannian geometry framework with applications to brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp 1107-1116, 2017.

[25] Yair O., Ben-Chen M., Talmon R.,"Parallel transport on the cone manifold of SPD matrices for domain adaptation," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp 1797-1811, 2019.

[26] Rodrigues P. L. C., Jutten C., Congedo M.,"Riemannian procrustes analysis: transfer learning for brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp 2390-2401, 2018.

[27] Cai Y., She Q., Ji J., et al.,"Motor imagery EEG decoding using manifold embedded transfer learning," *Journal of Neuroscience Methods*, vol. 370, 109489, 2022.

[28] Zhang W., Wu D.,"Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp 1117-1127, 2020.

[29] Xu Y., Huang X., Lan Q.,"Selective cross-subject transfer learning based on Riemannian tangent space for motor imagery brain-computer interface," *Frontiers in Neuroscience*, vol. 15, pp 1487-1497, 2021.

[30] Kumar S., Mamun K., Sharma A.,"CSP-TSM: Optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI," *Computers in Biology and Medicine*, vol. 91, pp 231-242, 2017.

[31] Yger F., Berar M., Lotte F.,"Riemannian approaches in brain-computer interfaces: a review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp 1753-1762, 2016.

[32] Borgwardt K. M., Gretton A., Rasch M. J., et al.," Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp e49-e57, 2006.

[33] Pan S. J., Tsang I. W., Kwok J. T., et al.,"Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp 199-210, 2010.

[34] Long M., Wang J., Ding G., et al.,"Transfer feature learning with joint distribution adaptation[C]. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, pp 2200-2207, 2013.

[35] Wang J., Chen Y., Hao S., et al.,"Balanced distribution adaptation for transfer learning," In Proceedings of 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, pp 1129-1134, 2017.

[36] Zhang W., Wu D. (2019). Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation. (Online). Available: https://arxiv.org/pdf/ 1912.00320v4.pdf.

[37] Wang J., Feng W., Chen Y., et al.,"Visual domain adaptation with manifold embedded distribution alignment," In Proceedings of the 26th ACM international conference on Multimedia, Seoul Republic of Korea, pp 402-410, 2018.

[38] Gong B., Shi Y., Sha F., et al.,"Geodesic flow kernel for unsupervised domain adaptation," In Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp 2066-2073, 2012.

[39] Sun B., Feng J., Saenko K. (2015). Return of frustratingly easy domain adaptation. (Online). Available: https://arxiv.org/pdf/1511.05547.pd.

[40] Chen K. X., Ren J. Y., Wu X. J., et al.,"Covariance descriptors on a gaussian manifold and their application to image set classification. *Pattern Recognition*, vol. 107, pp. 107463-107470, 2020.

[41] Jayasumana S., Hartley R., Salzmann M., et al.," Kernel methods on Riemannian manifolds with gaussian RBF kernels," *EEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 12, pp. 2464-2477, 2015.

[42] Jiang Q., Zhang Y., Zheng K., "Motor Imagery Classification via Kernel-Based Domain Adaptation on an SPD Manifold," *Brain Sciences*, vol. 12, pp. 659, 2022