A Novel Transfer Support Matrix Machine for Motor Imagery-Based Brain Computer Interface

Yan Chen[1,2†], Wenlong Hang[2†], Shuang Liang[3], Xuejun Liu[1], Guanglin Li[4], Qiong Wang[5*], Jing Qin[6] and Kup-Sze Choi[6]

[1] School of Computer Science and Technology, Nanjing Tech University, Nanjing, China, [2] Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University, Nanjing, China, [3] Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing, China, [4] CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [5] Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [6] School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

In recent years, emerging matrix learning methods have shown promising performance in motor imagery (MI)-based brain-computer interfaces (BCIs). Nonetheless, the electroencephalography (EEG) pattern variations among different subjects necessitates collecting a large amount of labeled individual data for model training, which prolongs the calibration session. From the perspective of transfer learning, the model knowledge inherent in reference subjects incorporating few target EEG data have the potential to solve the above issue. Thus, a novel knowledge-leverage-based support matrix machine (KL-SMM) was developed to improve the classification performance when only a few labeled EEG data in the target domain (target subject) were available. The proposed KL-SMM possesses the powerful capability of a matrix learning machine, which allows it to directly learn the structural information from matrix-form EEG data. In addition, the KL-SMM can not only fully leverage few labeled EEG data from the target domain during the learning procedure but can also leverage the existing model knowledge from the source domain (source subject). Therefore, the KL-SMM can enhance the generalization performance of the target classifier while guaranteeing privacy protection to a certain extent. Finally, the objective function of the KL-SMM can be easily optimized using the alternating direction method of multipliers method. Extensive experiments were conducted to evaluate the effectiveness of the KL-SMM on publicly available MI-based EEG datasets. Experimental results demonstrated that the KL-SMM outperformed the comparable methods when the EEG data were insufficient.

Keywords: motor imagery, brain-computer interface, electroencephalography, support matrix machine, transfer learning

INTRODUCTION

Brain-computer interface (BCI) systems enable machines to accurately perceive the mental states of human beings, thereby establishing an effective user interface between humans and machines. There are several kinds of BCI paradigms, such as steady-state visual evoked potentials (Allison et al., 2008), P300 (Salvaris and Sepulveda, 2009), and motor imagery (MI)

(Pfurtscheller and Neuper, 2001). Among them, the MI-based BCI is widely used because of its self-paced fashion, and it does not require any external stimuli (Pfurtscheller and Da Silva, 1999). Electroencephalography (EEG) is the most extensively used technique to record neuronal activity in the brain due to its high temporal resolution, portability, and non-invasiveness. EEG-based motor imagery BCI has shown great potential in many applications, such as rehabilitating the sensory-motor functions of disabled patients (Ang et al., 2011; Al-Qaysi et al., 2018) and facilitating smart living for healthy people (Vourvopoulos et al., 2017; Wang et al., 2019).

Although many machine learning algorithms have been developed to implement MI-based BCI with great success, most of them need to collect a considerable amount of labeled EEG data for model training, which is exceedingly time-consuming and labor-intensive. Insufficient labeled EEG data weaken the generalization capability of the classifier in the prediction. An intuitive solution to this problem is to leverage historical EEG data from the source domain (source subject) in modeling the target domain (target subject). However, this approach may engender some challenges. Owing to the EEG pattern variations between different subjects (Morioka et al., 2015), directly using the EEG data of the source domain may cause performance degradation. Furthermore, because the original EEG data contains personal information, the data of other subjects may not always be available for constructing the classifier for privacy reasons (Agarwal et al., 2019). Thus, exploring an effective knowledge transfer strategy that can protect the personal information of a source subject is highly desirable in the MI-based BCI.

From the perspective of transfer learning (Pan and Yang, 2009), the model knowledge of the source domain can potentially be leveraged to address these problems. Generally, EEG-based learning methods involve two steps: EEG feature extraction and classification. The model knowledge of the source domain can either be integrated into the feature extraction process (Kang et al., 2009; Samek et al., 2013), or be used in modeling the classifier (Azab et al., 2019). Specifically, Kang et al. (2009) proposed leveraging the linear combination of covariance matrices of the source subjects as reference during the feature extraction of the target EEG data. Azab et al. (2019) proposed the construction of multiple-source models and transfer of the weighted multiple-source model knowledge to the target domain. Deng et al. (2013) proposed a knowledge-leverage-based fuzzy system that can leverage the model knowledge from the source domain in order to make up for the lack of labeled target data as well as privacy protection.

Although it has been empirically demonstrated that the aforementioned methods are effective in dealing with EEG classification in scenarios where the labeled data are limited, these methods always need to transform the input data into vectors before classification. It is well known that EEG signals record brain activities over a period of time from multiple channels, which are naturally represented as matrices. Transforming the input matrices into vectors may destroy the correlation of rows or columns within matrix-form EEG features. Thus, several

classification methods that can directly handle these matrix-form data have been developed accordingly. For example, Wolf et al. (2007) proposed modeling the regression matrix of a support vector machine (SVM), which is the sum of the k rank-one orthogonal matrices (rank-k SVM). Pirsiavash et al. (2009) proposed a bilinear SVM (BSVM) based on factorizing the regression matrix into the product of two low-rank matrices. Although these methods can capture the correlation within matrix data, pre-determining the rank of the regression matrix requires a tedious tuning procedure. Luo et al. (2015) proposed combining the nuclear norm and squared Frobenius norm of the regression matrix to derive the support matrix machine (SMM). The cornerstone of the SMM uses the nuclear norm of the regression matrix as the convex approximation of the matrix rank; thus, its optimization problem becomes more tractable and can be solved using the alternating direction method of multipliers (ADMM) method. Based on SMM, Zheng et al. proposed multiclass SMM (Zheng et al., 2018c) and sparse SMM (Zheng et al., 2018b) for EEG data. Although existing matrix classification methods can effectively deal with the matrix-form EEG data, they have not taken the transferrable knowledge into consideration to improve EEG classification performance. They may suffer from the weak generalization capability when the available EEG data are insufficient.

We propose a novel knowledge-leverage-based matrix classification method for MI-based EEG classification at the first time. The proposed knowledge-leverage-based SMM (KL-SMM) can address the above-mentioned problems by integrating the model knowledge from the source domain and a few labeled target EEG data. It possesses the powerful capability of the SMM for learning matrix-form data. Furthermore, the model knowledge of the source domain can be used to compensate for the deficiency in learning due to the lack of labeled target EEG data. Different from most current model parameter transfer learning methods, the proposed method can propagate the structural information from the source model to the target model. Hence, the generalization capability can be greatly enhanced by transferring the model knowledge and structural information of the source domain. Instead of directly using the source EEG data, the KL-SMM can afford privacy protection by leveraging only the model knowledge of the source domain. In addition, it can be efficiently optimized through the ADMM method. We conducted extensive experiments on two publicly available EEG datasets to validate the effectiveness of the proposed method. As demonstrated by the experimental results, the KL-SMM can achieve promising results in scenarios with few labeled target EEG data.

The remainder of this paper is organized as follows: Section "Related Works" is a review of related works. In Section "Matrix Learning Preliminaries", the notations and preliminaries of the SMM are introduced. The KL-SMM model and its learning algorithm are described in Section "Knowledge-Leverage-Based SMM". In Section "Experiments", the details of extensive experiments and analyses are presented. The conclusions of the paper are presented in Section "Conclusion".

RELATED WORKS

Transfer learning has emerged as a novel technique for retaining and reusing knowledge learned from historical tasks for new tasks. As described above, transfer learning generally refers to the knowledge-leverage-based learning mechanism, which can extract useful knowledge from the source domain and propagate them as the supervision information for modeling the target domain. According to the types of transferred knowledge of the source domain, most current research on transfer learning for EEG classification can be broadly divided into the following categories: (1) instance transfer, (2) feature representation transfer, and (3) model parameter transfer (Wang et al., 2015).

For the first category, it is assumed that the partial source EEG data can be selected and considered together with few labeled target EEG data. The source EEG data are obtained through either instance selection or importance sampling cross-validation (Li et al., 2010; Hossain et al., 2016, 2018; Zanini et al., 2018). For example, Hossain et al. (2016) proposed an instance selection strategy based on active learning. The selected source EEG data were then used together with available target-labeled EEG data to train the target model. Li et al. (2010) demonstrated the possibility of weighing the source EEG data through the importance sampling cross-validation strategy, following which the source data with high weights were used to estimate the target classifier.

The aim of the feature representation transfer method is to learn a good feature representation, which has some relevant source knowledge encoded within it, for the target subject. Most feature representation transfer learning methods were developed based on the common spatial patterns (CSP) through the modification of the covariance matrix or optimization function (Kang et al., 2009; Lotte and Guan, 2010; Samek et al., 2013). For example, Samek et al. (2013) developed an extension of the CSP. They proposed learning a stationary subspace in which the stationary information of multiple subjects can be transferred. In addition to the above-mentioned shallow feature representation transfer learning methods, several deep transfer learning methods (Fahimi et al., 2019; Hang et al., 2019) have been proposed. In general, these methods apply the domain adaptation techniques in a task-specific layer to incorporate the learned source and target deep features into a common feature space. For example, Hang et al. (2019) proposed leveraging the maximum mean discrepancy and the center-based discriminative feature learning techniques simultaneously to reduce the domain shift, demonstrating a performance improvement in the MI-based BCI.

The third category is the model parameter transfer, which assumes that the source subjects and target subjects share some parameters or prior distributions of the models. Model parameter transfer learning methods always leverage source models to model the target subjects in EEG classification. For example, Azab et al. (2019) proposed a logistic regression-based transfer learning method. The linear combination of multiple-source models was transferred for the construction of the target domain. In Alamgir et al. (2010); Jayaram et al. (2016), Alamgir et al. proposed a multi-task learning method to improve the generalization

performance of the EEG classification for individual subjects using subject-specific information, as well as the shared model knowledge inherent in all available subjects.

Despite these successes, most current transfer learning methods require the direct use of source EEG data, which may cause the issue of privacy disclosure, especially for biomedical information. Furthermore, existing transfer learning methods used for EEG recognition always built on that the input data are vectors. However, transforming EEG data, which are naturally represented as matrices, into vectors will destroy their structural information. The proposed method belongs to the third category. Unlike the previous transfer learning methods, the KL-SMM can incorporate model knowledge from the source domain, thereby guaranteeing privacy protection to some extent; it can also directly handle matrix-form EEG data.

MATRIX LEARNING PRELIMINARIES

Among the current matrix learning methods, the SMM (Luo et al., 2015) and its variants [e.g., (Zheng et al., 2018a)] are applied in many fields, owing to their simplicity and effectiveness. In this section, we present some notations and preliminary knowledge on the SMM, which are the foundation of the proposed KL-SMM method.

Mathematical Notations

Matrices are denoted by bold uppercase letters (i.e., $\mathbf{X}$) in the following. For a matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ of rank r, it can be expressed as $rank(\mathbf{X}) = r$. The condensed singular value decomposition (SVD) of $\mathbf{X}$ is denoted as $\mathbf{X} = \mathbf{U_X} \sum_{\mathbf{X}} \mathbf{V_X}^T$, where $\mathbf{U_X} \in \mathbb{R}^{d_1 \times r}$ and $\mathbf{V_X} \in \mathbb{R}^{d_2 \times r}$ satisfy $\mathbf{U_X}^T \mathbf{U_X} = \mathbf{I}_r$ and $\mathbf{V_X}^T \mathbf{V_X} = \mathbf{I}_r$, and $\sum_{\mathbf{X}} = \text{diag}\left(\sigma_1(\mathbf{X}), \sigma_2(\mathbf{X}), \ldots, \sigma_r(\mathbf{X})\right)$ with $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \cdots \geq \sigma_r(\mathbf{X}) > 0$.

Definition 1. Given any $\tau > 0$, the singular value thresholding (SVT) (Cai et al., 2010) of matrix $\mathbf{X}$ is defined as

$$D_\tau[\mathbf{X}] = \mathbf{U_X} \mathbf{S}_\tau \left[\sum_{\mathbf{X}} \right] \mathbf{V_X}^T \qquad (1)$$

where $S_\tau\left[\sum_{\mathbf{X}}\right] = \text{diag}\left(\{\sigma_1(\mathbf{X}) - \tau\}_+, \{\sigma_2(\mathbf{X}) - \tau\}_+, \ldots, \{\sigma_r(\mathbf{X}) - \tau\}_+\right)$ and $\{z\}_+ = \max(z, 0)$.

Let $\| \mathbf{X} \|_F = \sqrt{\sum_{i=1}^r \sigma_i(\mathbf{X})^2}$ be the Frobenius norm of $\mathbf{X}$, $\| \mathbf{X} \|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$ denotes the nuclear norm of $\mathbf{X}$, and the subdifferential of $\| \mathbf{X} \|_*$ can be defined as follows.

Definition 2. The subdifferential of $\| \mathbf{X} \|_*$, that is, $\partial \| \mathbf{X} \|_*$ (Candes and Recht, 2009), can be formulated as

$$\partial \| \mathbf{X} \|_* = \left\{ \mathbf{U_X} \mathbf{V_X}^T + \mathbf{Z} \middle| \mathbf{Z} \in \mathbb{R}^{d_1 \times d_2}, \mathbf{U_X}^T \mathbf{Z} = \mathbf{0}, \right.$$

$$\left. \mathbf{Z} \mathbf{V_X} = 0, \| \mathbf{Z} \|_2 \leq 1 \right\}, \qquad (2)$$

where $\| \mathbf{Z} \|_2 = \sigma_1(\mathbf{Z})$ denotes the spectral norm of $\mathbf{Z}$.

SMM

The matrix classifier, SMM, is defined as a penalty function plus a hinge loss. The penalty function, i.e., spectral elastic net, which

enjoys the property of grouping effect as well as keeping a low-rank representation. The hinge loss enjoys the property of large margin while contributing to the sparseness and robustness of the classifier. The objective function of the SMM can be formulated as follows:

$$\min_{\mathbf{w},b} \frac{1}{2} tr\left(\mathbf{W}^T \mathbf{W}\right) + \tau \parallel \mathbf{W} \parallel_* + C \sum_{i=1}^{N} \xi_i$$
$$\text{s.t. } y_i \left[tr\left(\mathbf{W}^T \mathbf{X}_i\right) + b \right] \geq 1 - \xi_i, \xi_i \geq 0, \; \forall i = 1, 2, \ldots, N$$
(3)

Specifically, the spectral elastic net can be represented as a combination of the squared Frobenius matrix norm $\parallel \mathbf{W} \parallel_F^2 = tr\left(\mathbf{W}^T \mathbf{W}\right)$ and nuclear norm $\parallel \mathbf{W} \parallel_*$ on the regression matrix W.

The objective function of the SMM can be solved through the ADMM method; thus, Eq. (3) is reformulated as

$$\min_{\mathbf{W},b,\mathbf{S}} F\left(\mathbf{W}, b\right) + G\left(\mathbf{S}\right)$$
$$\text{s.t. } \mathbf{S} - \mathbf{W} = \mathbf{0}$$
(4)

where $F\left(\mathbf{W}, b\right) = \frac{1}{2} tr\left(\mathbf{W}^T \mathbf{W}\right) + C \sum_{i=1}^{N} \{1 - y_i[tr(\mathbf{W}^T \mathbf{X}_i) + b]\}_+$ and $G\left(\mathbf{S}\right) = \tau \parallel \mathbf{S} \parallel_*$.

According to the augmented Lagrangian function in Eq. (4), $\left(\mathbf{W}, b\right)$ and $\mathbf{S}$ can be iteratively computed in two steps:

$$(\mathbf{W}^{(k)}, b^{(k)}) = \operatorname*{argmin}_{(\mathbf{W},b)} F(\mathbf{W}, b) - tr((\mathbf{\Lambda}^{(k-1)})^T \mathbf{W})$$
$$+ \frac{\rho}{2} \parallel \mathbf{W} - \mathbf{S}^{(k-1)} \parallel_F^2,$$
(5)

$$\mathbf{S}^{(k)} = \operatorname*{arg\,min}_{\mathbf{S}} G(\mathbf{S}) + tr((\mathbf{\Lambda}^{(k-1)})^T \mathbf{S}) + \frac{\rho}{2} \parallel \mathbf{W}^{(k)} - \mathbf{S} \parallel_F^2, \quad (6)$$

where k denotes the iteration index. $\rho > 0$ is a hyperparameter, and $\mathbf{\Lambda}$ is a Lagrangian multiplier.

Knowledge-Leverage-Based SMM

Generally, the current SMM and its variants belong to the data-driven method that always focuses on achieving impressive classification performance with sufficient training data. In practice, it is necessary to collect sufficient EEG data for each subject to establish a subject-specific classifier. However, long-term recording EEG data may exhaust the subject. Therefore, to model the target domain using insufficient EEG data, we proposed a novel algorithm to enhance the generalization capability of the SMM on the target domain by leveraging the useful knowledge underlying the source domain.

The framework of the KL-SMM for an EEG-based MI BCI is illustrated in **Figure 1**. To model the target domain, two main types of information, the model knowledge of the source domain and few labeled target EEG data, are used simultaneously.

KL-SMM Model

A dataset $D_s = \left\{\left(\mathbf{X}_{s,1}, y_{s,1}\right), \left(\mathbf{X}_{s,2}, y_{s,2}\right), \cdots, \left(\mathbf{X}_{s,N_s}, y_{s,N_s}\right)\right\}$ in source domain, it consists of N_s trials labeled EEG signals. $\mathbf{X}_{s,i} \in \mathbb{R}^{p \times q}$ denotes the ith trial with $p \times q$ dimension. $y_{s,i}$ is the corresponding class label of $\mathbf{X}_{s,i}$. A dataset $D_t = \left\{\left(\mathbf{X}_{t,1}, y_{t,1}\right), \cdots, \left(\mathbf{X}_{t,N}, y_{t,N}\right), \mathbf{X}_{t,N+1}, \cdots, \mathbf{X}_{t,N_t}\right\}$ in the target domain, it consists of N labeled EEG trials and $(N_t - N)$ unlabeled trials, where $N \ll (N_t - N)$.

For modeling the target domain, we proposed to integrate the labeled target EEG data and source model as follows:

$$\min_{f} \sum_{i=1}^{N} L\left(f\left(\mathbf{X}_{t,i}\right), y_{t,i}\right) + \lambda tr\left[\left(\mathbf{W}_s - \mathbf{W}_t\right)^T \left(\mathbf{W}_s - \mathbf{W}_t\right)\right], \quad (7)$$

Here, $L\left(\cdot, \cdot\right)$ denotes the loss function. $f\left(\mathbf{X}_{t,i}\right) = tr\left(\mathbf{W}_t^T \mathbf{X}_{t,i}\right) + b$ denotes the matrix classifier to be learned. Eq. (7) includes two terms, where the first term is used to learn from labeled target EEG data, and the second term is designed to leverage the model knowledge (i.e., $\mathbf{W}_s$) underlying the source domain. The goal is to exploit the desired KL-SMM by approximating its model to the source model. The parameter λ is adopted to balance the influence between the two terms above.

As in the SMM, we introduced the spectral elastic net penalty to capture the correlation information within the matrix-form
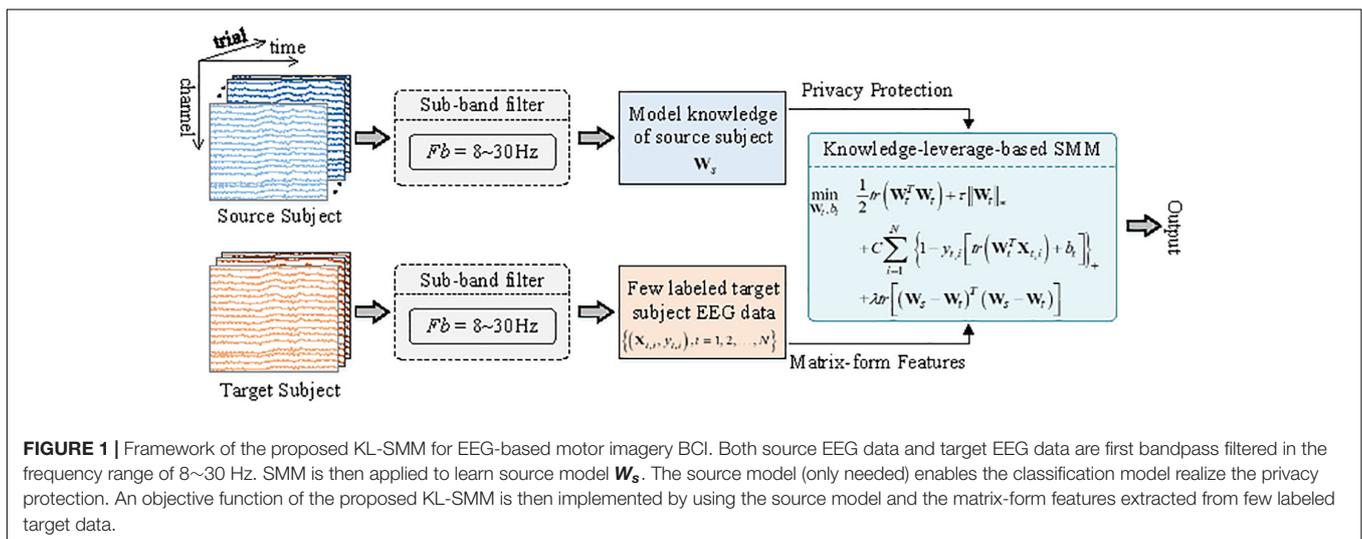
FIGURE 1 | Framework of the proposed KL-SMM for EEG-based motor imagery BCI. Both source EEG data and target EEG data are first bandpass filtered in the frequency range of 8~30 Hz. SMM is then applied to learn source model $\boldsymbol{W_s}$. The source model (only needed) enables the classification model realize the privacy protection. An objective function of the proposed KL-SMM is then implemented by using the source model and the matrix-form features extracted from few labeled target data.

EEG data. Furthermore, the hinge loss function was adopted, owing to its inherent characteristic of sparseness and robustness. Above all, the objective function of the proposed KL-SMM can be formulated as follows:

$$\min_{\mathbf{W}_t, b_t} \frac{1}{2} tr(\mathbf{W}_t^T \mathbf{W}_t) + \tau \parallel \mathbf{W}_t \parallel_* + C \sum_{i=1}^{N} \xi_i^t$$
$$+ \lambda tr[(\mathbf{W}_s - \mathbf{W}_t)^T (\mathbf{W}_s - \mathbf{W}_t)] \quad , \quad (8)$$
$$\text{s.t.} \quad y_{t,i}[tr(\mathbf{W}_t^T \mathbf{X}_{t,i}) + b_t] \geq 1 - \xi_i^t, \xi_i^t \geq 0,$$
$$\forall i = 1, 2, \ldots, N$$

where the parameter $C > 0$ is used to maintain a balance between fitting the labeled target EEG data and minimizing the complexity of the solution.

Parameter Learning for KL-SMM

Because the Eq. (8) is convex in both $\mathbf{W}_t$ and b_t, an alternating iterative strategy based on the ADMM method can be used to derive the learning algorithm of the KL-SMM. Specifically, by introducing an auxiliary variable S_t, the objective function of the KL-SMM can be equivalently reformulated as

$$\min_{\mathbf{W}_t, b_t, \mathbf{S}_t} F(\mathbf{W}_t, b_t) + G(\mathbf{S}_t) \quad , \quad (9)$$
$$\text{s.t.} \quad \mathbf{S}_t - \mathbf{W}_t = \mathbf{0}$$

where $F(\mathbf{W}_t, b_t) = \frac{1}{2} tr(\mathbf{W}_t^T \mathbf{W}_t) + C \sum_{i=1}^{N} \{1 - y_{t,i}[tr(\mathbf{W}_t^T \mathbf{X}_{t,i}) + b_t]\}_+ + \lambda tr[(\mathbf{W}_s - \mathbf{W}_t)^T (\mathbf{W}_s - \mathbf{W}_t)]$ and $G(\mathbf{S}_t) = \tau \parallel \mathbf{S}_t \parallel_*$.

The parameter optimization of Eq. (9) can be solved using the augmented Lagrangian algorithm

$$L_a(\mathbf{W}_t, b_t, \mathbf{S}_t, \mathbf{\Lambda}) = F(\mathbf{W}_t, b_t) + G(\mathbf{S}_t) + tr[\mathbf{\Lambda}^T(\mathbf{S}_t - \mathbf{W}_t)]$$
$$+ \frac{\rho}{2} \parallel \mathbf{S}_t - \mathbf{W}_t \parallel_F^2, \quad (10)$$

where $\mathbf{\Lambda}$ is the Lagrangian multiplier, and $\rho > 0$ is a hyperparameter. Theorems 1 and 2 provide the calculations of parameters $\mathbf{S}_t$ and $(\mathbf{W}_t, b_t)$.

Theorem 1. For the fixed $\mathbf{W}_t$, using the Lagrangian multiplier $\mathbf{\Lambda}$ and any positive scalar $\tau > 0$, $\rho > 0$ in Eq. (10), $\mathbf{S}_t$ can be optimized using the following update rule:

$$\mathbf{S}_t = D_\tau [\rho \mathbf{W}_t - \mathbf{\Lambda}]. \quad (11)$$

Proof of Theorem 1: Supposing $\mathbf{W}_t$ is fixed, the optimization problem in Eq. (10) is equivalent to minimizing the function as follows:

$$J(\mathbf{S}_t) = \tau \parallel \mathbf{S}_t \parallel_* + tr(\mathbf{\Lambda}^T \mathbf{S}_t) + \frac{\rho}{2} \parallel \mathbf{S}_t - \mathbf{W}_t \parallel_F^2. \quad (12)$$

Because $J(\mathbf{S}_t)$ is convex, with respect to $\mathbf{S}_t$, if $0 \in \partial J(\mathbf{S}_t^*)$ with $\mathbf{S}_t^* = (1/\rho) \cdot D_\tau [\rho \mathbf{W}_t - \mathbf{\Lambda}]$ can be proven, we can conclude that $\mathbf{S}_t^*$ is a solution to Eq. (12). The derivation of $J(\mathbf{S}_t)$ with respect to $\mathbf{S}_t$ can be expressed as

$$\partial J(\mathbf{S}_t) = \mathbf{\Lambda} - \rho \mathbf{W}_t + \rho \mathbf{S}_t + \tau \cdot \partial \parallel \mathbf{S}_t \parallel_*. \quad (13)$$

To further simplify this equation, let the SVD of $(\rho \mathbf{W}_t - \mathbf{\Lambda})$ be denoted as $\rho \mathbf{W}_t - \mathbf{\Lambda} = \mathbf{U}_a \sum_a \mathbf{V}_a^T + \mathbf{U}_b \sum_b \mathbf{V}_b^T$. In the equation,

$\sum_a$ represents the diagonal matrix with diagonal entries greater than τ. $\sum_b$ represents the remaining part of the SVD with diagonal entries less than or equal to τ. $\mathbf{U}_a$ and $\mathbf{V}_a$ ($\mathbf{U}_b$ and $\mathbf{V}_b$) are matrices that correspond to the left and right singular vectors of the diagonal matrix $\sum_a (\sum_b)$. In terms of Definition 1, $\mathbf{S}_t^*$ can be reformulated as $(1/\rho) \cdot \mathbf{U}_a (\sum_a - \tau \mathbf{I}) \mathbf{V}_a^T$. Substituting $\mathbf{S}_t^*$ into Eq. (13), we have

$$\partial J(\mathbf{S}_t)|_{\mathbf{S}_t^*} = \mathbf{\Lambda} - \rho \mathbf{W}_t + \rho \mathbf{S}_t^* + \tau \cdot \partial \parallel \mathbf{S}_t^* \parallel_*$$
$$= -\tau \left(\mathbf{U}_a \mathbf{V}_a^T + \frac{1}{\tau} \mathbf{U}_b \sum_b \mathbf{V}_b^T \right) + \tau \cdot \partial \parallel \mathbf{S}_t^* \parallel_*. \quad (14)$$

Let $\mathbf{Z} = (1/\tau) \cdot \mathbf{U}_b \sum_b \mathbf{V}_b^T$, because $\mathbf{U}_a$, $\mathbf{U}_b$, $\mathbf{V}_a$, $\mathbf{V}_b$ are column orthogonal, we can easily verify that $\mathbf{U}_a^T \mathbf{Z} = 0$, $\mathbf{Z} \mathbf{V}_a = 0$, and $\parallel \mathbf{Z} \parallel_2 \leq 1$. Thus, we have $0 \in \partial J(\mathbf{S}_t^*)$. Theorem 1 is proved.

Theorem 2. For the fixed $\mathbf{S}_t$, $(\mathbf{W}_t, b_t)$ can be optimized using the following update rule:

$$\mathbf{W}_t = \frac{1}{2\lambda + \rho + 1} \left(\mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s + \sum_{i=1}^{N_t} \alpha_i y_{t,i} \mathbf{X}_{t,i} \right), \quad (15)$$

$$b_t = \frac{1}{|\Delta|} \sum_{i \in \Delta} \left\{ y_{t,i} - [tr(\mathbf{W}_t^T \mathbf{X}_{t,i})] \right\}, \quad (16)$$

where $\Delta = \{i | 0 \leq \alpha_i \leq C, \forall i = 1, 2, \ldots, N\}$ refers to the Lagrangian multipliers, and $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T \in \mathbb{R}^N$ can be obtained using the box constraint quadratic programming solver:

$$\max_{\alpha} -\frac{1}{2} \alpha^T \mathbf{K} \alpha + \mathbf{H}^T \alpha \quad , \quad (17)$$
$$\text{s.t.} 0 \leq \alpha \leq C, \alpha^T \mathbf{Y} = 0$$

where $\mathbf{K} = [K_{ij}] \in \mathbb{R}^{N \times N}$ and $\mathbf{H} = [h_i] \in \mathbb{R}^N$ with

$$K_{ij} = \frac{1}{2\lambda + \rho + 1} y_{t,i} y_{t,j} tr(\mathbf{X}_{t,i}^T \mathbf{X}_{t,j}), \quad (18)$$

$$h_i = 1 - \frac{1}{2\lambda + \rho + 1} y_{t,i} tr[(\mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s)^T \mathbf{X}_{t,i}]. \quad (19)$$

Proof of Theorem 2: Given the fixed variable $\mathbf{S}_t$, the optimization problem in Eq. (10) equals to optimize the following objective function:

$$\min_{\mathbf{W}_t, b_t} \frac{1}{2} tr(\mathbf{W}_t^T \mathbf{W}_t) + C \sum_{i=1}^{N} \xi_i^t - tr(\mathbf{\Lambda}^T \mathbf{W}_t) + \frac{\rho}{2} \parallel \mathbf{S}_t - \mathbf{W}_t \parallel_F^2$$
$$+ \lambda tr[(\mathbf{W}_s - \mathbf{W}_t)^T (\mathbf{W}_s - \mathbf{W}_t)] \quad .$$
$$\text{s.t.} \quad y_{t,i}[tr(\mathbf{W}_t^T \mathbf{X}_{t,i}) + b_t] \geq 1 - \xi_i^t, \; \xi_i^t \geq 0, \forall i = 1, 2, \ldots, N \quad (20)$$

The augmented Lagrangian function of Eq. (20) is denoted as

$$L_a' = \frac{1}{2} tr(\mathbf{W}_t^T \mathbf{W}_t) + C \sum_{i=1}^{N} \xi_i^t - tr(\mathbf{\Lambda}^T \mathbf{W}_t) + \frac{\rho}{2} \parallel \mathbf{S}_t - \mathbf{W}_t \parallel_F^2$$
$$+ \lambda tr[(\mathbf{W}_s - \mathbf{W}_t)^T (\mathbf{W}_s - \mathbf{W}_t)] \quad .$$
$$- \sum_{i=1}^{N} \alpha_i \{y_{t,i}[tr(\mathbf{W}_t^T \mathbf{X}_{t,i}) + b_t] - 1 + \xi_i^t\} - \sum_{i=1}^{N} \beta_i \xi_i^t \quad (21)$$

Setting the derivative of L'_a with respect to ξ^t_i and b_t, to 0, we can obtain

$$\frac{\partial L'_a}{\partial \xi^t_i} = 0 \Rightarrow C - \alpha_i - \beta_i = 0 \text{ and } \frac{\partial L'_a}{\partial b_t} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_{t,i} = 0. \tag{22}$$

Substituting Eq. (22) into Eq. (21), and then setting the derivative of L'_a with respect to W_t to 0, we obtain

$$\frac{\partial L'_a}{\partial \mathbf{W}_t} = 0 \Rightarrow \mathbf{W}_t$$

$$= \frac{1}{2\lambda + \rho + 1} \left(\mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s + \sum_{i=1}^{N} \alpha_i y_{t,i} \mathbf{X}_{t,i} \right). \tag{23}$$

Substituting Eq. (22) and Eq. (23) into Eq. (21),

$$L'_a = \sum_{i=1}^{N} \left(1 - \frac{1}{2\lambda + \rho + 1} y_{t,i} tr \left[(\mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s)^T \mathbf{X}_i \right] \right) \alpha_i$$

$$- \frac{1}{2(2\lambda + \rho + 1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_{t,i} y_{t,j} tr(\mathbf{X}^T_{t,i} \mathbf{X}_{t,j}) + \theta. \tag{24}$$

Here, θ is a constant, which can be represented as $\theta = \frac{\rho}{2} tr \left(\mathbf{S}^T_t \mathbf{S}_t \right) + \lambda tr \left(\mathbf{W}^T_s \mathbf{W}_s \right) - \frac{1}{2(2\lambda + \rho + 1)} \| \mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s \|^2_F$. Thus, the dual problem of Eq. (24) can be denoted as

$$\max_{\alpha} \sum_{i=1}^{N} \left(1 - \frac{1}{2\lambda + \rho + 1} y_{t,i} tr \left[(\mathbf{\Lambda} + \rho \mathbf{S}_t + 2\lambda \mathbf{W}_s)^T \mathbf{X}_i \right] \right) \alpha_i$$

$$- \frac{1}{2(2\lambda + \rho + 1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_{t,i} y_{t,j} tr(\mathbf{X}^T_{t,i} \mathbf{X}_{t,j})$$

$$\text{s.t. } \sum_{i=1}^{N} \alpha_i y_{t,i} = 0, 0 \le \alpha_i \le C, \forall i = 1, 2, \ldots, N \tag{25}$$

Algorithm 1: The learning procedure for KL-SMM

Input: Training dataset $D_T = \{\mathbf{X}_{t,i}, y_{t,i}\}_{i=1}^{N}$, source model $\mathbf{W}_s$, parameter τ and λ;

Output: $\mathbf{W}_t, b_t$;

Initialize: $\mathbf{S}_t^{(-1)} = \hat{\mathbf{S}}_t^{(0)} \in \mathbf{R}^{p \times q}$, $\mathbf{\Lambda}^{(-1)} = \hat{\mathbf{\Lambda}}^{(0)} \in \mathbf{R}^{p \times q}$, $\nu^{(1)} = 1$, $\eta \in (0, 1), \rho > 0, l = 1$

   Repeat

1 | Computing $\mathbf{W}_t^{(l)}, b_t^{(l)}$ with $\hat{\mathbf{S}}_t^{(l)}$ and $\hat{\mathbf{\Lambda}}^{(l)}$ using Theorem 2;

2 | Computing $\mathbf{S}_t^{(l)}$ with $\mathbf{W}_t^{(l)}$ and $\hat{\mathbf{\Lambda}}^{(l)}$ using Theorem 1;

3 | Update $\mathbf{\Lambda}^{(l)} \leftarrow \hat{\mathbf{\Lambda}}^{(l)} + \rho \left(\mathbf{S}_t^{(l)} - \mathbf{W}_t^{(l)} \right)$;

4 | $l \leftarrow l + 1$.

   Until convergence

The optimization problem of Eq. (21) can finally be transformed into a QP problem. Substituting the obtained optimal solution α into Eq. (23), it is easy to obtain the value of $\mathbf{W}_t$. Finally, the optimal b_t can be calculated as follows:

$$b_t = y_{t,i} - \left[tr \left(\mathbf{W}_t^T \mathbf{X}_{t,i} \right) \right]. \tag{26}$$

In practice, averaging these optimal solutions, we can obtain

$$b_t = \frac{1}{|\Delta|} \sum_{i \in \Delta} \left\{ y_{t,i} - \left[tr \left((\mathbf{W}_t)^T \mathbf{X}_{t,i} \right) \right] \right\}, \tag{27}$$

where $\Delta = \{i \,|\, 0 \le \alpha_i \le C, \forall i = 1, 2, \ldots, N\}$. Theorem 2 is proved.

For the fixed $\mathbf{W}_t$ and $\mathbf{S}_t$, the Lagrangian multiplier $\mathbf{\Lambda}$ in Eq. (10) can be updated as follows:

$$\mathbf{\Lambda}^{(k)} = \mathbf{\Lambda}^{(k-1)} + \rho \left(\mathbf{S}_t - \mathbf{W}_t \right). \tag{28}$$

The optimal solution is estimated iteratively. The learning procedure for the KL-SMM is given in Algorithm 1.

Computational Complexity

We further analyzed the computational complexity of the KL-SMM. In Algorithm 1, Step 1 computes the parameter $(\mathbf{W}_t, b_t)$ by solving a QP problem in Eq. (17), which takes time $O(N^2 pq)$ with N samples of $p \times q$ dimension. Step 2 computes the eigen decomposition for $\mathbf{S}_t$ in Eq. (11), which takes time $O(\min(p^2 q, pq^2))$. In practice, the dimensions, p and q, of the extracted EEG features are not too high. Thus, the computational complexity of the KL-SMM is dominated by the QP, that is, $O(I \cdot N^2 pq)$, where I denotes the iteration number.

EXPERIMENTS

In this section, we evaluate the proposed KL-SMM on two publicly available MI EEG datasets [i.e., Datasets IIa and IIb of the BCI competition (Hang et al., 2020)], which can be found in http://www.bbci.de/competition/iv/. We first describe the EEG datasets. Then, the compared methods and corresponding parameter settings are provided. Finally, we present and discuss the experimental results.

EEG Data Description and Preprocessing

(1) *BCI competition IV Dataset IIa (Exp.1)*: This dataset includes 22-channel EEG signals recorded from nine subjects (denoted as *S01–S09*). During the experiment, each subject was required to perform four kinds of MI tasks, hand (left and right), foot, and tongue. A total of 576 trials of two sessions on different days were collected for each subject. In our experiment, we used the left-hand and right-hand EEG data. In addition, the EEG trials collected from the second day were adopted. Thus, the training and test datasets each contained 72 EEG trials.

(2) *BCI competition IV Dataset IIb (Exp.2)*: This dataset also contains the EEG signals of nine subjects (denoted as *B01–B09*), which were recorded using three electrodes, *C3, Cz,* and *C4*. During the experiment, each subject was instructed to perform left- and right-hand MI tasks for 4.5 s. For each subject, there were five sessions. Sessions 1, 2, and 3 were collected on the first day, and 4 and 5 were collected on the second day. Similar to *Exp.1*, the EEG trials collected from the second day were used. Specifically, Session 4 was used as the training data, and Session 5 was used as the test data.

With reference to Hang et al. (2020), there was a time interval of *[0.5, 3]*s after the visual cues in each trial for all the datasets. We bandpass-filtered the EEG signals to 8–30 Hz through a five-order Butterworth filter, which covers the dominated frequency band for MI tasks (Nam et al., 2011). Then, we adopted the spatial filters to detect the MI-related desynchronization/synchronization (ERD/ERS) patterns. Finally, the widely used band-power estimation method (Vidaurre et al., 2005) was used to extract the matrix-form EEG features for all the subjects. To construct the transfer learning tasks, each subject was considered the target domain, and the training data of the remaining subjects constituted the source domain. To evaluate the performance of the KL-SMM, we set three different numbers of labeled target EEG data, that is, the first 8, 14, and 20 training trials. The classification performances on the test data of all the subjects were reported.

Experimental Setup
Baseline Methods
To evaluate the classification performance of the aforementioned transfer tasks, the proposed method was compared with four methods in the experiments: (1) SVM (Vapnik, 1995), (2) BSVM (Pirsiavash et al., 2009), (3) SMM (Luo et al., 2015), and (4) Adaptive SVM (ASVM) (Yang et al., 2007).

Implementation Details
It is known that the format of the input data of both the SVM and ASVM should be vectors or scalars. Thus, we first had to reshape the extracted two-dimensional matrix features into vector features. For the BSVM, SMM, and proposed KL-SMM, the matrix features can be inputted directly. To evaluate the effect of the transfer learning mechanism, because the SVM, BSVM, and SMM are no-transfer baselines, we simply used the labeled target EEG trials as the training data to build these classification models. In addition, for the ASVM and KL-SMM, we also leveraged the source model knowledge in constructing the target classifier. However, unlike the ASVM, the traditional transfer learning method, the KL-SMM can directly process EEG matrix features and fully exploit the structural information.

The optimal parameters were selected using a five-fold cross-validation method on the training group for all comparison methods. Parameter C was assigned by selecting the value from the set $\{1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e0, 1e1\}$. For the SMM and KL-SMM, we tuned the parameter τ from the set $\{1e-5, 2e-5, 5e-5, 1e-4, 2e-4, 5e-4, 1e-3,$ $2e-3, 5e-3, 1e-2, 2e-2, 5e-2, 1e-1, 2e-1, 5e-1, 1e0\}$. To ensure a fair comparison with the ASVM, an adjustable parameter λ_A was added to the term of knowledge transfer, which can control how much the knowledge of source domain to transfer. For the KL-SMM and ASVM, parameters λ and λ_A were set by selecting the value from set $\{1e-4, 5e-4, 1e-3,$ $5e-3, 1e-2, 5e-2, 1e-1, 5e-1, 1e0\}$. To validate the classification performance of our method, we used the following metrics on the test dataset, i.e., Accuracy (ACC), $F1$ score (F1), and the area under the receiver operating characteristics curve (AUC). Herein, $\text{ACC} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})$ and $\text{F1} = 2 \times \text{PPV} \times \text{SEN} / (\text{PPV} + \text{SEN})$, where the positive predictive value $(\text{PPV}) = \text{TP} / (\text{TP} + \text{FP})$ and sensitivity $(\text{SEN}) = \text{TP} / (\text{TP} + \text{FN})$.

Experimental Results Analysis
The classification performances of all the comparison methods on 14 labeled target EEG trials on two datasets are given in **Tables 1**–**6**. The performance comparison of the KL-SMM with the compared methods in *Exp.1* is shown in **Tables 1**–**3**. The classification results of all compared methods in *Exp. 2* are shown in **Tables 4**–**6**. The best classification performance values are highlighted in bold. According to the results, the following conclusions can be drawn.

From the classification performances of all the comparison methods on the 14 labeled target EEG trials, we found that the proposed KL-SMM method achieved the highest average results in terms of the ACC, AUC, and F1. As shown in **Tables 1**–**3**, the proposed KL-SMM outperformed the baseline SMM on average by 5.87%, 5.66%, and 8.85% based on the ACC, AUC, and F1, respectively. As can be observed from the classification results in **Tables 4**–**6**, the KL-SMM outperformed the SMM on average by 3.68%, 3.78%, and 5.83% based on the ACC, AUC, and F1, respectively. The promising performances prove that the KL-SMM can leverage the model knowledge of the source subject to boost the generalization capability of the SMM when there are limited labeled EEG trials. In addition, the KL-SMM boosted the classification accuracy for six out of nine subjects in Exp.1, and eight out of nine subjects in Exp.2, respectively. These experimental results further demonstrate the effectiveness of the proposed KL-SMM that leveraged the knowledge underlying the source domain.

The BSVM and SMM outperformed the SVM in most cases. This confirms the ability of the BSVM and SMM to exploit the correlations between rows or columns of EEG matrix features.

TABLE 1 | Comparison of ACC using 14 labeled target EEG data in *Exp.1*.

	Target subjects									
Methods	S01	S02	S03	S04	S05	S06	S07	S08	S09	Avg.
SVM	0.5972	0.5000	0.9583	0.5556	0.5139	0.5556	0.5972	0.9306	0.8472	0.6728
BSVM	0.5556	0.4861	0.9583	0.5278	0.5972	0.5417	0.5972	**1.0000**	0.8472	0.6790
SMM	0.6111	0.5000	**0.9722**	0.5556	0.5556	0.5556	0.6111	0.9444	0.8472	0.6836
ASVM	0.6111	**0.5139**	0.9167	0.6667	0.6389	0.5972	0.7222	0.9444	0.8472	0.7176
KL-SMM	**0.6528**	0.5000	0.9306	**0.7361**	**0.6944**	**0.6111**	**0.7361**	0.9583	**0.8611**	**0.7423**

TABLE 2 | Comparison of AUC using 14 labeled target EEG data in *Exp.1*.

Methods	Target subjects									
	S01	S02	S03	S04	S05	S06	S07	S08	S09	Avg.
SVM	0.6265	0.5015	0.9776	0.5640	0.5316	0.5548	0.6019	0.9228	0.9190	0.6889
BSVM	0.5980	0.5023	0.9367	0.5239	0.6019	0.5579	0.6181	**1.0000**	0.9190	0.6953
SMM	0.6273	0.5015	**0.9830**	0.5640	0.5733	0.5548	0.6142	0.9406	0.9190	0.6975
ASVM	0.6273	**0.5177**	0.9313	0.6790	0.6481	0.6142	0.7261	0.9406	0.9190	0.7337
KL-SMM	**0.6744**	0.5015	0.9375	**0.7261**	**0.7068**	**0.6173**	**0.7415**	0.9606	**0.9213**	**0.7541**

TABLE 3 | Comparison of F1 using 14 labeled target EEG data in *Exp. 1*.

Methods	Target subjects									
	S01	S02	S03	S04	S05	S06	S07	S08	S09	Avg.
SVM	0.3256	0.5385	0.9589	0.4483	0.3636	0.4667	0.6234	0.9275	0.8308	0.6092
BSVM	0.2000	0.5195	0.9565	0.3462	0.5538	0.4407	0.5397	**1.0000**	0.8308	0.5986
SMM	0.3636	0.5385	**0.9722**	0.4483	0.4286	0.4667	0.6410	0.9429	0.8308	0.6258
ASVM	0.3636	**0.5570**	0.9091	0.6842	0.6579	**0.5672**	0.6970	0.9429	0.8308	0.6900
KL-SMM	**0.4681**	0.5385	0.9254	**0.7397**	**0.7027**	0.5333	**0.7164**	0.9565	**0.8485**	**0.7143**

TABLE 4 | Comparison of ACC using 14 labeled target EEG data in *Exp. 2*.

Methods	Target subjects									
	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg.
SVM	0.4750	0.5125	0.4500	0.8688	0.5813	0.6688	0.7000	0.9500	0.7250	0.6590
BSVM	0.5125	0.5125	0.4813	0.8750	0.5688	0.5813	0.6250	0.9375	0.7188	0.6458
SMM	0.4875	**0.5313**	0.4500	0.8750	0.5938	0.6750	0.6813	0.9563	0.7625	0.6681
ASVM	0.6063	0.5000	**0.5000**	0.9125	0.6063	0.6938	0.6813	0.9500	0.6688	0.6799
KL-SMM	**0.6125**	**0.5313**	0.4625	**0.9563**	**0.6438**	**0.7125**	**0.7250**	**0.9625**	0.7375	**0.7049**

TABLE 5 | Comparison of AUC using 14 labeled target EEG data in *Exp. 2*.

Methods	Target subjects									
	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg.
SVM	0.4614	0.5247	0.4655	0.8698	0.5948	0.6761	0.6980	0.9577	0.7263	0.6638
BSVM	0.4963	0.5102	0.5048	0.8652	0.5781	0.5833	0.6159	0.9453	0.7080	0.6452
SMM	0.4747	0.5431	0.4655	0.8814	0.6081	0.6825	0.6841	0.9605	**0.7678**	0.6742
ASVM	0.5922	0.5103	**0.5356**	0.9377	0.6172	0.7138	0.6664	0.9577	0.6580	0.6876
KL-SMM	**0.5939**	**0.5470**	0.4697	**0.9673**	**0.6627**	**0.7389**	**0.7202**	**0.9670**	0.7416	**0.7120**

TABLE 6 | Comparison of F1 using 14 labeled target EEG data in *Exp. 2*.

Methods	Target subjects									
	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg.
SVM	0.5172	0.4348	0.3125	0.8800	0.6082	0.6788	0.7303	0.9518	0.7755	0.6544
BSVM	0.6100	**0.5938**	0.3566	0.8837	0.5767	0.5839	0.7170	0.9390	0.7368	0.6664
SMM	0.5119	0.4604	0.3125	0.8851	0.6061	0.6829	0.7052	0.9576	**0.7935**	0.6572
ASVM	0.6358	0.3651	**0.5918**	0.9176	0.6557	0.7322	0.7243	0.9518	0.7039	0.6976
KL-SMM	**0.6702**	0.4000	0.4691	**0.9571**	**0.6885**	**0.7527**	**0.7528**	**0.9634**	0.7857	**0.7155**

Notably, the SMM has better classification performance than the BSVM because of its convex objective function that can be effectively optimized by the ADMM method. Furthermore, it can be observed that, owing to its ability to leverage the source model knowledge in modeling the target domain when the labeled target EEG data is very limited, the ASVM yielded
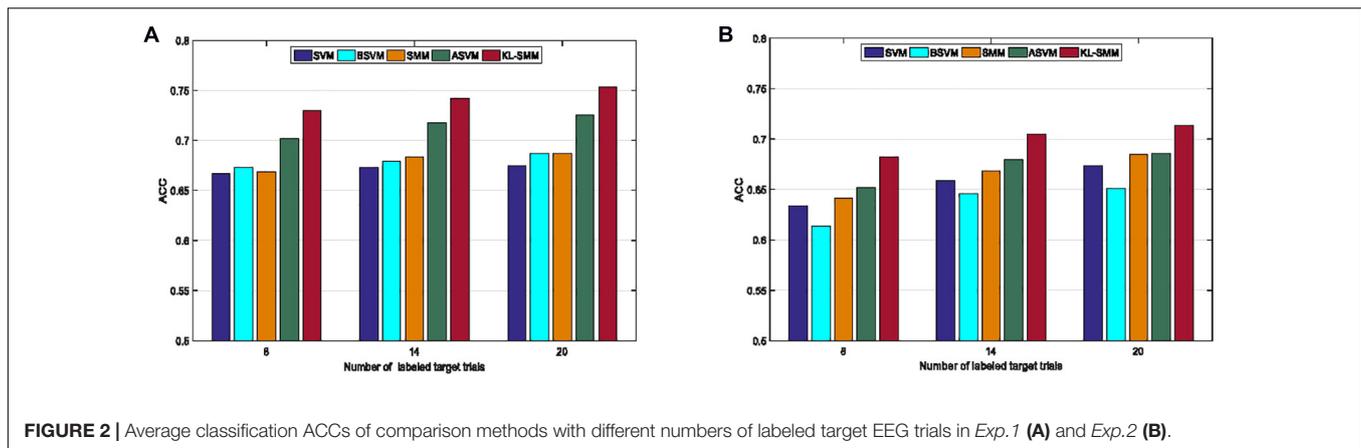
FIGURE 2 | Average classification ACCs of comparison methods with different numbers of labeled target EEG trials in *Exp.1* **(A)** and *Exp.2* **(B)**.

TABLE 7 | Statistical significance comparisons of ACC of KL-SMM and other methods in *Exp.1* and *Exp.2*.

Num. of labeled trials	Exp.1				Exp.2			
	KL-SMM vs. SVM	KL-SMM vs. BSVM	KL-SMM vs. SMM	KL-SMM vs. ASVM	KL-SMM vs. SVM	KL-SMM vs. BSVM	KL-SMM vs. SMM	KL-SMM vs. ASVM
8	**0.0167**	**0.0116**	**0.0085**	**0.0085**	**0.0164**	**0.0107**	**0.0465**	**0.0307**
14	**0.0287**	**0.0485**	**0.0442**	**0.0207**	**0.0129**	**0.0074**	**0.0419**	**0.0376**
20	**0.0039**	**0.0227**	**0.0060**	**0.0249**	**0.0014**	**0.0053**	**0.0215**	**0.0328**
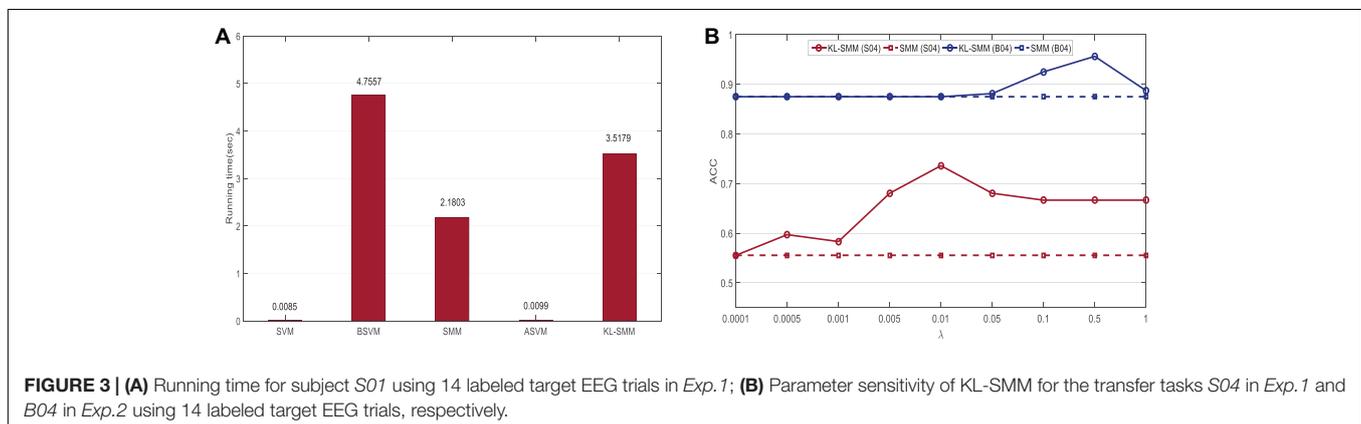
FIGURE 3 | (A) Running time for subject *S01* using 14 labeled target EEG trials in *Exp.1*; **(B)** Parameter sensitivity of KL-SMM for the transfer tasks *S04* in *Exp.1* and *B04* in *Exp.2* using 14 labeled target EEG trials, respectively.

better classification results than the SVM. The foundation of the KL-SMM is the SMM, which can leverage the source model knowledge and exploit the structural information within the EEG feature matrices. The experimental results prove that structural information can indeed improve classification performance.

We further studied the effects of different numbers of labeled target EEG instances on the classification performance of the KL-SMM. **Figure 2** shows the average classification ACCs when 8, 14, and 20 labeled target EEG trials were available from the target subject. **Figures 2A,B** show the average classification results of all the compared methods for *Exp.1* and *Exp.2*, respectively. It can be observed that the KL-SMM outperformed the other methods in all the cases. Specifically, the improvement was more pronounced when few labeled target EEG trials were available, as shown in **Figure 2B**. From these results, we can observe that increasing the number of the labeled target EEG instances

improved the average classification ACCs of all the compared methods. This is mainly because more training data may enhance the generalization performance of the classification model. The average ACC of the KL-SMM was significantly better than those of the other methods when there was no knowledge transfer, especially when the labeled target EEG instances were very limited. Overall, compared to other methods, the classification performance of the proposed KL-SMM was superior. The encouraging results were mainly attributed to the fact that the KL-SMM method possessed the matrix learning capability derived from the matrix learning machine, which allowed it to directly handle the matrix-form features, thus retaining the structural information of EEG data. In addition, the KL-SMM achieved a more outstanding classification performance because of its ability to leverage the useful model knowledge of the source domain.

DISCUSSION

Statistical Analysis

We further performed a t-test statistical analysis to verify whether there was a significant difference with a confidence level of 95% between the KL-SMM and the other methods. The results of the t-test using different numbers of labeled target EEG trials are shown in **Table 3**. A p-value less than 0.05 indicates that there are significant differences between the KL-SMM and the other methods. We highlighted the statistically significant differences in boldface. From **Table 7**, it can be observed that in all cases, the null hypothesis can be rejected. This proves that the KL-SMM significantly outperformed the other methods. This further demonstrated the ability of the KL-SMM to capture the structural information within the EEG data, in addition to a strong transfer learning capability. Therefore, it is suitable for the classification of complex matrix-form EEG data with cross-subject variability.

Running Time

Figure 3A shows the running time of the KL-SMM and other methods on a subject, *S01*, in *Exp.1* using 14 labeled target EEG trials. Except for the SVM and ASVM, the KL-SMM achieved comparable computational cost with the traditional matrix leaning method SMM. Furthermore, the KL-SMM required less computational time, compared to the BSVM. It was proven that the running time of the KL-SMM was approximately 1.6 times less that of the SMM. The KL-SMM achieved better classification results, without the increase in computational costs becoming unacceptable. This shows the potential value of the KL-SMM for real-world BCI applications.

Parameter Sensitivity

We further show the effect of free parameter on the performance of KL-SMM, i.e., the knowledge transfer penalty λ. We conduct parameter sensitivity experiments on the transfer tasks *S04* in *Exp.1* and *B04* in *Exp.2* using 14 labeled target EEG trials, respectively. We vary the parameter of interest in $\{1e-4, 5e-4, 1e-3, 5e-3, 1e-2, 5e-2, 1e-1, 5e-1, 1e0\}$. **Figure 3B** shows the classification accuracy of our KL-SMM in contrast to SMM represented as dashed lines. It can be found that the accuracy of KL-SMM is improved with the increase of parameter λ, suggesting that taking the model knowledge of source domain into account can benefit for EEG classification. As the parameter value is further increased, the classification performance will decrease due to the distribution discrepancy between the source domain and the target domain.

CONCLUSION

In this study, we proposed a KL-SMM method for MI-based BCIs. The proposed KL-SMM belongs to the matrix classifier, which can exploit the structural information of EEG data in matrix form. Furthermore, it can leverage the existing source model knowledge in modeling the construction of the target subjects in scenarios of limited labeled target training data. Similar to the SMM, the KL-SMM can be easily optimized using the ADMM. Extensive experimental results on two publicly available MI datasets demonstrate the superiority of the KL-SMM to the compared methods in most cases. However, despite its promising performance, there is still room for improvement. For example, adaptively controlling the penalty λ is critical to determining how much knowledge is transferred. In addition, how to extend KL-SMM to multi-class classification will be investigated in future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.bbci.de/competition/iv/.

AUTHOR CONTRIBUTIONS

YC is responsible for data processing and data analysis. WH and SL are responsible for manuscript writing. XL and QW is responsible for study design. GL is responsible for experimental design. JQ and K-SC are responsible for manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Agarwal, A., Dowsley, R., McKinney, N. D., Wu, D., Lin, C. T., De Cock, M., et al. (2019). Protecting privacy of users in brain-computer interface applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1546–1555. doi: 10.1109/tnsre.2019. 2926965

Alamgir, M., Grosse-Wentrup, M., and Altun, Y. (2010). "Multitask learning for brain-computer interfaces," in *Processing of the 13th International Conference on Artificial Intelligence and Statistics. (AISTATS)*, Vol. 9, Washington, DC, 17–24.

Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., and Wolpaw, J. R. (2008). Towards an independent brain–computer interface using steady state visual evoked potentials. *Clin. Neurophysiol.* 119, 399–408. doi: 10.1016/j.clinph.2007.09.121

Al-Qaysi, Z. T., Zaidan, B. B., Zaidan, A. A., and Suzani, M. S. (2018). A review of disability EEG based wheelchair control system: coherent taxonomy, open challenges and recommendations. *Comput. Methods Progr. Biomed.* 164, 221–237. doi: 10.1016/j.cmpb.2018.06.012

Ang, K. K., Guan, C., Chua, K. S. G., Ang, B. T., Kuah, C. W. K., Wang, C., et al. (2011). A large clinical study on the ability of stroke patients to use an

EEG-based motor imagery brain-computer interface. *Clin. EEG Neurosci.* 42, 253–258. doi: 10.1177/155005941104200411

Azab, A. M., Mihaylova, L., Ang, K. K., and Arvaneh, M. (2019). Weighted transfer learning for improving motor imagery-based brain–computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1352–1359. doi: 10.1109/tnsre.2019.2923315

Cai, J. F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20, 1956–1982. doi: 10.1137/080738970

Candes, E. J., and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* 9, 717–772. doi: 10.1007/s10208-009-9045-5

Deng, Z., Jiang, Y., Choi, K.-S., Chung, F.-L., and Wang, S. (2013). Knowledge-leverage-based TSK fuzzy system modeling. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 1200–1212. doi: 10.1109/tnnls.2013.2253617

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T.-S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6

Hang, W., Feng, W., Du, R., Liang, S., Chen, Y., Wang, Q., et al. (2019). Cross-subject EEG signal recognition using deep domain adaptation network. *IEEE Access* 7, 273–282. doi: 10.1109/ACCESS.2019.2939288

Hang, W., Feng, W., Liang, S., Wang, Q., Liu, X., and Choi, K. S. (2020). Deep stacked support matrix machine based representation learning for motor imagery EEG classification. *Compu. Methods Prog. Biomed.* 193:105466. doi: 10.1016/j.cmpb.2020.105466

Hossain, I., Khosravi, A., Hettiarachchi, I., and Nahavandi, S. (2018). Multiclass informative instance transfer learning framework for motor imagery-based brain-computer interface. *Comput. Intell. Neurosci.* 2018:6323414. doi: 10.1155/2018/6323414

Hossain, I., Khosravi, A., and Nahavandhi, S. (2016). "Active transfer learning and selective instance transfer with active learning for motor imagery based BCI," in *Proceeding of the IEEE International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, 4048–4055. doi: 10.1109/IJCNN.2016.7727726

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545

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

Li, Y., Kambara, H., Koike, Y., and Sugiyama, M. (2010). Application of covariate shift adaptation techniques in brain–computer interfaces. *IEEE Trans. Biomed. Eng.* 57, 1318–1324. doi: 10.1109/tbme.2009.2039997

Lotte, F., and Guan, C. (2010). "Learning from other subjects helps reducing brain-computer interface calibration time," in *Proceeding of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, 614–617. doi: 10.1109/ICASSP.2010.5495183

Luo, L., Xie, Y., Zhang, Z., and Li, W. J. (2015). "Support matrix machines," in *Proceeding of the International Conference on International Conference on Machine Learning*, France, 938–947.

Morioka, H., Kanemura, A., Hirayama, J.-I., Shikauchi, M., Ogawa, T., Ikeda, S., et al. (2015). Learning a common dictionary for subject-transfer decoding with resting calibration. *NeuroImage* 111, 167–178. doi: 10.1016/j.neuroimage.2015.02.015

Nam, C. S., Jeon, Y., Kim, Y. J., Lee, I., and Park, K. (2011). Movement imagery-related lateralization of event-related (de) synchronization (ERD/ERS): motor-imagery duration effects. *Clin. Neurophysiol.* 122, 567–577. doi: 10.1016/j.clinph.2010.08.002

Pan, S., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Pfurtscheller, G., and Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Pirsiavash, H., Ramanan, D., and Fowlkes, C. C. (2009). Bilinear classifiers for visual recognition. *Adv. Neural Inform. Process. Syst.* 22, 1482–1490.

Salvaris, M., and Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *J. Neural Eng.* 6:046011. doi: 10.1088/1741-2560/6/4/046011

Samek, W., Meinecke, F. C., and Müller, K.-R. (2013). Transferring subspaces between subjects in brain–computer interfacing. *IEEE Trans. Biomed. Eng.* 60, 2289–2298. doi: 10.1109/tbme.2013.2253608

Vapnik, V. (1995). *The Nature of Statistical Learning.* Berlin: Springer-Verlag. doi: 10.1007/978-1-4757-2440-0

Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., and Pfurtscheller, G. (2005). Adaptive on-line classification for EEG-based brain computer interfaces with AAR parameters and band power estimates. *Biomed. Technik Biomed. Eng.* 50, 350–354. doi: 10.1515/bmt.2005.049

Vourvopoulos, A., Badia, S. B. I., and Liarokapis, F. (2017). EEG correlates of video game experience and user profile in motor-imagery-based brain–computer interaction. *Vis. Comput.* 33, 533–546. doi: 10.1007/s00371-016-1304-2

Wang, P., Lu, J., Zhang, B., and Tang, Z. (2015). "A review on transfer learning for brain-computer interface classification," in *Proceeding of the 2015 5th International Conference on Information Science and Technology (ICIST)*, Piscataway, NJ: IEEE, 315–322. doi: 10.1109/ICIST.2015.7288989

Wang, Z., Yu, Y., Xu, M., Liu, Y., Yin, E., and Zhou, Z. (2019). Towards a hybrid BCI gaming paradigm based on motor imagery and SSVEP. *Int. J. of Hum. Comput. Interact.* 35, 197–205. doi: 10.1080/10447318.2018.1445068

Wolf, L., Jhuang, H., and Hazan, T. (2007). "Modeling appearances with low rank svm," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, 1–6. doi: 10.1007/978-3-319-12000-3_1

Yang, J., Yan, R., and Hauptmann, A. G. (2007). "Adapting SVM classifiers to data with shifted distributions," in *Proceeding of the Seventh IEEE International Conference on Data Mining Workshops*, Omaha, NE, 69–76. doi: 10.1109/ICDMW.2007.37

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

Zheng, Q., Zhu, F., and Heng, P. A. (2018a). Robust support matrix machine for single trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 551–562. doi: 10.1109/tnsre.2018.2794534

Zheng, Q., Zhu, F., Qin, J., Chen, B. D., and Heng, P. A. (2018b). Sparse support matrix machine. *Pattern Recognit.* 76, 715–726. doi: 10.1016/j.patcog.2017.10.003

Zheng, Q., Zhu, F., Qin, J., and Heng, P. (2018c). Multiclass support matrix machine for single trial EEG classification. *Neurocomputing* 275, 869–880. doi: 10.1016/j.neucom.2017.09.030