

Article

# Regularized RKHS-Based Subspace Learning for Motor Imagery Classification

Linzhi Jiang<sup>1</sup>, Shuyu Liu<sup>2,\*</sup>, Zhengming Ma<sup>1,\*</sup> , Wenjie Lei<sup>1</sup> and Cheng Chen<sup>1</sup>

<sup>1</sup> School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China; Jianglzh5@mail2.sysu.edu.cn (L.J.); leiwj@mail2.sysu.edu.cn (W.L.); Chench385@mail2.sysu.edu.cn (C.C.)

<sup>2</sup> Public Experimental Teaching Center, Sun Yat-sen University, Guangzhou 510006, China

\* Correspondence: ljie@mail.sysu.edu.cn (S.L.); issmzm@mail.sysu.edu.cn (Z.M.)

**Abstract:** Brain–computer interface (BCI) technology allows people with disabilities to communicate with the physical environment. One of the most promising signals is the non-invasive electroencephalogram (EEG) signal. However, due to the non-stationary nature of EEGs, a subject’s signal may change over time, which poses a challenge for models that work across time. Recently, domain adaptive learning (DAL) has shown its superior performance in various classification tasks. In this paper, we propose a regularized reproducing kernel Hilbert space (RKHS) subspace learning algorithm with K-nearest neighbors (KNNs) as a classifier for the task of motion imagery signal classification. First, we reformulate the framework of RKHS subspace learning with a rigorous mathematical inference. Secondly, since the commonly used maximum mean difference (MMD) criterion measures the distribution variance based on the mean value only and ignores the local information of the distribution, a regularization term of source domain linear discriminant analysis (SLDA) is proposed for the first time, which reduces the variance of similar data and increases the variance of dissimilar data to optimize the distribution of source domain data. Finally, the RKHS subspace framework was constructed sparsely considering the sensitivity of the BCI data. We test the proposed algorithm in this paper, first on four standard datasets, and the experimental results show that the other baseline algorithms improve the average accuracy by 2–9% after adding SLDA. In the motion imagery classification experiments, the average accuracy of our algorithm is 3% higher than the other algorithms, demonstrating the adaptability and effectiveness of the proposed algorithm.

**Keywords:** EEG; brain–computer interfaces; domain adaptation; reproducing kernel Hilbert space; SLDA



**Citation:** Jiang, L.; Liu, S.; Ma, Z.; Lei, W.; Chen, C. Regularized RKHS-Based Subspace Learning for Motor Imagery Classification. *Entropy* **2022**, *24*, 195. <https://doi.org/10.3390/e24020195>

Academic Editor: Carlos M. Travieso-González

Received: 25 November 2021

Accepted: 17 January 2022

Published: 27 January 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Non-invasive BCIs enable people to communicate with electronic devices by analyzing the electrical or magnetic signals generated by the brain’s nervous system. Due to the advantages of non-invasiveness, low cost, portability and high temporal resolution for different brain activity monitoring modalities, electroencephalography (EEG) has been used in many non-invasive BCI studies [1]. Depending on the strategy used to control the device, BCIs systems can be classified as endogenous or exogenous [2]. Exogenous task BCIs systems are based on evoked activities that require external stimuli, such as visual evoked potentials. In contrast, endogenous BCIs are based on spontaneous activities, such as motor imagery (MI), in which the subject needs to focus on a specific mental task [3].

Motor imagery signals are body parts that imagine movement in the absence of actual movement. Different MI tasks lead to oscillatory activity observed in different areas of the sensorimotor cortex of the brain [4]. Various MI-based BCI applications have been used as rehabilitation for wheelchair and prosthetic control in disabled patients [5–8], and for recreation in healthy individuals [9,10].

EEG signals are prone to be affected by individual mental states, such as mood and attention. In the BCI’s MI experiment, subjects were asked to repeat the motor imagery tasks

of their left hand, right hand, foot and tongue on two different days. This concentration constraint is a very tiring mental task for the subjects [2]. The BCIs system uses a fixed time interval for all subjects, which is considered as one of the drawbacks of this model [11]. The MI experiment depends on the subject and there is no way to define exactly when the effect of motor imagery appears after the cue [12], i.e., it can appear immediately after the cue or after a period of time. Therefore, the time interval for the onset of imagery may vary across subjects. Human consciousness is complex and unstable, and once the subject's consciousness is separated in an experiment, the results can be severely biased. It has been shown that the EEG signal of the BCI-based MI task has high variability in the subprojects [13]. Therefore, when the data include measurements from different time periods, there is no guarantee that the spatial distribution of EEG data is consistent across days, even when the same task is performed. This multi-domain data poses a major challenge for machine learning methods.

Domain adaptation learning (DAL) is a branch of transfer learning for solving cross-domain learning problems, in which the training data and test data are from different domains. In general, the problem of domain adaptation processing involves a well-labeled source domain and an unlabeled target domain with different probability distributions. The goal is to solve the task in the target domain by knowledge transfer using the labeled information from the source domain [14]. In the BCI experience, the experimental data of the first day can be considered as the source domain and the experimental data of the second day can be considered as the target domain, and the knowledge obtained from the source domain can be transferred to the target domain. This is a typical domain adaptation problem. The methods of domain adaptation are minimizing the distribution differences in the feature space, such as joint distribution adaptation (JDA) [15], joint geometric and statistical alignment (JGSA) [16], and manifold embedding distribution alignment (MEDA) [17]. Domain adaptation methods have been applied in many fields, such as image classification, object recognition [18], text classification [14] and video event detection [19]. Previous studies [20,21] have shown the effectiveness of domain adaptation methods in mitigating the data distribution of different subjects or different stages in BCI. Therefore, domain adaptation learning (DAL) is the best choice for signal recognition in BCI systems.

The source and target domains can be transformed into another feature space by a transformation (e.g., kernel function), and then applied by the machine learning method. Kernel functions are a suitable class of feature mapping functions that implicitly map data to a high-dimensional RKHS and explicitly provide the inner product of the data in the space. RKHS subspace learning is a common framework for transfer learning, which learns a suitable subspace in the RKHS according to a specific machine learning task.

The most commonly used nonparametric distance estimation method for measuring the distance of statistical feature distribution between the source and target domain data is the maximum mean difference (MMD), which was proposed by Gretton et al. [22] and Smola et al. [23]. Based on MMD, Pan et al. [24] proposed transfer component analysis (TCA), which maps data from the source and target domains to a high-dimensional RKHS. However, MMD uses the first-order moments of the origin of two random variables to measure the distance between two probability distributions, which does not accurately describe the local differences between the two distributions. Therefore, it is common practice to add regularization terms to compensate for the shortcomings of MMD. For example, semi-supervised transfer component analysis (SSTCA) [24] adds a streamwise regularization term to TCA, which can reduce the distance in data distributions between domains and maximize label dependence in a latent space. Jiang et al. [25] proposed to integrate the global and local metrics for domain adaptive learning (IGLDA). Based on TCA, IGLDA considers both the local data information and overall information to make the source and target domain data as close as possible while preserving the geometric properties of the source domain data. Li et al. [26] proposed a domain adaptation algorithm framework that maps data from two domains to RKHS with feature selection and a

maximum regularization term for the variance of the target domain data in the subspace. Our experiments show that their algorithm improves the classification accuracy to some extent, but ignores the optimization problem of the source domain data and its labels. In domain adaptation, source domain data with label is an important information source. How to use label information is always the focus of various domain adaptation algorithms. Lei et al. [27] applied the dictionary learning to the source domain while we borrowed the idea of LDA.

In this paper, we develop a new approach based on RKHS subspace learning and apply it to motor imagery recognition. It attempts to learn the coefficients of the RKHS subspace so that the differences in data distribution across domains can be reduced when projecting to that subspace. Machine learning approaches, such as classification and regression models, can be used in this subspace. Additionally, to make full use of the source domain information, we propose a source linear discriminant analysis (SLDA) regularization term. Specifically, considering the sensitivity of BCI data, we sparsely construct the RKHS subspace framework using the L2.1 criterion. The primary contributions of this paper are summarized as follows:

- (1) We reformulate the RKHS subspace learning framework (RKHS-DA), and propose the SLDA regularization term to remedy the deficiency of MMD in domain adaptation.
- (2) To address the problem of complex and unstable EEG signal, we choose features wisely in the low-dimensional subspace projected to the data through the L2.1 criterion to constrain the coefficient matrix.
- (3) Experimental results show that the average accuracy of our algorithm is 3% higher than other algorithms.

The remainder of this paper is organized as follows. Section 2 presents a general description of our approach. Section 3 describes the proposed framework and the SLDA regularization terms in detail. We validate our SLDA regularization and RKHS subspace learning framework, and the experimental results are presented in Section 4.

## 2. Preliminaries

### 2.1. Notations

In this paper, we use a combination of letters and numbers to represent data. A sample is denoted as a vector, e.g., the  $i$ th sample of  $x$  in a set is denoted as  $x_i$ . We also use the subscripts  $s$  and  $t$  to indicate the source domain and the target domain, respectively. For a matrix  $M$ , the trace of matrix  $M$  is denoted by  $\text{tr}(M)$ . For clarity, the frequently used notations and corresponding descriptions are shown in Table 1.

**Table 1.** Notation and description.

Notation	Description
$X_s, Y_s$	Original/subspace source data
$X_t, Y_t$	Original/subspace target data
$L$	MMD matrix
$\lambda, \mu, \gamma, \beta$	Penalty parameters
$K$	Kernel matrix
$W$	Projection matrix
$I$	Identity matrix

### 2.2. Reproducing Kernel Hilbert Spaces (RKHS)

#### 2.2.1. Hilbert Spaces

Definition (inner product space [28]): let  $H$  be the linear space on the real number domain  $R$ ,  $\langle \bullet, \bullet \rangle : H \rightarrow R$ , with the following properties:

- (1) Positive definiteness: for all  $x \in H$ ,  $\langle x, x \rangle$  and  $\Leftrightarrow x = 0$ ;
- (2) Symmetry: For all  $x, y \in H$ ,  $\langle x, y \rangle = \langle y, x \rangle$ ;
- (3) Bilinear: For all  $x, y, z \in H$  and  $\alpha, \beta \in R$ ,

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$$

Then, we consider that  $\langle \bullet, \bullet \rangle$  is the inner product of  $H$ , and  $(H, \langle \bullet, \bullet \rangle)$  is an inner product space.

Let  $x$  be an element of inner product space  $(H, \langle \bullet, \bullet \rangle)$ . In the inner product space, the norm is defined by the inner product:

$$\|x\| = \sqrt{\langle x, x \rangle} \tag{1}$$

According to the nature of the positive definite inner product:

$$\|x\| = \sqrt{\langle x, x \rangle} = 0 \Leftrightarrow x = 0 \tag{2}$$

then we have

$$\|x - y\| = 0 \Leftrightarrow x = y \tag{3}$$

If all the basic sequences are convergent in this inner product space which is known as a Hilbert space.

### 2.2.2. Definition of Reproducing Kernel Hilbert Space (RKHS)

Let  $H = \{f | f : \Omega \rightarrow R, \int_{\Omega} |f(x)|^2 < +\infty\}$  be a square integrable function space. It is clear that  $H$  is a linear space. We define  $\langle \bullet, \bullet \rangle : H \times H \rightarrow R$ , for any  $f, g \in H$

$$\langle f, g \rangle = \int_{\Omega} f(x)g(x)dx \tag{4}$$

It can be shown that  $\langle \bullet, \bullet \rangle$  is an inner product and  $(H, \langle \bullet, \bullet \rangle)$  is a Hilbert space. Further, if there is  $k : \Omega \times \Omega \rightarrow R$ , satisfy

- (1) For any  $x \in \Omega, k_x = k(\bullet, x) \in H$ ;
- (2) For any  $x \in \Omega$  and  $f \in H$ , we have

$$f(x) = \langle f, k_x \rangle = \langle f(\bullet), k(\bullet, x) \rangle \tag{5}$$

Therefore,  $H$  can be called a reproducing kernel Hilbert space (RKHS), and  $k$  is the reproducing kernel of  $H$ . Using reproducing kernel  $k$ , we can define mapping  $\varphi : \Omega \rightarrow H$ : for any  $x \in \Omega$ , we have

$$\varphi(x) = k(\bullet, x) = k_x \in H \tag{6}$$

From Equation (6), it can be proved that

$$\langle \varphi(x), \varphi(y) \rangle = \langle k_x, k(\bullet, y) \rangle = k_x(y) = k(y, x) = k(x, y) \tag{7}$$

### 2.3. Hilbert Subspace Projection Theorem

Definition of projection: let  $(H, \langle \bullet, \bullet \rangle)$  be an inner product space and  $A$  be a subspace of  $H$ . For  $x_0 \in H$ . If  $x_0$  can be decomposed into  $x_0 = x'_0 + x''_0$ , where  $x'_0 \in A, \langle x'_0, x''_0 \rangle = 0$ , then  $x'_0$  is called the projection of  $x_0$  in subspace  $A$ .

Projection theorem:  $(H, \langle \bullet, \bullet \rangle)$  is an inner product space.  $A$  is a finite dimensional subspace of  $H$  and  $\{e_1, \dots, e_d\}$  is the standard orthogonal basis of  $A$ . For any  $x_0 \in H$ , the projection  $x'_0$  of  $x_0$  in  $A$  is as follows:

$$x'_0 = \sum_{i=1}^d \langle x_0, e_i \rangle e_i \in A \tag{8}$$

**Remark 1.** *A is a finite-dimensional subspace of H; therefore, A is complete, i.e., A is a Hilbert subspace, so the projection of any point in H onto A exists.*

2.4. Domain Adaptation Learning and MMD

There are two datasets in data space  $\Omega$ : the labeled source domain data  $X_s = \{x_1^s, \dots, x_{N_s}^s\} \subseteq \Omega$ , and the unlabeled target domain data  $X_t = \{x_1^t, \dots, x_{N_t}^t\} \subseteq \Omega$ , and the distributions of  $X_s$  and  $X_t$  in the data space are different. We need to classify  $X_t$  based on  $X_s$ . This problem is domain adaptation learning. In our work, we resorted to MMD [22], a nonparametric metric to measure the distance between distributions, which can transform the source domain data  $X_s$  and target domain data  $X_t$  onto RKHS  $H$  generated by the reproducing kernel  $k$ , i.e.,

$$\Phi(X_s) = \{\Phi(x_1^s), \dots, \Phi(x_{N_s}^s)\} \subseteq H, \Phi(X_t) = \{\Phi(x_1^t), \dots, \Phi(x_{N_t}^t)\} \subseteq H \tag{9}$$

In this way, the distribution of  $\Phi(X_s)$  and  $\Phi(X_t)$  in RKHS  $H$  can be as similar as possible. Moreover, the similarity here can exactly be measured by MMD:

$$\text{MMD}^2(X_s, X_t) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \Phi(x_i^s) - \frac{1}{N_t} \sum_{i=1}^{N_t} \Phi(x_i^t) \right\|^2 \tag{10}$$

where  $\varphi(\cdot)$  is the mapping defined by reproducing kernel  $k$ .

In practice, it is not easy to learn an optimal RKHS  $H$  based on MMD. Most methods based on MMD choose to learn a linear subspace  $\text{span}\Theta$  of RKHS  $H$ , so the MMD distance can be expressed as follows:

$$\text{MMD}^2(X_s, X_t) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \Phi_{\text{span}\Theta}(x_i^s) - \frac{1}{N_t} \sum_{i=1}^{N_t} \Phi_{\text{span}\Theta}(x_i^t) \right\|^2 \tag{11}$$

where  $\Phi_{\text{span}\Theta}(X_s)$  and  $\Phi_{\text{span}\Theta}(X_t)$  mean the projection of  $\Phi(X_s)$  and  $\Phi(X_t)$  in the subspace, respectively.

**3. Domain Adaptation Based on Source LDA Regularized RKHS Subspace Learning and Its Application in BCI**

3.1. Reformulation of the RKHS Subspace Learning Framework

3.1.1. Construction of RKHS

The regenerated kernel of RKHS is used to construct the transformation from the original data space to RKHS, rather than defining the transformation first and then using the transformation and the inner product of RKHS to define the so-called “kernel function”, which is not actually the reproducing kernel of RKHS. However, many studies have used the reproducing kernel to define the transformations from original data space to RKHS, ignoring the connection between the original data space and RKHS. Therefore, we reformulated a mathematical framework model of RKHS in this section.

Let  $(H, \langle \bullet, \bullet \rangle)$  be the RKHS on the data space  $\Omega$ , and use the reproducing kernel  $k$  of  $H$  to define the transformation from the data space  $\Omega$  to  $H$ :  $\varphi : \Omega \rightarrow H$ , for any  $x \in \Omega$ , we define  $\varphi(x) = k(\bullet, x) \in H$ , so for any  $x, y \in \Omega$ , we have  $\langle \varphi(x), \varphi(y) \rangle = k(x, y)$ .

Now, given a set of data on data space  $\Omega$ ,

$$X = \{x_1, \dots, x_N\} \subseteq \Omega \tag{12}$$

Feature map  $\varphi(\cdot)$  is used to transform  $X$  to  $H$

$$\varphi(X) = \{\varphi(x_1), \dots, \varphi(x_N)\} \subseteq H \tag{13}$$

The kernel matrix  $K$  is represented as

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix} = [K_{1Col} \cdots K_{NCol}] \in R^{N \times N} \quad (14)$$

where  $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ , and  $K_{iCol}$  is the  $i$ th column vector of  $K$ ,  $i = 1, \dots, N$ .

### 3.1.2. The Construction and Restraint of the RKHS Subspace

$\varphi(X)$  is used to construct a basis of subspace of  $H$ :

$$\theta_i = \sum_{j=1}^N w_{ji} \varphi(x_j) \quad i = 1, \dots, d \quad (15)$$

We define

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{Nd} \end{bmatrix} = [W_{1Col} \cdots W_{dCol}] \in R^{N \times d} \quad (16)$$

where  $W_{iCol}$  is the  $i$ th column vector of  $W$ , and  $i = 1, \dots, d$ . We use  $\Theta = \{\theta_1, \dots, \theta_d\}$  to span a subspace of  $H$ :

$$span\Theta = \left\{ \sum_{i=1}^d \alpha_i \theta_i \mid \alpha_i \in R, i = 1, \dots, d \right\} \quad (17)$$

To constitute the orthonormal basis of subspace  $span\Theta$ ,  $\Theta$  satisfies the condition of bivariate orthogonality:

$$\begin{bmatrix} \langle \theta_1, \theta_1 \rangle & \cdots & \langle \theta_1, \theta_d \rangle \\ \vdots & \ddots & \vdots \\ \langle \theta_d, \theta_1 \rangle & \cdots & \langle \theta_d, \theta_d \rangle \end{bmatrix} = \begin{bmatrix} W_{1Col}^T K W_{1Col} & \cdots & W_{1Col}^T K W_{dCol} \\ \vdots & \ddots & \vdots \\ W_{dCol}^T K W_{1Col} & \cdots & W_{dCol}^T K W_{dCol} \end{bmatrix} = W^T K W = I_d \quad (18)$$

Subspace  $span\Theta$  is a  $d$ -dimensional subspace and is determined by the data transformed to subspace and combination coefficient  $W$ , which satisfies the above constraints.

### 3.1.3. Representation of Data in the RKHS Subspace

Since RKHS is an infinite dimensional space, machine learning algorithms cannot be directly applied to such space, so it needs to project the data in RKHS into the subspace of RKHS. According to the projection theorem, if  $\{\theta_1, \dots, \theta_d\}$  is the orthonormal basis of subspace  $span\Theta$ , then the coordinates of the projection of  $\varphi(x_i)$  on subspace  $span\Theta$  are

$$y_i = \begin{bmatrix} \langle \varphi(x_i), \theta_1 \rangle \\ \vdots \\ \langle \varphi(x_i), \theta_d \rangle \end{bmatrix} = \begin{bmatrix} W_{1Col}^T K_{iCol} \\ \vdots \\ W_{dCol}^T K_{iCol} \end{bmatrix} = W^T K_{iCol} \quad (19)$$

where  $i = 1, \dots, d$ . By constructing the subspace of RKHS, we implemented the transformation of data from the original data space  $\Omega$  to the Euclidean space  $R^d$ :

$$X = \{x_1, \dots, x_N\} \subseteq \Omega \Rightarrow Y = \{y_1, \dots, y_N\} \in R^d \quad (20)$$

The working space is Euclidean space  $R^d$ , of which  $W$  will be determined according to the specific machine learning task. The orthonormal basis of the subspace is constructed by the linear combination of transformed samples. Due to the requirements of the orthonormal

basis, the constraints of the combination coefficient of the subspace are obtained, rather than based on some assumptions. The original data is transformed to RKHS and projected to the subspace. The coordinates of this projection on the orthonormal basis of the subspace are the final representation of the original data.

### 3.2. Domain Adaptation Based on SLDA Regularized RKHS Subspace Learning

#### 3.2.1. Domain Adaptation Based on RKHS Subspace Learning and MMD

Given a set of source domain data and a set of target domain data in the original data space  $\Omega$ :

$$X_s = \{x_1^s, \dots, x_{N_s}^s\} \subseteq \Omega, X_t = \{x_1^t, \dots, x_{N_t}^t\} \subseteq \Omega \tag{21}$$

The source domain data  $X_s$  is labeled while the target domain data  $X_t$  is unlabeled. We define the following:

$$X = X_s \cup X_t = \{x_1, \dots, x_N\} = \{x_1^s, \dots, x_{N_s}^s, x_1^t, \dots, x_{N_t}^t\} \subseteq \Omega, N = N_s + N_t, \tag{22}$$

Using the RKHS subspace learning framework proposed in Section 3.1, we have

$$Y_s = \{y_1^s, \dots, y_{N_s}^s\} \subseteq R^d, Y_t = \{y_1^t, \dots, y_{N_t}^t\} \subseteq R^d \tag{23}$$

where  $y_i^s = W^T K_{iCol}$ ,  $i = 1, \dots, N_s$  and  $y_i^t = W^T K_{(N_s+i)Col}$ ,  $i = 1, \dots, N_t$ . In the expressions of  $Y_s$  and  $Y_t$ , the matrix  $W$  is unknown and represents the subspace of RKHS, desired distribution of  $Y_s$  and  $Y_t$  in the Euclidean space  $R^d$  can be achieved by learning  $W$ . To measure the difference between two distributions, MMD between  $X_s$  and  $X_t$  can be calculated as follows:

$$MMD^2(X_s, X_t) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} y_i^s - \frac{1}{N_t} \sum_{j=1}^{N_t} y_j^t \right\|^2 \tag{24}$$

where  $y_i^s = W^T K_{iCol}$  and  $y_j^t = W^T K_{(N_s+j)Col}$ , the MMD distance in (25) can be rewritten as

$$MMD^2(X_s, X_t) = tr(W^T L W) \tag{25}$$

and

$$L = \left( \frac{1}{N_s} \sum_{i=1}^{N_s} K_{iCol} - \frac{1}{N_t} \sum_{j=1}^{N_t} K_{(N_s+j)Col} \right) \left( \frac{1}{N_s} \sum_{i=1}^{N_s} K_{iCol} - \frac{1}{N_t} \sum_{j=1}^{N_t} K_{(N_s+j)Col} \right)^T \tag{26}$$

#### 3.2.2. Domain Adaption Based on Source LDA Regularized RKHS Subspace Learning (SLDARKHS-DA)

MMD is an approximate criterion rather than an exact one. Therefore, it is common practice to add regularization terms to compensate for the deficiency of MMD. In transfer learning, KNN is a commonly used classifier. To improve the classification efficiency of KNN, we considered the reduction of the within-class scatter between the source domain and the target domain, while increasing the between-class scatter. Since the target domain is usually unlabeled, the SLDA proposed in this section only applies to the source domain data.

During the distribution matching process, it would be helpful to keep samples of the same class close to each other while the samples of different classes are far from each other. For this purpose, we define the transformed source domain data as  $C$  categories, and each category has  $N_c$  data samples, which can be expressed as:

$$\{y_1, \dots, y_N\} = \left\{ y_{11}^s, \dots, y_{1N_1}^s, \dots, y_{C1}^s, \dots, y_{cN_c}^s \right\}, N_s = \sum_{c=1}^C N_c \tag{27}$$

where  $y_{ci}^s$ ,  $c = 1, \dots, C$ ,  $i = 1, \dots, N_c$ .

The center of the  $c$ th class can be computed as follows:

$$\bar{y}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} y_{ci} = \frac{1}{N_c} \sum_{i=1}^{N_c} W^T K_{c(i)Col} = W^T \left( \frac{1}{N_c} \sum_{i=1}^{N_c} K_{c(i)Col} \right) = W^T \bar{K}_{Col}^c \quad (28)$$

Moreover, the center of all the samples can also be computed as follows:

$$\bar{y} = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{N_c} y_{ci} = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{N_c} W^T K_{c(i)Col} = W^T \left( \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{N_c} K_{c(i)Col} \right) = W^T \bar{K}_{Col} \quad (29)$$

where  $\bar{K}_{Col}^c = \frac{1}{N_c} \sum_{i=1}^{N_c} K_{c(i)Col}$ ,  $\bar{K}_{Col} = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{N_c} K_{c(i)Col}$

- (1) To increase the distance between the different types of source domain data, the between-class scatter can be defined and rewritten as:

$$S_b = \sum_{c=1}^C \frac{N_c}{N_S} \|\bar{y}_c - \bar{y}\|^2 = tr(W^T \Psi W) \quad (30)$$

where  $\Psi = \sum_{c=1}^C \frac{N_c}{N_S} (\bar{K}_{Col}^c - \bar{K}_{Col})(\bar{K}_{Col}^c - \bar{K}_{Col})^T$

- (2) To improve the discriminative efficiency of the same category data in the subspace, the intra-class divergence can be expressed as follows:

$$S_w = \frac{1}{N_S} \sum_{c=1}^C \sum_{i=1}^{N_i} \|y_{ci} - \bar{y}_c\|^2 = tr(W^T \Phi W) \quad (31)$$

where  $\Phi = \frac{1}{N_S} \sum_{c=1}^C \sum_{i=1}^{N_i} (K_{c(i)Col} - \bar{K}_{Col}^c)(K_{c(i)Col} - \bar{K}_{Col}^c)^T$ . The distance between the same types of data in the source domain is reduced, so that the same type of data will be more concentrated.

### 3.2.3. Solution

Since the target domain data is used to rely the source domain data in the subspace, the optimization of the target domain data in the subspace will not be effective if KNN is used to identify the target domain data in the subspace. By adding the regularization term of SLDA, the overall objective function of our proposed SLDARKHS-DA can be formulated as follows:

$$\begin{aligned} \min_W \quad & tr(W^T L W) + \lambda tr(W^T (\Phi - \Psi) W) + \mu tr(W^T W) = tr(W^T N W) \\ \text{s.t.} \quad & W^T K W = I_d \end{aligned} \quad (32)$$

where  $N = L + \lambda(\Phi - \Psi) + \mu I$ . This model can be solved by the properties of generalized Rayleigh entropy. Since  $K$  is symmetric positive definite, it can be expressed as

$$K = U \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} U^T \quad (33)$$

where  $U U^T = I$ ,  $\Sigma^{\frac{1}{2}} = \text{diag}(\sqrt{\sigma_1}, \dots, \sqrt{\sigma_N})$ ,  $\sigma_i$  are the eigenvalues of  $K$ ,  $i = 1, \dots, N$ .  $W^T K W$  can be rewritten as  $W^T K W = W^T U \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} U^T W = V^T V = I_d$  according to the above restrictions, and  $V = \Sigma^{\frac{1}{2}} U^T W$ . We reformulate (32) by  $V$ :

$$tr(W^T N W) = tr(V^T M V) \quad (34)$$



where  $M = \Sigma^{-\frac{1}{2}}U^T(L + \gamma(\Phi - \Psi) + \mu I)U\Sigma^{-\frac{1}{2}}$ . It can be solved by the generalized Rayleigh entropy.  $V$  is a  $d$ -dimensional row vector, which is the eigenvector corresponding to the first  $d$  smallest eigenvalues of matrix  $M$ .

### 3.2.4. Computational Complexity

The computational complexity of the SLDARKHS-DA Algorithm 1 consists of the following three main components: (1) the complexity of the feature problem optimization in step 2, (2) computing  $\Phi$  and  $\Psi$  and (3) the computing of  $K$  and  $L$ . The complexity is usually expressed in terms of  $O$ , and the complexity of the generalized eigen-decomposition is  $O(dn^2)$  ( $d$  is the dimension of the subspace). By computing  $\Phi$ ,  $\Psi$  is  $O(dn^2)$ , and by computing  $K$ ,  $L$  is  $O(n^2)$ . Therefore, the total complexity of the algorithm is  $O((4d + 1)n^2)$ .

---

#### Algorithm 1: SLDARKHS-DA

---

**Input:** source domain data set  $X_s$  and target domain data set  $X_t$ , label information of  $X_s$ ; parameters  $\lambda, \mu$  and subspace dimension  $d$ .

**Output:** projection matrix  $W$  and the label information of  $X_t$ .

1. Combine source domain data set and domain data set:  $X = [X_s, X_t]$ ;
  2. Compute  $K, L, \Phi, \Psi, M$ ;
  3. Eigendecompose the matrix  $M$  and select the  $d$  leading eigenvectors to construct the projection matrix  $W$ ;
  4. Project both  $X_s$  and  $X_t$  to obtain data in the subspace,  $y_i^s = W^T K_{iCol}$  and  $y_i^t = W^T K_{(N_s+i)Col}$ . Classify  $y_i^t$  in the subspace by KNN, and  $y_i^s$  is used as the reference.
- 

### 3.3. Application of SLDARKHS-DA to EEG Motor Imagery Recognition

#### 3.3.1. Description of BCI IV 2a Data

Nowadays, many BCI data recognition tasks are handled by domain adaptation methods. Previous studies [20,21] have shown the effectiveness of domain adaptation approaches in reducing the differences in data distribution between subjects or sessions.

The BCI competition dataset is commonly used as a benchmark dataset for BCI domains. This 2a dataset consisted of nine subjects recorded [29]. Subjects were asked to imagine moving four parts of their body: left hand, right hand, foot and tongue. Addressing multi-class problems is an important challenge for the BCI system.

#### 3.3.2. Domain Adaptation Subspace Learning Based on Sparse Regularized RKHS

Considering the complexity of BCI data, when the transformed data are projected into the subspace, the dimensionality reduction will be performed and some irrelevant data features should be discarded. We selected the most favorable data to improve the recognition effect, construct the subspace by row sparse projection matrix and minimize the geometric offset of the data.

We used the  $L_{2,1}$  norm to constrain the  $W$  matrix so that the rows were sparse. The  $L_{2,1}$  norm of matrix  $A$  is defined as follows:

$$\|A\|_{2,1} = \sum_{j=1}^D \sqrt{\sum_{i=1}^C A_{ij}^2}, A \in \mathbb{R}^{C \times D} \tag{35}$$

The  $L_{2,1}$  norm makes the  $L_2$  norm of each line as small as possible, and as many zeros appear in the line as possible to achieve sparsity.

### 3.3.3. Solution

By adding the  $L_{2,1}$  norm of matrix  $W$  as a sparse regularization term, the overall objective function of our proposed SLDARKHS-DA based on sparse regularization terms is formulated as follows, show in Algorithm 2.

$$\begin{aligned} \min_W \quad & tr(W^T(L + \gamma(\Phi - \Psi) + \mu I)W) + \lambda \|W\|_{2,1} \\ \text{s.t.} \quad & W^T K W = I_d \end{aligned} \quad (36)$$

where  $tr(W^T L W)$  represents the MMD distance between the source domain sample and the target domain sample in the subspace. The purpose of  $tr(W^T(\Phi - \Psi)W)$  is to increase the inter-class divergence and reduce the intra-class divergence of the data in the subspace.  $\|W\|_{2,1}$  is the  $L_{2,1}$  norm of matrix  $W$  to the sparse elements that make up the basis of the subspace. The regularization term  $tr(W^T W)$  can avoid the over-fitting of the model. The constraint  $W^T K W = I_d$  serves two purposes: (1) to make the basis of the subspace orthogonal, and (2) to avoid trivial solutions and ensure that  $W$  is not 0.  $\lambda$ ,  $\mu$ ,  $\gamma$  represents the coefficient of the regularization term.

To solve the optimization problem, we introduced a Lagrange multiplier  $\Lambda$ , and the Lagrange function for the model can be obtained as follows:

$$L(W, \Lambda) = tr(W^T(L + \gamma(\Phi - \Psi) + \mu I)W) + \lambda \|W\|_{2,1} - tr((W^T K W - I_d)\Lambda) \quad (37)$$

Then, by taking the derivative of (37) with respect to  $W$ , and setting the derivative to zero, we obtain

$$((L + \gamma(\Phi - \Psi) + \mu I) + \lambda G)W = K W \Lambda \quad (38)$$

Note that  $\|W\|_{2,1}$  is not smooth, so we computed its subgradient  $G$ , which is a diagonal matrix with the  $i$ th diagonal element that equals to

$$G_{ii} = \begin{cases} 0, & \text{if } W^i = 0 \\ \frac{1}{2\|W^i\|}, & \text{otherwise} \end{cases} \quad (39)$$

where  $W^i$  denotes the  $i$ th row of  $W$ . Thus the concatenated multiple transformations can be solved by calculating the  $d$  smallest eigenvectors of  $K W \Lambda$ .

---

#### Algorithm 2: SLDARKHS-DA (Sparse)

---

**Input:** source domain data set  $X_s$  and target domain data set  $X_t$ , label information of  $X_s$ ; parameters  $\gamma$ ,  $\lambda$ ,  $\mu$  and subspace dimension  $d$ .

**Output:** projection matrix  $W$  and the label information of  $X_t$ .

1. Combine the source domain data set and domain data set:  $X = [X_s, X_t]$ ;
2. Computer matrix  $K, L, \Phi, \Psi$  and initialize  $G = I$ .

#### Repeat

3. Optimize  $W$  by solving the eigen-decomposition problem in (38);
4. Update  $G$  by (39).

**Until** convergence or max iteration

5. Project both  $X_s$  and  $X_t$  to obtain the data in the subspace,  $y_i^s = W^T K_{iCol}$  and  $y_i^t = W^T K_{(N_s+i)Col}$ . Classify  $y_i^t$  in the subspace by KNN, and  $y_i^s$  is used as the reference.
- 

## 4. Experiments

To verify the fitness of the SLDA regularization term, we first conducted experiments on four commonly used standard benchmark datasets (faces, objects, handwritten digits and text). We added the SLDA regularization term to the comparison algorithm. For example, we added the SLDA regularization terms to TCA and performed comparison

experiments with the original TCA. Second, we tested the performance of our SLDARKHSDA on the 4th BCI competition 2A dataset and compared it with some classical baselines published in recent years, respectively. All the methods were programmed in MATLAB 2019 and executed on a PC (CPU: Intel i9) with 3.50 GHz and memory: 16 GB. The source programs of the baseline methods used for the comparison can be downloaded from GitHub. The following are available online at <https://github.com/viggin/domain-adaptation-toolbox> (TCA, downloaded in April 2021), <https://github.com/minjiang/iglda> (IGLDA, downloaded in April 2021) and <https://github.com/lijin118/tit> (TIT, downloaded in May 2021).

#### 4.1. Baseline and Parameter Settings

We compared the proposed method with typical subspace learning methods in domain adaptation, such as TCA [24], IGLDA [25] and TIT [26]. The details of each baseline methods are summarized below:

(1) TCA [24] is a typical example of RKHS subspace learning in domain adaptation. TCA converts the original data into RKHS, then finds a subspace in this space to reduce the dimensionality of the data, and then uses MMD to measure the distance between the two domains. Its objective function is as follows:

$$\begin{aligned} \min_W \quad & \text{tr}(W^T K L K W) + \mu \text{tr}(W^T W) \\ \text{s.t.} \quad & W^T K H K W = I_m \end{aligned} \quad (40)$$

where  $\text{tr}(W^T W)$  prevents the over-fitting of data and  $W^T K H K W = I_m$  can maintain the data characteristics of the source domain data and target domain data.

(2) Jiang et al. proposed the integration of global and local metrics for domain adaptation learning (IGLDA) [25], which added the regularization term based on TCA. IGLDA introduces data label information to keep source domain and target domain data as close as possible while preserving the geometric properties of source domain data. The objective function of this method is as follows:

$$\begin{aligned} \min_W \quad & \text{tr}(W^T K L K W) + \alpha \text{tr}(W^T K L_w K W) + \beta \text{tr}(W^T W) \\ \text{s.t.} \quad & W^T K H K W = I_d \end{aligned} \quad (41)$$

where  $L_w$  represents the within-class divergence matrix.

(3) Li proposed another approach in 2019 [26], which combines the manifold regularization terms used in SSTCA, a kind of regularization terms to select features, and regularization terms to minimize the variance in the target domain. In the experiments, the sample selection is performed by iterative experiments, and the final objective function is as follows.

$$\begin{aligned} \min_W \quad & \text{tr}(W^T K L K W) + \alpha \text{tr}(W^T K \xi K W) - \beta \text{tr}(W^T K C K W) + \gamma \|W\|_{2,1} \\ \text{s.t.} \quad & W^T K H K W = I_d \end{aligned} \quad (42)$$

where  $\text{tr}(W^T K C K W)$  can be used to minimize the variance of the target domain data in the subspace, and  $\|W\|_{2,1}$  selects the feature from the data.

In our experiment, we use the K-nearest neighbor (KNN) as a classifier to evaluate the performance of the proposed method. We obtained the parameter setting with the best classification accuracy through grid search and applied the same parameter selection process to the baseline methods. Each of the hyper-parameters used in our experiments was chosen. We chose the best parameter by searching in the range of  $[10^{-15}, 10^2]$ . For simplicity and clarity, we chose an acceptable common set of them, as shown in Table 2. In the experiments of the first four data sets, we used the linear kernel as kernel function, while in the experiments of the BCI data set, we used the radial basis function (RBF).

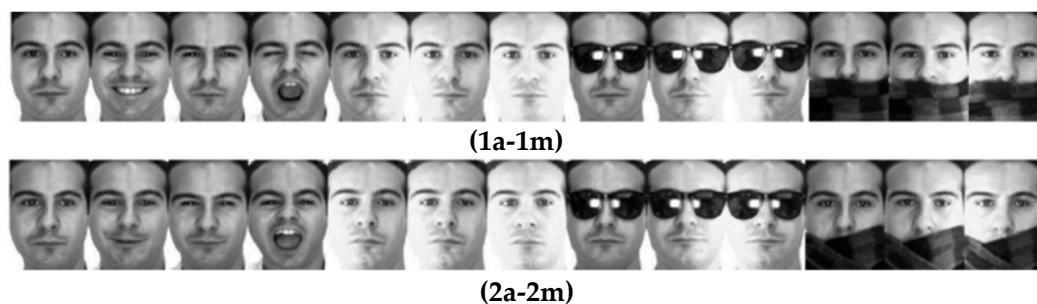
**Table 2.** PARAMETER SETTINGS.

DS	OD	SD	NoN	$\mu$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$
AR	2580	10–100	1	1	$10^2$	$10^{-2}$	$10^{-7}$	$10^{-12}$
4DA	800	80	5	1	$10^2$	$10^{-2}$	$10^{-7}$	$10^{-12}$
MNIST and USPS	256	30–150	5	1	$10^2$	$10^{-2}$	$10^{-2}$	$10^{-12}$
Reuters-215789	$4593 \pm 200$	10–50	5	1	$10^2$	$10^{-15}$	$10^{-7}$	$10^{-9}$
BCI-2a	288	10–110	5	1	$10^{-2}$	$10^{-3}$	$10^{-3}$	$10^{-2}$

DS = dataset, OD = original dimensionality, SD = subspace dimensionality, NoN = number of neighbors in KNN,  $\lambda_1$  for SLDARKHS-DA,  $\lambda_2$  for SLDATCA,  $\lambda_3$  for SLDAIGLDA and  $\lambda_4$  for SLDATIT.

#### 4.2. Face Recognition

In this section, we evaluate the effectiveness of the proposed algorithm in face recognition tasks. The AR dataset [30] is widely used in experiments in the field of face recognition. In this section, we select a subset of AR data set with a total of 2600 face images. This dataset consists of 100 people, with 50 men and 50 women. Each subject has 26 images. The AR face images were captured twice, with an interval of two weeks between the two shots. Each shot collected 13 pictures of different modes with different light brightness, light angle, facial expression and occlusion (sunglasses or scarf). In this experiment, each face image was normalized as a gray level image of pixels. The training set and test set directly used the gray value and vectorization of the image as the input. According to the different shooting times and states, 26 face images of each subject corresponded to 26 patterns, which were numbered as 1a to 1m and 2a to 2m. Figure 1 shows a sample of the AR data set with 26 face images from the same subject. Figure 1a–m belongs to one group, while 2a–2m is from another group, which is under the same conditions taken two weeks later. We used the notation  $\mathbb{C}_{1,a}$  and  $\mathbb{C}_{2,a}$  to represent a collection of natural expressions of the face images. In this section,  $\mathbb{C}_{1,a}$  and  $\mathbb{C}_{2,a}$  are combined as source domain data set  $X_S$ .



**Figure 1.** Sample images of the AR face dataset. (1a–1m) and (2a–2m) represent two groups of 13 face pictures from left to right.

From the other 24 patterns except  $\mathbb{C}_{1,a}$  and  $\mathbb{C}_{2,a}$ , the first 18 patterns were selected, including  $\mathbb{C}_{1,b}$  to  $\mathbb{C}_{1,j}$  and  $\mathbb{C}_{2,b}$  to  $\mathbb{C}_{2,j}$ , and the data of these patterns were taken as 18 target domain data sets respectively and 18 classification tasks were set.

In the first experiment, we studied how our proposed SLDARKHS-DA affected the distribution of the source and target domains. We took  $\mathbb{C}_{1,f}$  and  $\mathbb{C}_{2,f}$  as target domain  $X_T$ , respectively, and calculated the distance between the geometric center of the source domain and target domain and the variance of the source domain data in the original space and subspace, to prove the effectiveness of domain adaptation. As shown in Table 3, in the experiment with  $\mathbb{C}_{1,f}$  as the target domain, after the data of the source domain and target domain are transformed from the original space to the subspace, not only do the geometric centers of the data of the two almost coincide, but also the variance of the data is greater. As the distance between the classes in the source domain becomes larger, the classification

efficiency of the KNN algorithm improves. Similarly, the geometric distribution of data with  $\mathbb{C}_{2,f}$  as the target domain, also shows a similar change.

**Table 3.** Data distribution in the original space and subspace for face recognition.

Task	Original Space			Subspace		
	D (S, T)	Var (S)	Var (T)	D (S, T)	Var (S)	Var (T)
1.f	3182	1867	2065	$9 \times 10^{-3}$	5393	5561
2.f	3060	1867	2036	$1 \times 10^{-12}$	5448	5624

D (S, T) = distance between the source domain and target domain, and Var = variance.

In the second experiment, we regard the data in the source domain as labeled and the data in the target domain as unlabeled. Similarly, we combined  $\mathbb{C}_{1,a}$  and  $\mathbb{C}_{2,a}$  as the source domains and set the target domains as  $\mathbb{C}_{1,b}$  to  $\mathbb{C}_{1,j}$  and  $\mathbb{C}_{2,b}$  to  $\mathbb{C}_{2,j}$ . A total of 30% of the images from the target domain were crystals selected as training data and the transformation function by the domain adaptation methods obtained by  $X_T$  and  $X_S$ . We set the KNN as the default classifier and the expected subspace dimensionality was fixed at 90 for the classification experiments. The experimental results are shown in Table A1 in Appendix A.

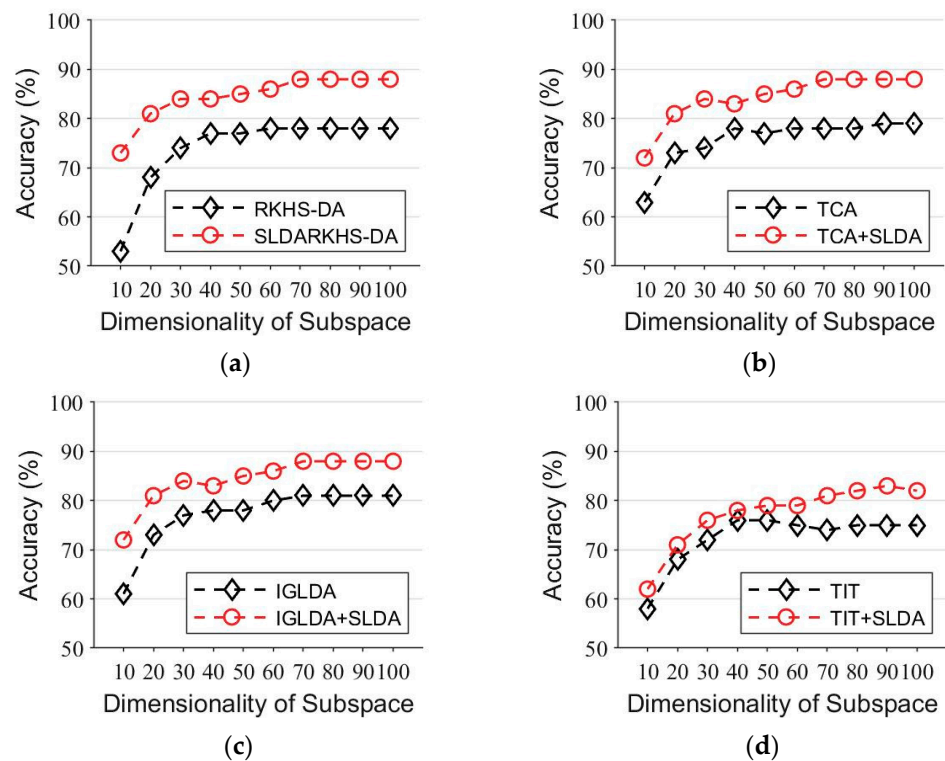
We noted that the direct classification of  $X_T$  data with the KNN was worse than the classification of the mapped  $\Gamma(X_T)$  data obtained by RKHS-DA with KNN. Moreover, the classification accuracy of the SLDARKHSDA algorithm combining the regularization terms SLDA and RKHS-DA improved by 4% on average.

Furthermore, we combined the proposed regularization term SLDA with the baseline algorithm for comparison to form a new domain adaptation algorithm, which was compared with the original baseline algorithm. As shown in Table A3, in terms of average classification accuracy, the SLDA improves the TCA by 3.1%, IGLDA by 1.2% and TIT by 2.8%, respectively. This agrees with our idea of RKHS subspace learning, because the baseline algorithm compared with our RKHS-DA algorithm is similar in terms of domain adaptation; therefore, SLDA also improves the performance of such an algorithm.

In the third experiment, to investigate how the dimensionality of the subspace of the feature map affects the final performance of our algorithm, we combined  $\mathbb{C}_{1,a}$  and  $\mathbb{C}_{2,a}$  as the source domain and took  $\mathbb{C}_{1,f}$  as the target domain for the classification experiment. We mapped the data into different dimensionalities in subspace from 10-dimensional to 100-dimensional, the step size was set to 10 and other parameters were set to the same values as in the second experiment. The experimental results are shown in Figure 2. We observed that the larger the subspace dimension, the higher the classification accuracy. However, the curve of classification accuracy tends to flatten out as the subspace dimension keeps increasing. Compared with the original baseline algorithm, the baseline algorithm combining the SLDA regularization term achieved a higher accuracy in different subspace dimensions, which means that the SLDA regularization term proposed in this paper is robust and stable.

#### 4.3. Object Recognition

Caltech-256 (C, collected by the California Institute of Technology), Amazon (A, images downloaded from amazon.com in October 2020), webcam (W, low resolution images captured by a Web camera) and DSLR (D, high-resolution images captured by a digital SLR camera) 4 datasets domain adaptation (4DA) are the most popular benchmarks in domain adaptation. The number of common categories in the 4 domains is 10, indicating that the number of categories in the 4DA dataset is 10. Each category in each domain has 8 to 151 samples, with a total of 2533 images. Figure 3 shows some samples selected from the 4DA.



**Figure 2.** Comparison of baseline and the SLDA regularized baseline in the subspace of 10–100 dimensionality on the face recognition task. (a) RKHS-DA and RKHS-DA+SLDA. (b) TCA and TCA+SLDA. (c) IGLA and IGLDA+SLDA. (d) TIT and TIT+SLDA.



**Figure 3.** Sample pictures from the four datasets: (a) Caltech-256, (b) Amazon, (c) webcam and (d) DSLR.

For all datasets, we followed [31] to preprocess the data using a similar feature extraction and experimentation protocol. By randomly selecting two different domains as the source and target domains, a total of  $4 \times 3 = 12$  cross-domain object recognition tasks were constructed. In each task, we randomly selected a certain number of samples from each category as the source domain data for the training set.

When D was the source domain, we drew 8 samples from each category; when A, C and W were the source domains, we drew 20 samples from each category. Then, the source domain samples were used as the training set data and the target domain samples were used as the test set.

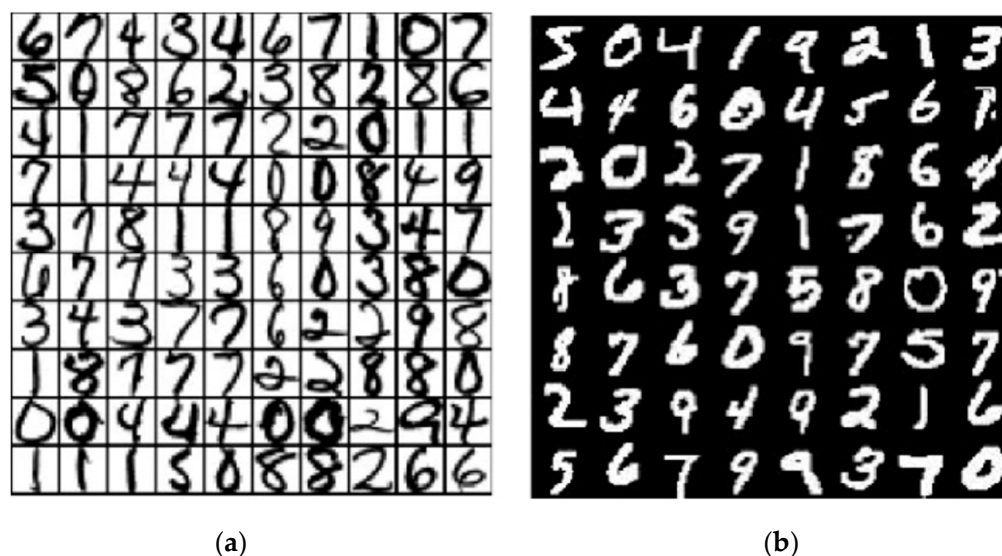
The results of the first experiment are shown in Appendix A, Table A2. Compared with the original space, the geometric center distance between the source and target domains in the subspace is greatly reduced, and the number variance of the source and target domains is greatly increased.

The results of the classification experiments are shown in Appendix A, Table A3. The classification accuracy of the SLDARKHSDA algorithm with the addition of the SLDA regularization term is about 2% higher than that of the KNN and RKHS-DA algorithms.

#### 4.4. Handwritten Numeral Classification

In this section, the USPS+MNIST dataset is used for handwritten digit classification experiments. The USPS dataset consists of 7291 training images and 2007 test images of size  $16 \times 16$ . The MNIST dataset has a training set of 60,000 examples and a test set of 10,000 examples of size  $28 \times 28$ .

The images of both the MNIST and USPS datasets share 10 grayscale images of handwritten Arabic numerals. These images were rescaled to a size of  $16 \times 16$ , which allowed the numbers to be fixed in the center of the entire image and the images to be of the same size. Figure 4 shows an example of MNIST and USPS data sets.



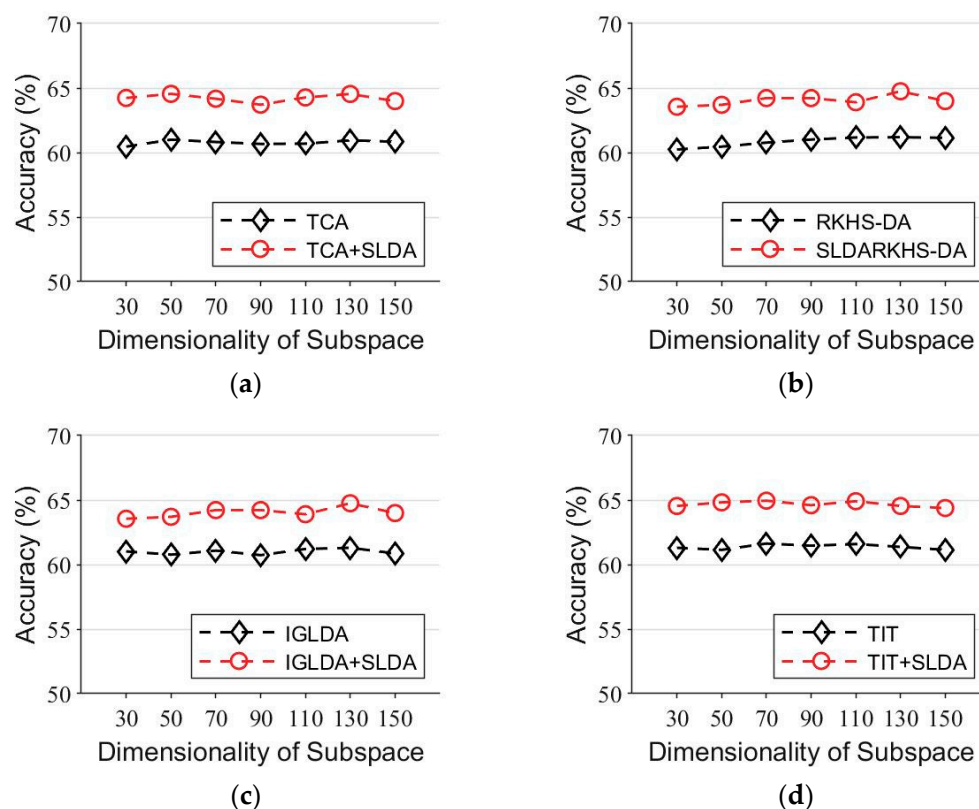
**Figure 4.** Sample pictures from MNIST and USPS. (a) MNIST. (b) USPS.

The experiment in this section was conducted on a subset of MNIST+USPS data set, which consisted of two parts: the first part was 2000 images randomly selected from the MNIST data set, and the second part was 1800 images randomly selected from the USPS data set.

Similarly, all the images in the subset were uniformly resized to  $16 \times 16$  pixels and the gray value of the pixels was used as a feature vector to represent each image. Thus, the samples of MNIST and USPS lie in the same 256-dimensional feature space. To speed up the experiments, we constructed a dataset MINST vs. USPS, randomly selected 50 sets of digital images in MINST, with a total of 500 images to form the source data, and used all the images in USPS to form the target data.

Like we did on the other data sets, in the first experiment, we fixed the dimension of the subspace as 150, and after our algorithm transformation,  $D(S, T)$  was reduced from 2.96 to 1.15,  $\text{Var}(S)$  was changed from 3.6 to 3.2, and  $\text{Var}(T)$  was changed from 4 to 6. Although the  $\text{Var}(S)$  is smaller, which is not what we expected, the ratio of  $\text{Var}(S)/D(S, T)$  is larger, so it still verifies the effectiveness of our algorithm.

In the second experiment, we trained the KNN classifier to repeat the classification experiment 100 times, and used a linear kernel function. The subspace dimensions were set to 30 to 150 and the step size was 20. Figure 5 shows the experimental results. The SLDARKHS-DA algorithm with the SLDA regularization term improves the classification accuracy of RKHS-DA algorithm by about 3%, which is much higher than the classification accuracy of KNN directly (51.18%). Similar results were found for other baseline methods: the accuracy of the baseline algorithm with the SLDA regularization term was higher than that of the original baseline algorithm. In addition, the variation of subspace dimensions had little effect on the classification accuracy of each algorithm.



**Figure 5.** Comparison of the baseline and the SLDA regularized baseline in 30–150 dimensionality of subspace on handwritten numeral classification. (a) RKHS-DA and RKHS-DA+SLDA. (b) TCA and TCA+SLDA. (c) IGLDA and IGLDA+SLDA. (d) TIT and TIT+SLDA.

#### 4.5. Text Categorization

Reuters-21578 dataset (Dai et al., 2007) contains three cross-domain document categorization tasks, *Orgs vs. People*, *Orgs vs. Places* and *People vs. Places*. The notation “orgs vs. Place” indicates that we have the *Org* subtype as the source domain data and the *Place* subtype as the target domain. There are 1237 source documents and 1208 target documents for the task of *Orgs vs. People*, 1016 source documents and 1043 target documents for the task of *Orgs vs. Places* and 1077 source documents and 1077 target documents for the task of *People vs. Places*. We randomly selected 50% of the source domain data as the training set and used all the target domain data as the testing set.

In the first experiment, we set the subspace dimensions from 10 to 50 with a step size of 10, and calculated the variance and the distance between the source domain and the geometric center of the target domain. The experimental results are shown in Appendix A, Table A4.

In the second experiment, we used the KNN classifier to verify the effect of the methods.

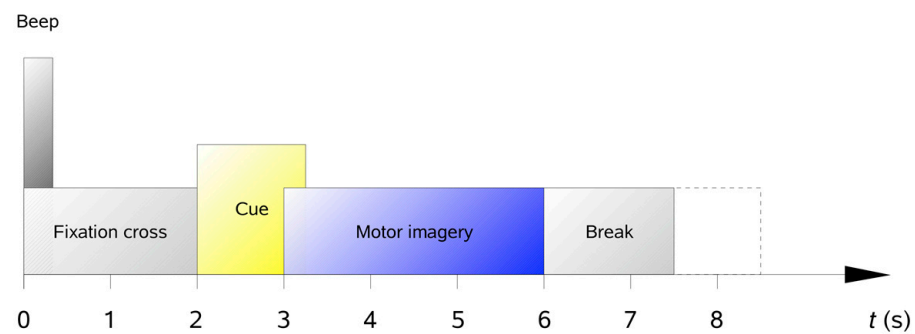
The experimental results are shown in Appendix A, Table A5. In almost all the dimensions and all the experiments, the recognition rate of RKHS-DA was improved to some extent by the SLDA regularization term. In addition, SLDA also improved the classification accuracy of the other baseline methods used for comparison.

#### 4.6. Motor Imagery Classification

As described in Section 3, we used the 2a dataset from the BCI competition IV, which consists of nine subjects [32]. The subjects were sitting in an armchair in front of a computer screen. As shown in Figure 6, at the beginning of the trial ( $t = 0$  s), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds ( $t = 2$  s), a cue appeared and stayed on the screen for 1.25 s. This prompted the subjects to perform the desired motor imagery task (left hand, right hand, both foot and



tongue). No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at  $t = 6$  s.

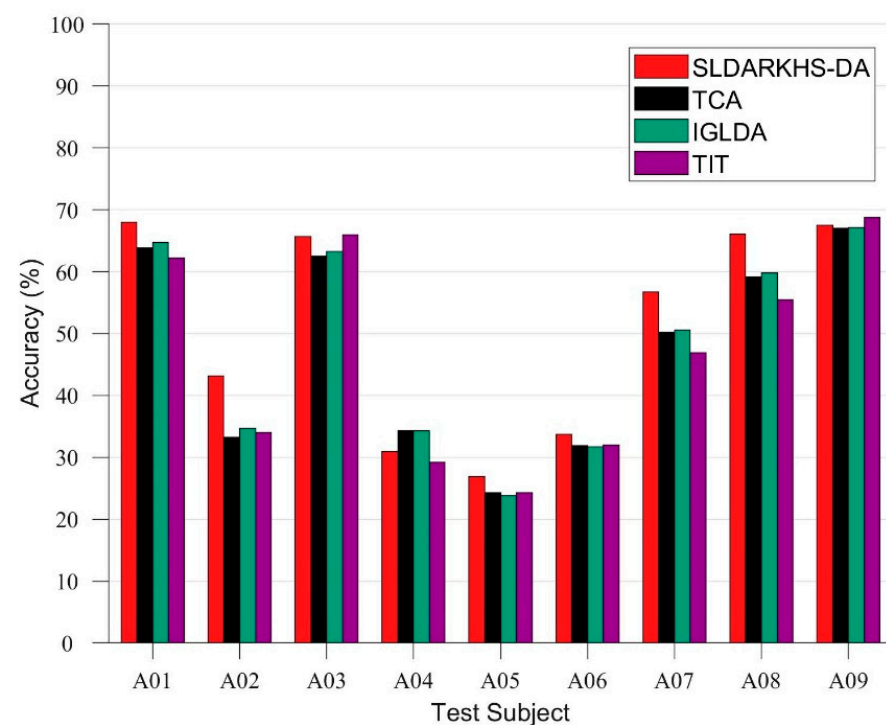


**Figure 6.** Timing scheme of one trial.

For each subject, two periods of data were recorded on two different days, with 288 trials for each period and 72 trajectories for each category. We captured data from 1.5 to 6.5 s for one trial. The recorded EEG signals were sampled with 250 Hz and filtered by a fifth-order Butterworth filter band in the 8–30 Hz frequency band.

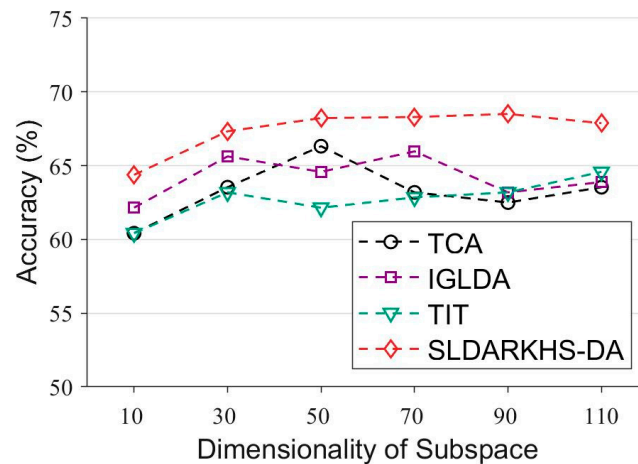
We took A01T to A09T as the source domain and A01E to A09E as the target domain, respectively, while a total of 9 experiments were set up. In this experiment, the PCA algorithm as the baseline method was added for the dimensionality reduction of the original spatial data, and KNN was used as the default classifier of all algorithms. We set the parameter  $\gamma$  of the sparse regularization item to  $10^{-2}$ , and the other parameter settings are shown in Section 4.1.

Since our ultimate goal was to compare the performance of our method with the other baseline methods on the BCI 2a dataset, in the first experiment, we fixed the dimension of our subspace to 25 and performed the classification on different subjects from A01 to A09 for comparison. Figure 7 shows that our method outperforms the baseline algorithm in all experiments, except for the result recorded in A04.



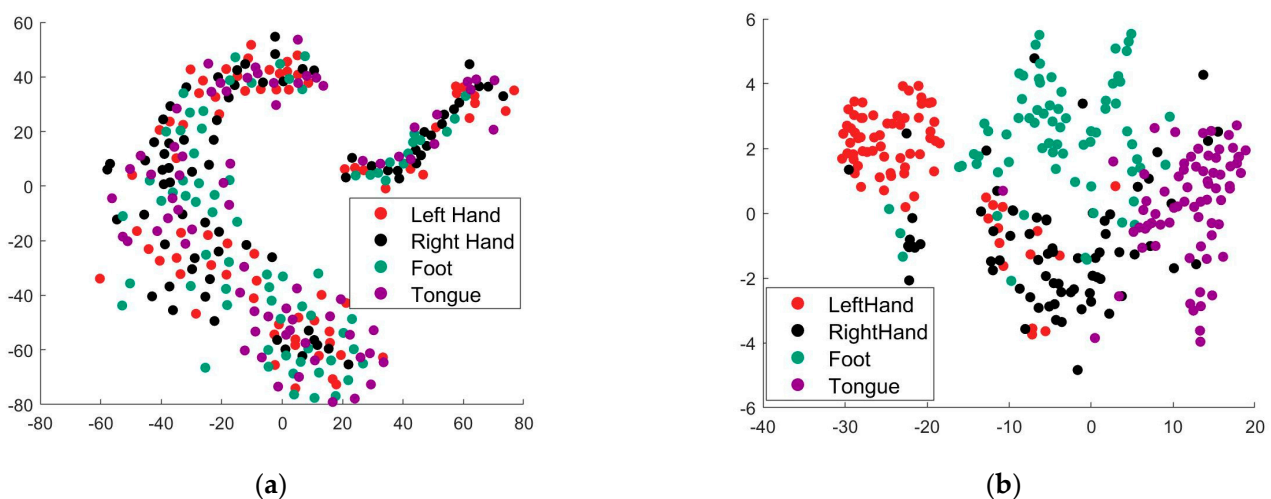
**Figure 7.** Comparison of SLDARKHS-DA(Sparse) and the various baseline methods on different test subjects in the BCI 2a datasets. The classification (by KNN) accuracies (In %) are shown here.

In the second experiment, we compared our method with the baseline method in terms of the dimensionality reduction. We used A01T as the source domain and A01E as the target domain, and the dimensionality of the subspace varies from 10 to 110. As shown in Figure 8, our SLDARKHS-DA (Sparse) outperforms the other baseline methods.



**Figure 8.** Comparison of SLDARKHS-DA (Sparse) and the various baseline methods on BCI 2a datasets in different dimensionalities of the subspace.

In the third experiment, we investigated the impact of our algorithm on the source domain data distribution. For visualization purposes, we applied tSNE to both the original data and the transformed data. Figure 9a shows a two-dimensional representation of the original data vector, i.e., each point in the figure is a representative of a trial. Moreover, Figure 9b shows the representation of the transformed data vector obtained by our SLDARKHS-DA (Sparse). In Figure 9a,b, the points are colored according to the mental task. We observe that the source domain data are chaotic in the original space, while our algorithm separates the four classes of data, which facilitates the accuracy of the KNN classifier.

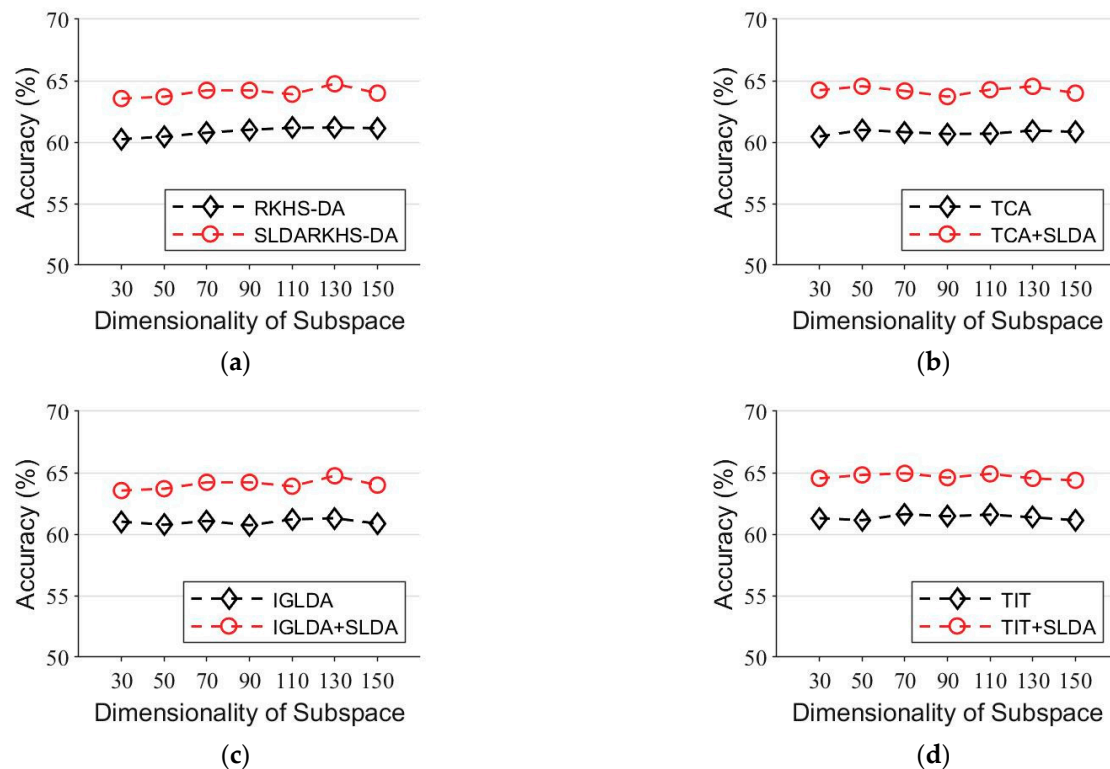


**Figure 9.** Visualization of the feature distribution in source domain by tSNE from subject A09. (a) The original feature distribution. (b) The transformed feature distribution obtained by SLDARKHS-DA (Sparse).

In the fourth experiment, Table A6 in Appendix A shows the classification accuracy of various original baseline algorithms and the baseline algorithm after adding the SLDA regularization term. Firstly, from the perspective of the domain adaptation framework, RKHS-DA and the other baseline algorithms of the domain adaptation are better than

PAC+KNN, and the SLDARKHS-DA (Sparse) is better than the other algorithms. In addition, it can be seen that the SLDA regularization term has certain improvements over the other baseline algorithms used for comparison.

In the fifth experiment, we conducted experiments on subject A01, with A01T as the source domain and A01E as the target domain. We set the range of the subspace dimension from 10 to 110. The average classification results are shown in Figure 10. Based on these results, we observe that the classification performance of the algorithm with regularization SLDA is better than that of the original baseline algorithm in all the dimensions.



**Figure 10.** Comparison of the baseline and the SLDA regularized baseline in face recognition with a subspace dimensionality of 10–100. (a) RKHS-DA and RKHS-DA+SLDA. (b) TCA and TCA+SLDA. (c) IGLDA and IGLDA+SLDA. (d) TIT and TIT+SLDA.

## 5. Conclusions

In this paper, we reorganized the RKHS subspace learning framework based on the theory of RKHS, which consists of functions defined on the original data space instead of the Hilbert space that is independent of the original data space. We first proposed an SLDA regularization term based on the discriminant analysis of the source domain data. The regularization term can increase the inter-class distance and decrease the intra-class distance. Based on the SLDA and RKHS subspace learning framework, we proposed a domain adaptation algorithm. Based on the application of BCI, we selected the most desired data to form the basis of the subspace by adding sparse constraints, i.e.,  $L_{2,1}$  norm. Extensive experiments validated the effectiveness of our algorithm.

In the future, we plan to continue our work by pursuing several avenues. First, SLDARKHS-DA uses parametric kernels for the MMD, and we plan to develop an efficient algorithm for kernel choice in SLDARKHS-DA. Second, to improve the sensitivity of the MI data, we will use the frequency domain features of the MI data. Moreover, we plan to extend SLDARKHS-DA to other BCI experiments with cross-subject settings.

**Author Contributions:** Conceptualization, S.L. and Z.M.; methodology, L.J. and Z.M.; software, L.J.; validation, L.J., W.L. and C.C.; formal analysis, W.L.; investigation, L.J.; resources, L.J.; data curation, L.J.; writing—original draft preparation, L.J.; writing—review and editing, W.L. and C.C.; visualization, C.C.; supervision, S.L. and Z.M.; project administration, S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China grant number 61773022 and Science and Technology Program of Guangzhou grant number 68000-42050001.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Classification accuracy (in %) of face recognition in different tasks (1(a) and 2(a) for source domain).

Task	KNN	RKHS-DA	SLDA RKHS-DA	TCA	TCA + SLDA	IGLDA	IGLDA + SLDA	TIT	TIT + SLDA
1.b	98.00	97.00	97.00	97.00	97.00	99.00	97.00	97.00	99.00
1.c	89.00	89.00	95.00	89.00	94.00	90.00	94.00	88.00	94.00
1.d	57.00	63.00	64.00	63.00	65.00	64.00	65.00	60.00	66.00
1.e	15.00	88.00	91.00	88.00	91.00	90.00	92.00	85.00	90.00
1.f	11.00	74.00	84.00	74.00	88.00	81.00	88.00	75.00	83.00
1.g	3.00	63.00	67.00	63.00	67.00	64.00	67.00	59.00	69.00
1.h	48.00	63.00	64.00	63.00	65.00	68.00	65.00	61.00	61.00
1.i	33.00	57.00	49.00	57.00	54.00	58.00	54.00	52.00	49.00
2.b	96.00	97.00	98.00	97.00	98.00	98.00	98.00	95.00	98.00
2.c	92.00	95.00	99.00	95.00	99.00	95.00	99.00	93.00	99.00
2.d	65.00	69.00	64.00	69.00	68.00	70.00	68.00	67.00	63.00
2.e	21.00	88.00	93.00	88.00	95.00	90.00	95.00	88.00	93.00
2.f	17.00	78.00	83.00	78.00	86.00	79.00	86.00	74.00	82.00
2.g	2.00	48.00	72.00	48.00	73.00	55.00	73.00	43.00	72.00
2.h	38.00	68.00	56.00	68.00	61.00	67.00	61.00	63.00	56.00
2.i	41.00	59.00	54.00	59.00	58.00	61.00	58.00	57.00	49.00
Average	45.38	74.75	76.88	74.75	78.69	76.81	78.75	72.31	76.44

**Table A2.** Data distribution in the original space and subspace for object recognition.

Task	Original Space			Subspace		
	D(S, T)	Var(S)	Var(T)	D(S, T)	Var(S)	Var(T)
A→W	1.665	27.19	27.97	0.0068	1030.64	1046.91
A→C	1.704	27.68	27.58	0.0052	1564.31	1567.77
A→D	1.678	27.78	28.03	0.0097	1017.99	979.06
C→W	1.776	27.40	27.97	0.0081	1054.88	1043.30
C→A	1.826	28.11	27.75	0.0080	1389.24	1411.04
C→D	1.851	27.03	28.03	0.0088	975.87	975.18
D→W	2.225	28.14	27.97	0.0085	1032.31	1000.64
D→A	2.298	27.62	27.75	0.0053	1242.08	1427.02
D→C	2.315	27.97	27.58	0.0049	1359.13	1502.02
W→D	1.257	27.62	28.03	0.0165	932.61	934.83
W→A	1.236	27.73	27.75	0.0129	1215.04	1399.26
W→C	1.283	28.04	27.58	0.0112	1272.64	1456.59

**Table A3.** Classification accuracy (in %) of object recognition in different tasks.

Task	KNN	RKHS-DA	SLDA RKHS-DA	TCA	TCA + SLDA	IGLDA	IGLDA + SLDA	TIT	TIT + SLDA
A→W	29.83	33.14	35.45	34.83	35.92	35.28	35.70	37.13	37.32
A→C	26.00	33.09	34.91	34.83	35.41	34.99	35.01	33.44	33.90
A→D	25.48	28.83	30.25	35.61	36.74	30.76	31.23	35.43	35.79
C→W	25.76	31.82	32.54	32.90	33.10	32.82	33.66	37.07	37.82
C→A	23.70	35.02	36.68	36.61	37.09	36.62	37.26	33.84	33.77
C→D	25.48	32.96	34.69	35.57	36.37	35.09	35.40	39.97	40.78
D→W	63.39	59.44	60.79	61.93	62.13	60.58	60.84	70.42	71.40
D→A	28.49	31.89	33.92	34.03	34.23	33.96	34.05	27.90	28.39
D→C	26.27	31.04	33.15	32.91	33.46	32.57	33.06	26.26	26.61
W→D	59.24	66.93	68.59	68.97	69.82	68.90	69.06	77.34	78.35
W→A	22.96	33.41	35.64	35.71	35.80	35.48	35.86	31.55	32.14
W→C	19.86	28.50	30.65	30.86	31.39	30.61	30.88	27.99	28.56
Average	31.37	37.17	38.94	39.56	40.12	38.97	39.33	39.86	40.40

**Table A4.** Data distribution in the original space and subspace for text categorization.

Task	SD	Original Space			Subspace		
		D(S, T)	Var(S)	Var(T)	D(S, T)	Var(S)	Var(T)
People vs. Places	10	1.96	60.62	64.14	0.005	844.13	1283.17
	20	1.98	59.95	64.14	0.005	1002.98	1763.11
	30	1.96	60.44	64.14	0.010	1346.93	2049.31
	40	1.93	61.05	64.14	0.008	1406.10	2240.79
	50	1.92	61.56	64.14	0.002	2530.73	2463.35
Orgs vs. People	10	1.90	61.65	65.70	0.004	688.33	1190.34
	20	1.85	62.72	65.70	0.005	1023.54	947.86
	30	1.89	63.19	65.70	0.002	2046.41	2210.36
	40	1.89	62.08	65.70	0.002	1299.08	2131.33
	50	1.89	63.01	65.70	0.008	1447.91	1459.49
Orgs vs. Places	10	1.99	61.44	63.41	0.005	690.08	1161.74
	20	1.99	61.18	63.41	0.007	976.39	1591.35
	30	1.98	61.14	63.41	0.002	1878.04	1939.63
	40	2.00	61.37	63.41	0.008	1390.50	2116.67
	50	1.97	61.78	63.41	0.007	1408.70	2276.13

**Table A5.** Classification accuracy (in %) of text categorization in different tasks.

Task	SD	KNN	RKHS-DA	SLDA RKHS-DA	TCA	TCA + SLDA	IGLDA	IGLDA + SLDA	TIT	TIT + SLDA
People vs. Places	10		61.42	61.72	61.30	61.70	62.09	62.72	51.88	57.96
	20		61.33	60.81	60.92	61.57	61.38	61.81	53.62	57.47
	30	45.18	59.20	59.78	59.34	60.35	59.69	60.78	54.55	57.17
	40		57.99	58.15	58.48	58.84	58.57	59.15	54.88	56.25
	50		57.41	57.85	57.44	58.12	57.96	58.85	54.15	57.22
Orgs vs. People	10		76.14	76.97	76.80	77.19	77.14	76.97	58.22	68.09
	20		77.99	78.91	79.16	79.77	79.17	78.91	62.85	72.51
	30	45.32	78.07	79.03	78.45	79.88	79.07	79.03	63.38	74.80
	40		78.57	79.41	79.18	80.21	79.26	79.41	65.56	75.82
	50		78.31	79.07	79.03	79.85	79.13	79.07	66.33	75.69
Orgs vs. Places	10		70.27	69.88	69.77	69.79	70.60	69.88	57.03	61.94
	20		72.37	72.45	72.44	73.45	71.97	72.45	60.91	63.61
	30	54.39	71.84	71.81	71.71	72.38	71.76	71.81	62.69	65.48
	40		71.18	71.24	71.37	71.72	71.38	71.24	63.95	65.80
	50		70.97	71.55	70.89	71.58	71.30	71.55	64.66	65.88
Average		48.30	69.54	69.91	60.75	70.43	70.03	70.24	59.64	65.05

**Table A6.** Classification accuracy (in %) of BCI—motor imagery on different subjects.

Subject	KNN	RKHS-DA	SLDA RKHS-DA(Sparse)	TCA	TCA + SLDA	IGLDA	IGLDA + SLDA	TIT	TIT + SLDA
A01	38.89	64.24	67.98	63.82	68.42	64.69	63.65	62.26	63.53
A02	35.07	31.94	43.19	33.26	44.71	34.69	33.85	34.03	46.48
A03	29.51	61.81	65.72	62.53	67.42	63.33	62.99	65.94	67.03
A04	26.74	33.68	30.97	34.34	32.24	34.31	29.86	29.20	30.89
A05	24.65	24.31	26.90	24.34	26.31	23.78	23.89	24.27	26.24
A06	24.31	31.94	33.78	31.91	34.57	31.74	32.29	31.98	36.58
A07	49.65	50.69	56.76	50.17	57.80	50.56	49.17	46.88	57.35
A08	26.04	57.99	66.03	59.13	66.13	59.83	59.58	55.52	55.96
A09	25.00	64.58	67.56	67.08	68.42	67.12	66.53	68.75	71.03
Average	27.99	46.80	50.99	47.40	51.78	47.78	46.87	46.54	50.56

## References

- Nicolas-Alonso, L.F.; Gomez-Gil, J. Brain Computer Interfaces, a Review. *Sensors* **2012**, *12*, 1211–1279. [\[CrossRef\]](#)
- Nijholt, A.; Bos, D.P.-O.; Reuderink, B. Turning Shortcomings into Challenges: Brain—Computer Interfaces for Games. *Entertain. Comput.* **2009**, *1*, 85–94. [\[CrossRef\]](#)
- Blankertz, B.; Dornhege, G.; Krauledat, M.; Müller, K.-R.; Curio, G. The Non-Invasive Berlin Brain—Computer Interface: Fast Acquisition of Effective Performance in Untrained Subjects. *NeuroImage* **2007**, *37*, 539–550. [\[CrossRef\]](#)
- Al-Ani, T.; Trad, D.; Somerset, V.S. Signal Processing and Classification Approaches for Brain-Computer Interface. *Intell. Biosens.* **2010**, *25*–66. [\[CrossRef\]](#)
- Leeb, R.; Friedman, D.; Müller-Putz, G.R.; Scherer, R.; Slater, M.; Pfurtscheller, G. Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic. *Comput. Intell. Neurosci.* **2007**, *2007*, e79642. [\[CrossRef\]](#)
- Galán, F.; Nuttin, M.; Lew, E.; Ferrez, P.W.; Vanacker, G.; Philips, J.; Millán, J.D. A Brain-Actuated Wheelchair: Asynchronous and Non-Invasive Brain—Computer Interfaces for Continuous Control of Robots. *Clin. Neurophysiol.* **2008**, *119*, 2159–2169. [\[CrossRef\]](#)
- Krepki, R.; Blankertz, B.; Curio, G.; Müller, K.-R. The Berlin Brain-Computer Interface (BBCI)—towards a New Communication Channel for Online Control in Gaming Applications. *Multimed. Tools Appl.* **2007**, *33*, 73–90. [\[CrossRef\]](#)
- Hochberg, L.R.; Bacher, D.; Jarosiewicz, B.; Masse, N.Y.; Simeral, J.D.; Vogel, J.; Haddadin, S.; Liu, J.; Cash, S.S.; van der Smagt, P.; et al. Reach and Grasp by People with Tetraplegia Using a Neurally Controlled Robotic Arm. *Nature* **2012**, *485*, 372–375. [\[CrossRef\]](#)
- Pan, J.; Xie, Q.; He, Y.; Wang, F.; Di, H.; Laureys, S.; Yu, R.; Li, Y. Detecting Awareness in Patients with Disorders of Consciousness Using a Hybrid Brain—Computer Interface. *J. Neural Eng.* **2014**, *11*, 056007. [\[CrossRef\]](#)
- Pfurtscheller, G.; Da Silva, F.L. Event-Related EEG/MEG Synchronization and Desynchronization: Basic Principles. *Clin. Neurophysiol.* **1999**, *110*, 1842–1857. [\[CrossRef\]](#)
- Soleymanpour, R.; Arvaneh, M. Entropy-Based EEG Time Interval Selection for Improving Motor Imagery Classification. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 004034–004039.
- Feng, J.; Yin, E.; Jin, J.; Saab, R.; Daly, I.; Wang, X.; Hu, D.; Cichocki, A. Towards Correlation-Based Time Window Selection Method for Motor Imagery BCIs. *Neural Netw.* **2018**, *102*, 87–95. [\[CrossRef\]](#)
- Thomas, K.P.; Guan, C.; Tong, L.C.; Vinod, A.P. Discriminative FilterBank Selection and EEG Information Fusion for Brain Computer Interface. In Proceedings of the 2009 IEEE International Symposium on Circuits and Systems, Taipei, Taiwan, 24–27 May 2009; pp. 1469–1472.
- Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [\[CrossRef\]](#)
- Long, M.; Wang, J.; Ding, G.; Sun, J.; Yu, P.S. Transfer Feature Learning with Joint Distribution Adaptation. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 2200–2207.
- Zhang, J.; Li, W.; Ogunbona, P. Joint Geometrical and Statistical Alignment for Visual Domain Adaptation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1859–1867.
- Wang, J.; Feng, W.; Chen, Y.; Yu, H.; Huang, M.; Yu, P.S. Visual Domain Adaptation with Manifold Embedded Distribution Alignment. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 402–410.
- Patel, V.M.; Gopalan, R.; Li, R.; Chellappa, R. Visual Domain Adaptation: A Survey of Recent Advances. *IEEE Signal Process. Mag.* **2015**, *32*, 53–69. [\[CrossRef\]](#)
- Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. In *Domain Adaptation in Computer Vision Applications*; Csurka, G., Ed.; Advances in Computer Vision and Pattern Recognition; Springer International Publishing: Cham, Switzerland, 2017; pp. 189–209. ISBN 978-3-319-58346-4.
- Sugiyama, M.; Krauledat, M.; Müller, K.-R. Covariate Shift Adaptation by Importance Weighted Cross Validation. *J. Mach. Learn. Res.* **2007**, *8*, 985–1005.
- Chai, X.; Wang, Q.; Zhao, Y.; Liu, X.; Bai, O.; Li, Y. Unsupervised Domain Adaptation Techniques Based on Auto-Encoder for Non-Stationary EEG-Based Emotion Recognition. *Comput. Biol. Med.* **2016**, *79*, 205–214. [\[CrossRef\]](#)

22. Gretton, A.; Borgwardt, K.; Rasch, M.J.; Schölkopf, B.; Smola, A.J. A Kernel Method for the Two-Sample Problem. *arXiv* **2008**, arXiv:08052368.
23. Smola, A.; Gretton, A.; Song, L.; Schölkopf, B. A Hilbert Space Embedding for Distributions. In *Proceedings of the Algorithmic Learning Theory*; Hutter, M., Seredidio, R.A., Takimoto, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 13–31.
24. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **2011**, *22*, 199–210. [[CrossRef](#)]
25. Jiang, M.; Huang, W.; Huang, Z.; Yen, G.G. Integration of Global and Local Metrics for Domain Adaptation Learning Via Dimensionality Reduction. *IEEE Trans. Cybern.* **2017**, *47*, 38–51. [[CrossRef](#)]
26. Li, J.; Lu, K.; Huang, Z.; Zhu, L.; Shen, H.T. Transfer Independently Together: A Generalized Framework for Domain Adaptation. *IEEE Trans. Cybern.* **2019**, *49*, 2144–2155. [[CrossRef](#)]
27. Lei, W.; Ma, Z.; Liu, S.; Lin, Y. EEG Mental Recognition Based on RKHS Learning and Source Dictionary Regularized RKHS Subspace Learning. *IEEE Access* **2021**, *9*, 150545–150559. [[CrossRef](#)]
28. Larsen, R.J.; Marx, M.L. *Introduction to Mathematical Statistics and Its Applications: Pearson New International Edition PDF EBook*; Pearson Higher Ed: London, UK, 2013; ISBN 1-292-03672-9.
29. Seperability of Four-Class Motor Imagery Data Using Independent Components Analysis—IOPscience. Available online: <https://iopscience.iop.org/article/10.1088/1741-2560/3/3/003/meta> (accessed on 15 September 2021).
30. Martinez, A.; Benavente, R. The AR Face Database: CVC Technical Report #24; June 1998. Available online: <https://www2.ece.ohio-state.edu/~aleix/ARdatabase.html> (accessed on 24 November 2021).
31. Gong, B.; Shi, Y.; Sha, F.; Grauman, K. Geodesic Flow Kernel for Unsupervised Domain Adaptation. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 16–21 June 2012; pp. 2066–2073.
32. Brunner, C.; Leeb, R.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. BCI Competition 2008—Graz Data Set A. *Inst. Knowl. Discov. Lab. Brain-Comput. Interfaces Graz Univ. Technol.* **2008**, *16*, 1–6.