Single-Source to Single-Target Cross-Subject Motor Imagery Classification Based on Multisubdomain Adaptation Network

Yi Chen, Rui Yang, *Member, IEEE*, Mengjie Huang, *Member, IEEE*, Zidong Wang, *Fellow, IEEE*, and Xiaohui Liu

***Abstract*— In the electroencephalography (EEG) based cross-subject motor imagery (MI) classification task, the device and subject problems can cause the time-related data distribution shift problem. In a single-source to single-target (STS) MI classification task, such a shift problem will certainly provoke an increase in the overall data distribution difference between the source and target domains, giving rise to poor classification accuracy. In this paper, a novel multi-subdomain adaptation method (MSDAN) is proposed to solve the shift problem and improve the classification accuracy of the traditional approaches. In the proposed MSDAN, the adaptation losses in both class-related and time-related subdomains (that are divided by different data labels and session labels) are obtained by measuring the distribution differences between the source and target subdomains. Then, the adaptation and classification losses in the loss function of MSDAN are minimized concurrently. To illustrate the application value of the proposed method, our method is applied to solve the STS MI classification task about data analysis with respect to the brain-computer interface (BCI) competition III-IVa dataset. The resultant experiment results demonstrate that compared with other well-known domain adaptation and deep learning methods, the proposed method is capable of solving the time-related data distribution problem at higher classification accuracy.**

***Index Terms*— Electroencephalography classification, motor imagery, multi-subdomain adaptation, single-source to single-target, time-related distribution shift.**

Manuscript received 19 February 2022; revised 17 June 2022; accepted 13 July 2022. Date of publication 18 July 2022; date of current version 22 July 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61603223; in part by the Jiangsu Provincial Qinglan Project 2021; in part by the Suzhou Science and Technology Programme, under Grant SYG202106; in part by the Research Development Fund of Xi'an Jiaotong-Liverpool University (XJTLU) under Grant RDF-18-02-30 and Grant RDF-20-01-18; in part by the Key Program Special Fund in XJTLU under Grant KSF-E-34; and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 20KJB520034. *(Corresponding authors: Rui Yang; Mengjie Huang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University Ethics Committee of Xi'an Jiaotong-Liverpool University under Application No. EXT20-01-07 on March 31, 2020.

Yi Chen is with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China, and also with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, U.K. (e-mail: Yi.Chen1902@student.xjtlu.edu.cn).

Rui Yang is with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China, and also with the Research Institute of Big Data Analytics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: r.yang@xjtlu.edu.cn).

Mengjie Huang is with the Design School, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: mengjie.huang@xjtlu.edu.cn).

Zidong Wang and Xiaohui Liu are with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, U.K. (e-mail: zidong.wang@brunel.ac.uk; xiaohui.liu@brunel.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2022.3191869

I. Introduction

NUMEROUS neurons are constantly creating bioelectricity as the brain works in the human brain, which can change the electrical potentials on the scalp surface, which can change the electrical potentials on the human scalp surface [7]. Brain-computer interface (BCI) devices can be used to collect the electrical potentials change across time and obtain Electroencephalography (EEG) [58]. EEG is an efficient data acquisition method that is widely adopted in the field of brain science [7], [58], where the analysis of EEG signals is instrumental in understanding different brain activities [7], [59]. For example, EEG can be used to test the concentration level [30], evaluate sleep quality [68], and diagnose diseases including seizures [1] and schizophrenia [48]. It should be noted that the EEG-based motor imagery (MI) signal can be used as commands to control external systems [5], and the research on EEG-based MI classification has thus attracted much attention in recent years.

In recent MI classification studies, cross-subject classification has become one of the most popular directions where the generalization ability (of the classification model) can be enhanced by improving the cross-subject performance [56]. In cross-subject classification tasks, classical transfer learning methods such as transfer component analysis [39] and joint distribution adaptation [33] are widely applied in many situations. Recently, some new deep learning algorithms have attracted much attention, such as graph convolutional [51] networks and federated learning [22]. Depending on the number of subjects in the source and target domains, cross-subject MI classification tasks can be divided into three categories: multi-source to multi-target (MTM), multi-source to single-target (MTS), and single-source to single-target (STS). In the MTM MI classification task, the EEG data from several subjects are pooled together for training and testing, where the training set includes the labeled data from every subject.

TABLE I
SUMMARY OF CROSS-SUBJECT MI CLASSIFICATION METHODS

	MTM	MTS	STS
Deep learning w/o domain adaptation	[13], [22], [42], [51], [67]	[2], [17], [29], [40], [41], [60]	Nil
Global domain adaptation	[38]	[3], [12], [18], [24], [61], [65]	[44], [69]
Class-related subdomain adaptation	Nil	[66]	[16]
Time-related subdomain adaptation	Nil	Nil	Nil

In order to improve the model generalization in the MTM MI classification task, the deep learning [13], [22], [42], [51], [67] and global domain adaptation [38] methods have been applied by researchers. In the MTS MI classification task, the data from multiple source subjects are used for training, and the data from a new target subject are used for testing. The training data are fully labeled, and the test data have no label or are partially labeled. In order to solve the MTS problem, various approaches have been developed by researchers such as deep learning [2], [17], [29], [40], [41], [60], global domain adaptation [3], [12], [18], [24], [61], [65], and class-related subdomain adaptation [66] methods. In the STS MI classification task, only the data from a single source subject are used for training, and the model is tested on a new target subject. Global domain adaptation [44], [69] and class-related subdomain adaptation [16] have been introduced to improve classification performance. Some typical cross-subject MI classification methods in these three categories are listed in Table I.

On one hand, it should be noted that so far, in comparison with the MTM and MTS tasks, very little attention has been paid to the STS MI classification task, where only the data of a single source subject are available for training. Undoubtedly, the shortage of training data could increase the difficulty of model generalization. On the other hand, as shown in Table I, the existing domain adaptation studies in the cross-subject MI classification tasks have put their focus on the global or class-related approaches, whereas the time-related data distribution problem has not been investigated, not to mention the case where the time-related subdomain adaptation is also concerned.

The time-related data distribution shift problem is a common practical problem in EEG-related studies [4], [21], [23], [64]. It has been shown in a recent study that the classification accuracies varied from 7% to 30% across different experimental sessions for the same subject, and this indicates the existence of a significant EEG data distribution shift across time [43]. Moreover, it has been pointed out that the classification accuracy of this study has dropped by 18.2% when combining the data from two sessions together [43]. Another study showed that the classification accuracy dropped by 52.76% when the model training on data from one session was tested on data from another session [37]. Other previous studies have also stated that the time-related distribution shift problem is a critical factor and should be considered in the EEG analysis [6], [26]. According to the literature review, the time-related EEG data distribution shift can be associated with the following two factors:

1) *EEG device problems.* In typical noninvasive EEG-based MI experiments, electrodes on the scalp surface are used to collect the signals from the human brain when subjects are carrying out specific actions [36]. However, the connection quality between electrodes and scalp surface changes over time, and this dramatically affects the process of data collection. Generally, there are two kinds of electrodes that are frequently used in EEG signal collection devices: dry electrodes and wet electrodes [57]. For dry electrodes, although the experiment setup is convenient, the contact between electrodes and scalp surface may become unstable over time which could lead to the data distribution shift problem [47]. For wet electrodes (e.g., saline electrodes), the evaporation of electrolytes can decrease the connectivity between the electrodes and the human scalp, leading to an increase in impedance and data distribution shift over time [35]. As a result, the signal quality changes across time, giving rise to the time-related EEG data distribution shift problem.

2) *Subject problems.* The subject's emotion may change over the experiment, which would lead to the change of frontal EEG activity related to the intensity [45]. Some studies have shown that many emotional states can cause the change of several frequency bands, such as the beta power band change in the parietal lobe region [19], [45], [46]. Moreover, muscle fatigue during a long-time experiment could contribute to the activity changing of different hemispheres in beta and gamma rhythms [52]. The past studies have also verified that different observations of alpha and theta power are accompanied by mental fatigue increasing [8], [9], [27]. Therefore, it can be concluded that the emotion and fatigue changing of subjects in experiments can lead to the time-related EEG data distribution shift.

In general, the data distribution shift cannot be eliminated in experiments due to the equipment problems and the physiological state of human beings. Hence, there is an urgent need to reduce the time-related data distribution difference by improving the existing STS MI classification algorithm. To the best of the authors' knowledge, no previous work has focused on the time-related data distribution shift problem and the time-related subdomain adaptation in the STS MI classification tasks.

To solve the problems mentioned above, a novel multi-subdomain adaptation network (MSDAN) that combines both class-related and time-related subdomain adaptations is proposed in this paper. The main challenges in this study are: 1) both time-related and class-related subdomains are required to be adapted concurrently, which brings the difficulty in the loss function design for multi-objective optimization; and 2) the STS MI classification tasks have much less training data from the source domain compared with the MTM and MTS tasks, which increases the difficulty of model generalization.

Motivated by the above discussions, the main contributions of this paper can be summarized in three aspects as follows:

1) A novel time-related EEG distribution shift problem in the STS MI classification tasks is investigated in this paper.

2) A novel MSDAN method is introduced where the multi-subdomain adaptation terms help to minimize both time-related and class-related subdomain distribution differences, and moreover, find a proper balance between the adaptation and classification losses.
3) The proposed MSDAN method is successfully verified in the analysis of the STS MI classification tasks for the BCI competition III-IVa dataset, where the classification accuracies are proved to be higher than other well-known domain adaptation methods, and moreover, the efficiency of the MI classification model building in BCI applications is also improved based on the STS cross-subject transfer.

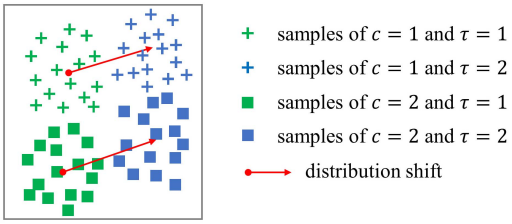
The rest of this paper is presented as follows. The problem of distribution shift in the EEG-based STS MI classification task is described in Section II. The proposed method with data preprocessing and MSDAN is presented in Section III. In Section IV, the dataset and experiments used in this paper are introduced, and the experimental results are analyzed in detail. Finally, conclusions and discussions on relevant future work are presented in Section V.

II. Problem Formulation

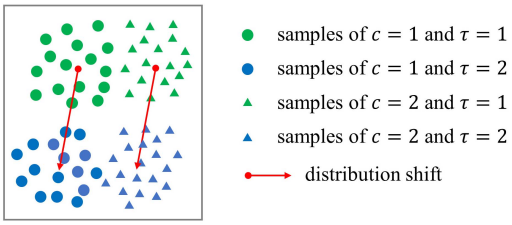
To illustrate the time-related EEG data distribution shift, Fig. 1 shows the changes in data samples with two classes ($c = 1$ or 2) and two experimental sessions ($\tau = 1$ or 2). The green and blue samples represent data distributions in different sessions. The red arrows represent the directions of data distribution shifts between two sessions. Fig. 1(a) shows the distribution shift of source domain data between two sessions, and the cross and square symbols represent different classes in the source domain. Fig. 1(b) shows the distribution shift of target domain data between two sessions, and the circle and triangle symbols represent different classes in the target domain. Due to the EEG device and subject problems, the data distribution of blue samples in the second session can differ from green samples in the first session. This time-related data distribution shift would increase the original distribution difference between the source and target domains, eventually affecting the STS MI classification task.

To further clarify the time-related EEG data distribution shift problem in practical application, the brain topographies of 22 components after independent component analysis (ICA) are shown in Fig. 2. The EEG signals are obtained from the same MI task of one subject at two different experimental sessions. The deep red and deep blue regions represent the high concentration power of positive and negative potential in components. The light color regions represent the low concentration power in components. The differences between Fig. 2(a) and Fig. 2(b) represent the time-related EEG data distribution shifts between two sessions in the same MI task.

From the illustrations of Fig. 1 and Fig. 2, it is shown that the time-related distribution shift problems in EEG signals could have a negative impact on cross-subject MI classification performance. The problem formulation is shown as follows. The original EEG signal measurement can be denoted as



(a) Source data distribution shift over time



(b) Target data distribution shift over time

Fig. 1. Illustration of time-related EEG data distribution shift.

$\hat{S} = \{\hat{s}_i\}$ that contains n samples:

$$\hat{s}_i = \hat{f}_i + \hat{v}_i, i \in n \tag{1}$$

where $\hat{f}_i$ represents the proper data of EEG signal, and $\hat{v}_i$ is the noise.

As the data distribution shifts over time, an example of the actual EEG signal measurement $S = \{s_i\}$, which contains distribution shifts, can be defined as:

$$s_i = \gamma_i \hat{f}_i + \omega_i + \hat{v}_i \tag{2}$$

where ω_i and γ_i represent the additive and multiplicative time-related distribution shifts, respectively.

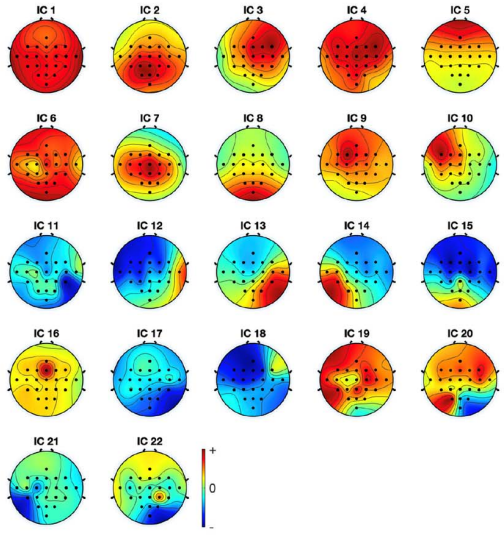
Both the source and target domain data have multiple sessions in the STS MI classification tasks. The EEG signals collected in different sessions may suffer from the illustrated time-related distribution shifts ω_i and γ_i. The time-related distribution shifts would increase the overall distribution difference between the source and target domains and increase the difficulty of the STS MI classification tasks. The problem of EEG data with time-related distribution shifts in STS MI classification tasks is investigated in this paper. A specific method needs to be proposed to solve this time-related distribution shift problem.

III. Methodology

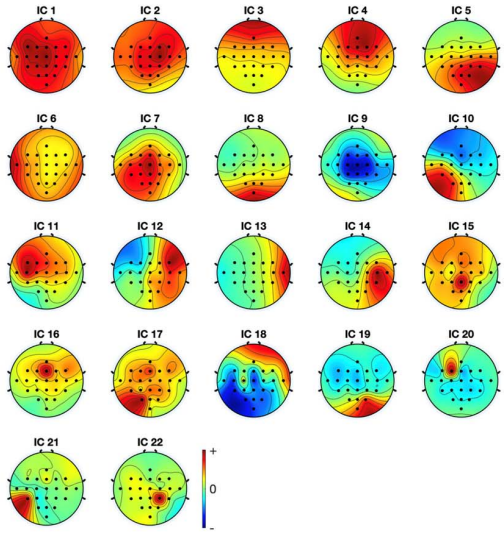
The structure of the proposed method is shown in Fig. 3. EEG data are preprocessed by signal filtering and artifacts removal first, which can increase the signal-to-noise ratio (SNR) and help improve the classification accuracy. Then, data features are extracted from the preprocessed data and classified with multi-subdomain adaptations by the MSDAN. In MSDAN, a notable loss function is designed by combining the classification loss and the subdomain adaptation losses and later optimized during the training process.

A. EEG Data Preprocessing

The preprocessing is shown in the grey part of Fig. 3. Preprocessing is essential due to redundant information in the



(a) Decomposed EEG signals by ICA for the first session



(b) Decomposed EEG signals by ICA for the second session

Fig. 2. Component maps of EEG signals after ICA.

input EEG data, such as high-frequency noise and artifacts generated by the body movement. In this study, the EEG data preprocessing contains signal filtering and artifacts removal. First, the EEG signals are bandpass filtered from 5 to 35 Hz by finite impulse response (FIR) to focus on the wave band that reflects the motor imagery. Second, EEG data are extracted for each sample around cues, reducing the input data size and speeding up the training process. Third, the ICA is implemented to remove the artifact components so the SNR of the data can be further improved. Last, the signals are bandpass filtered and extracted again to refine the data further.

The FIR filter is adopted to obtain the signal in the desired frequency range as shown in (3):

$$g(m) = \sum_{n=0}^{k-1} h(n)s(m-n) \tag{3}$$

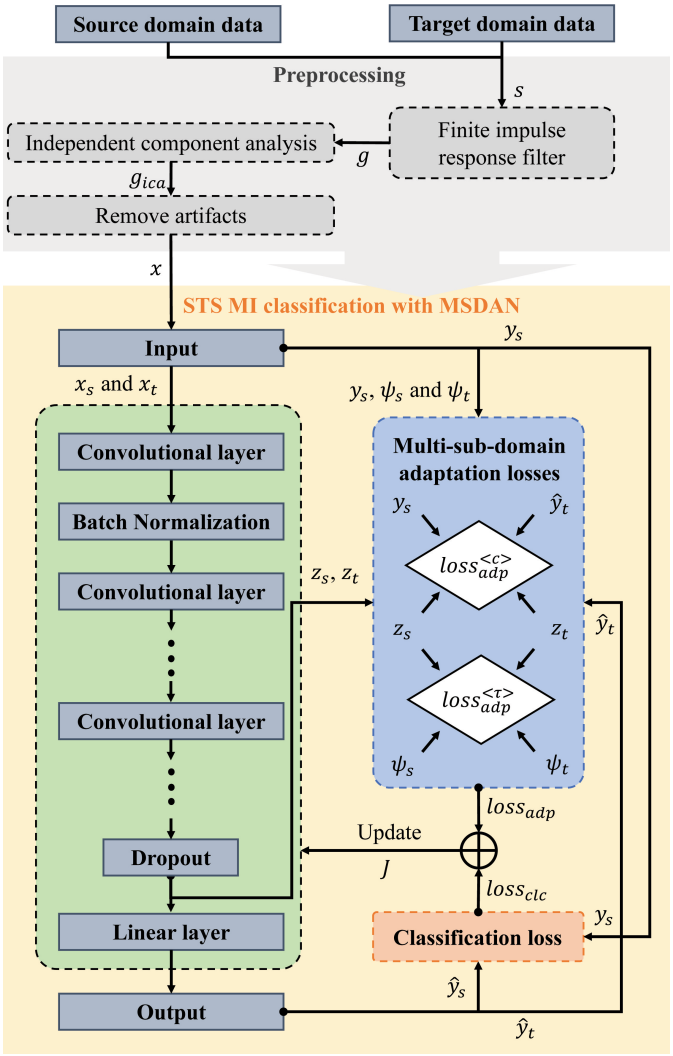


Fig. 3. Structure of the proposed method.

where $s(m)$ is the input signal of time step m, $g(m)$ is the output signal after filtering, $h(n)$ is the FIR filtering parameter, and $0 < k < m$.

ICA can be implemented to remove artifacts, including the electromyography (EMG) caused by the muscle of head and body, the electrocardiogram (ECG) caused by the heart, and the electrooculogram (EOG) caused by the eyeballs, ocular muscles, and eyelids. The signal matrix $g_{ica}(m)$ after ICA can be expressed by (4):

$$g_{ica}(m) = A \cdot g(m) \tag{4}$$

where A is the demixing matrix. Through ICA, the EEG signal can be decomposed into μ independent components. The decomposed matrix is $g_{ica}(m) = [g_1(m), \cdots, g_\mu(m)]$. The preprocessed data x are obtained after the removal of artifact components from $g_{ica}(m)$.

B. Multi-Subdomain Adaptation Method

The proposed MSDAN is shown in the yellow part of Fig. 3. The network has 13 layers, and the detailed architecture is shown in Table II. The green part of the network is inspired

TABLE II
DETAILS OF THE NETWORK STRUCTURE

No	Layer	Options[a]
	Input EEG	N.A.
1	Convolutional layer	kernel size=32×1, filters=8, strides=(1,1), padding='same', no bias
2	Batch normalization	epsilon=1e-05, momentum=0.1
3	Convolutional layer	kernel size=1×μ[b], filters=16, strides=(1,1), padding='valid', no bias
4	Batch normalization	epsilon=1e-05, momentum=0.1
5	ReLU	N.A.
6	Mean pooling	kernel size=4×1, strides=(4,1)
7	Dropout	ρ_{drop}=0.25
8	Convolutional layer	kernel size=16×1, filters=16, strides=(1,1), padding='same', no bias
9	Batch normalization	epsilon=1e-05, momentum=0.1
10	ReLU	N.A.
11	Mean pooling	kernel size=8×1, strides=(8,1)
12	Dropout	ρ_{drop}=0.25
13	Linear layer	N.A.
	Output	N.A.

[a] N.A. represents not applicable, 'valid' denotes no padding, and 'same' denotes that the input and output data matrixes have the same size.

[b] μ represents the number of channels of the input EEG data.

by EEGNet [25]. By features extraction and network training, the classification results can be obtained by:

$$\hat{y}_s = F(x_s) \tag{5}$$
$$\hat{y}_t = F(x_t) \tag{6}$$

where x_s and x_t are the preprocessed data of source and target domains, F denotes the network mapping, and $\hat{y}_s$ and $\hat{y}_t$ denote the predicted class labels of source and target domains, respectively.

The core difference between the proposed method and other methods is the design of the loss function. As illustrated in the blue and orange parts of Fig. 3, the objective function J of MSDAN consists of the classification loss $loss_{clc}$ and the multi-subdomain adaptation loss $loss_{adp}$. The objective function of MSDAN is expressed as follows:

$$\min J = loss_{clc} + loss_{adp} \tag{7}$$

where $loss_{clc}$ denotes the classification loss measuring the difference between the predicted class labels and actual class labels of the source domain, and $loss_{adp}$ denotes the multi-subdomain adaptation loss measuring the distribution difference across different subdomains.

The classification loss $loss_{clc}$ in (7) can be computed by:

$$\begin{aligned} loss_{clc} &= \frac{1}{n_s}\sum_{i=1}^{n_s} J(\hat{y}_{s(i)}, y_{s(i)}) \\ &= -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{c=1}^{C} \mathbb{1}_{\{y_{s(i)}=c\}} \log \hat{y}_{s(i)} \end{aligned} \tag{8}$$

where n_s denotes the number of samples of the source domain, y_s denotes the actual class labels of the source domain, $J(\cdot,\cdot)$ denotes the cross-entropy loss function, C denotes the number of classes, and $\mathbb{1}$ is an indicator function which equals 1 if the condition $y_{s(i)} = c$ $(c \in C)$ is satisfied or 0 if not.

The domain adaptation loss $loss_{adp}$ of MSDAN, as shown in the blue part of Fig. 3, has two parts: the class-related distribution difference $loss_{adp}^{<c>}$ and time-related distribution difference $loss_{adp}^{<\tau>}$:

$$\begin{aligned} loss_{adp} &= \lambda^{<c>} loss_{adp}^{<c>} + \lambda^{<\tau>} loss_{adp}^{<\tau>} \\ &= \lambda^{<c>} \hat{d}(p_s^{<c>}, p_t^{<c>}) \\ &\quad + \lambda^{<\tau>} \hat{d}(p_s^{<\tau>}, p_t^{<\tau>}) \end{aligned} \tag{9}$$

where $<c>$ and $<\tau>$ denote the class-related and the time-related subdomains, respectively; $\lambda^{<c>}$ and $\lambda^{<\tau>}$ are the positive scale factors; $p_s^{<c>}$ and $p_s^{<\tau>}$ represent the class-related and time-related subdomain distributions of the source domain data, respectively; $p_t^{<c>}$ and $p_t^{<\tau>}$ represent the class-related and time-related subdomain distributions of the target domain data, respectively; and $\hat{d}(\cdot,\cdot)$ represents the distribution difference between the source and target domains. The scale factors $\lambda^{<c>}$ and $\lambda^{<\tau>}$ can provide the trade-off among classification loss, class-related adaptation loss, and time-related adaptation loss. Therefore, the different focus of optimization can be shown at different stages of the training process.

The adaptation losses can be obtained with six inputs: actual class labels of source domain y_s, predicted class labels of target domain $\hat{y}_t$, session labels of source domain ψ_s, session labels of target domain ψ_t, and features z_s and z_t extracted from the source and target domains, respectively. Features z_s and z_t can be obtained before the linear layer. Session labels can be obtained based on the data collection order. Here, the maximum mean discrepancy (MMD) [70] is used to measure the distribution difference between the source and target subdomains. In this case, the class-related subdomain difference $\hat{d}(p_s^{<c>}, p_t^{<c>})$ and time-related subdomain difference $\hat{d}(p_s^{<\tau>}, p_t^{<\tau>})$ in (9) can be computed by (10) and (11) as follows:

$$\begin{aligned} &\hat{d}(p_s^{<c>}, p_t^{<c>}) \\ &\quad \triangleq E^{<c>} \| E_s^{<c>}[\phi(z_s)] - E_t^{<c>}[\phi(z_t)] \|_{\mathcal{H}}^2 \\ &\quad = \frac{1}{C}\sum_{c=1}^{C} \| \sum_{i=1}^{n_s} w_{s(i)}^c \phi(z_{s(i)}) - \sum_{j=1}^{n_t} w_{t(j)}^c \phi(z_{t(j)}) \|_{\mathcal{H}}^2 \\ &\quad = \frac{1}{C}\sum_{c=1}^{C} [\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_{s(i)}^c w_{s(j)}^c k(z_{s(i)}, z_{s(j)}) \\ &\qquad + \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} w_{t(i)}^c w_{t(j)}^c k(z_{t(i)}, z_{t(j)}) \\ &\qquad - 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} w_{s(i)}^c w_{t(j)}^c k(z_{s(i)}, z_{t(j)})] \end{aligned} \tag{10}$$

$$\begin{aligned} &\hat{d}(p_s^{<\tau>}, p_t^{<\tau>}) \\ &\quad \triangleq E^{<\tau>} \| E_s^{<\tau>}[\phi(z_s)] - E_t^{<\tau>}[\phi(z_t)] \|_{\mathcal{H}}^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{\tau=1}^{T}\|\sum_{i=1}^{n_s} w_{s(i)}^{\tau}\phi(z_{s(i)}) - \sum_{j=1}^{n_t} w_{t(j)}^{\tau}\phi(z_{t(j)})\|_{\mathcal{H}}^2 \\
&= \frac{1}{T}\sum_{\tau=1}^{T}[\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_{s(i)}^{\tau} w_{s(j)}^{\tau} k(z_{s(i)}, z_{s(j)}) \\
&\quad + \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} w_{t(i)}^{\tau} w_{t(j)}^{\tau} k(z_{t(i)}, z_{t(j)}) \\
&\quad - 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} w_{s(i)}^{\tau} w_{t(j)}^{\tau} k(z_{s(i)}, z_{t(j)})] \qquad (11)
\end{aligned}
$$

where $\mathcal{H}$ represents the reproducing kernel Hilbert space (RKHS) [49], $E[\cdot]$ denotes the expectation; $\phi(\cdot)$ is the mapping function that maps original features to RKHS; $k(z_s, z_t) = \langle\phi(z_s), \phi(z_t)\rangle$ represents the inner product of $\phi(z_s)$ and $\phi(z_t)$; T is the number of experimental sessions; n_s and n_t are the number of samples from the source and target domains, respectively; and $w_{s(i)}^{c}$, $w_{t(j)}^{c}$, $w_{s(i)}^{\tau}$, and $w_{t(j)}^{\tau}$ are the weights denoting the probability that a sample belongs to a specific subdomain.

The Gaussian kernel function $k(z_s, z_t) = e^{-\|z_s - z_t\|^2/\sigma}$ is used in the kernel mean embeddings to estimate the MMD, and the kernel bandwidth σ is set as the median squared distances between the input instances [15]. The sum of the weights satisfy $\sum_{i=1}^{n_s} w_{s(i)}^{c} = 1$, $\sum_{j=1}^{n_t} w_{t(j)}^{c} = 1$, $\sum_{i=1}^{n_s} w_{s(i)}^{\tau} = 1$, and $\sum_{j=1}^{n_t} w_{t(j)}^{\tau} = 1$ with

$$
w_{s(i)}^{c} = \begin{cases} \frac{1}{n_s^c}, & y_{s(i)} = c \\ 0, & y_{s(i)} \neq c \end{cases} \qquad (12)
$$

$$
w_{t(j)}^{c} = \begin{cases} \frac{1}{n_t^c}, & \hat{y}_{t(j)} = c \\ 0, & \hat{y}_{t(j)} \neq c \end{cases} \qquad (13)
$$

$$
w_{s(i)}^{\tau} = \begin{cases} \frac{1}{n_s^{\tau}}, & \psi_{s(i)} = \tau \\ 0, & \psi_{s(i)} \neq \tau \end{cases} \qquad (14)
$$

$$
w_{t(j)}^{\tau} = \begin{cases} \frac{1}{n_t^{\tau}}, & \psi_{t(j)} = \tau \\ 0, & \psi_{t(j)} \neq \tau \end{cases} \qquad (15)
$$

where n_s^c (n_s^{τ}) and n_t^c (n_t^{τ}) are the number of samples in the c-th (τ-th) class in the source and target domains, respectively.

After the combination of classification loss $loss_{clc}$ and adaptation loss $loss_{adp}$, the objective function J is optimized by the standard stochastic gradient (SGD) algorithm, and the network's parameter Θ is updated as follows:

$$
\Theta \leftarrow \Theta - \eta\nabla_{\Theta} \qquad (16)
$$

where η is the learning rate optimized by the grid search. ∇_{Θ} can be computed by:

$$
\begin{aligned}
\nabla_{\Theta} &= \frac{\partial loss_{clc}}{\partial\Theta} + \frac{\partial loss_{adp}}{\partial\Theta} \\
&= \frac{\partial loss_{clc}}{\partial\Theta} + \lambda^{<c>}\frac{\partial loss_{adp}^{<c>}}{\partial\Theta} + \lambda^{<\tau>}\frac{\partial loss_{adp}^{<\tau>}}{\partial\Theta} \\
&= \frac{1}{n_b}\sum_{i=1}^{n_b}\frac{\partial J(\hat{y}_{s(i)}, y_{s(i)})}{\partial\Theta} \\
&\quad + \lambda^{<c>}\frac{\partial E[\hat{d}(p_s^{<c>}, p_t^{<c>})]}{\partial\Theta} \\
&\quad + \lambda^{<\tau>}\frac{\partial E[\hat{d}(p_s^{<\tau>}, p_t^{<\tau>})]}{\partial\Theta} \qquad (17)
\end{aligned}
$$

where n_b denotes the batch size. Since the class-related and time-related subdomains both have an important impact on the classification task, the $\lambda^{<c>}$ and $\lambda^{<\tau>}$ are set to the same value in this paper.

The pseudocode of the proposed MSDAN is shown in Algorithm 1:

Algorithm 1 MSDAN

Input:
Data of source domain: $\{(x_{s(i)}, y_{s(i)}, \psi_{s(i)})\}_{i=1}^{n_s}$
Data of target domain: $\{(x_{t(j)}, \psi_{t(j)})\}_{j=1}^{n_t}$
Output:
Data labels of target domain: $\{\hat{y}_{t(j)}\}_{j=1}^{n_t}$

Initialize the network
for each *epoch* **do**
  for each *batch* **do**
    1. Extract features and obtain predicted data labels $\hat{y}_s$ and $\hat{y}_t$ by (5)-(6);
    2. Compute classification loss $loss_{clc}$ by (8);
    3. Obtain the features z_s and z_t from the network;
    4. Compute adaptation losses $loss_{adp}^{<c>}$ and $loss_{adp}^{<\tau>}$ by (9), (10) and (11) with y_s, $\hat{y}_t$, ψ_s, ψ_t, z_s, and z_t;
    5. Obtain the objective function J by combing $loss_{clc}$ and $loss_{adp}$;
    6. Compute gradient ∇_{Θ} and update network parameters Θ by (16) and (17).
  end for
  Record the best classification result $\{\hat{y}_{t(j)}\}_{j=1}^{n_t}$
end for
return Classification result $\{\hat{y}_{t(j)}\}_{j=1}^{n_t}$

Now, the proposed method has been formulated for solving the time-related data distribution shift problem in the STS MI classification task, where 1) the EEG data are preprocessed by the FIR filter and ICA; 2) the proposed MSDAN is adopted for features extraction and classification; and 3) the classification loss and two subdomain adaptation losses are optimized concurrently. Note that the network parameters are updated during the model training accordingly.

IV. Experiment and Result Analysis

In this section, in order to verify the performance of the proposed method, experiments are conducted on an open access MI dataset: brain-computer interface (BCI) competition III-IVa dataset [11]. In the experiments, different individual subjects are used as the source and target domains, and the performance of the proposed method is compared with that of some typical deep learning methods, global domain adaptation methods, and subdomain adaptation methods. This research has been approved by University Ethics Committee of Xi'an Jiaotong-Liverpool University with proposal number EXT20-01-07 on March 31 2020.

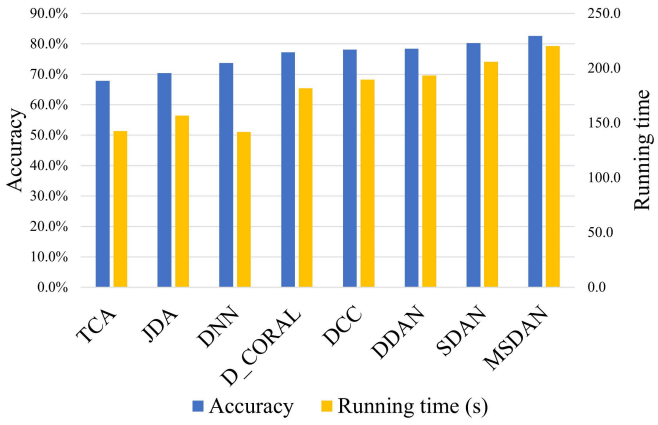


Fig. 4. Comparison of algorithm performance.

A. Dataset Description

The EEG signals of the BCI competition III-IVa dataset contain 118 channels which are measured at positions of the extended international 10/20 system. Signals are sampled at 100 Hz with a bandpass frequency filter that ranges from 0.05 to 200 Hz. The dataset includes signals from five healthy subjects: aa, al, av, aw, and ay. For each subject, there are 280 EEG data samples of two classes: class 1 for the right hand MI and class 2 for the right foot MI. In the EEG signal collection of the dataset, subjects should conduct one kind of motor imagery after each visual cue. Then subjects can relax for 1.75 to 2.25 seconds after 3.5 seconds of motor imagery.

The EEG data obtained from the BCI competition III-IVa dataset were preprocessed in the following sequence: firstly, the EEG signals were filtered from 5 to 35 Hz to focus on the wave band reflecting the activity of motor imagery; secondly, to reduce the input data size and speed up the training process, data segments of 5 seconds containing most of motor imagery information were extracted from each sample from 1 second before cue to 3 seconds after cue; thirdly, the ICA was implemented to remove the artifact components and improve the SNR of the data; lastly, the data is band filtered from 8 to 30 Hz to focus on the core wave band of motor imagery.

B. Experiment Settings

Bidirectional cross-subject experiments are implemented for every two of the five subjects in the dataset to test the performance of the proposed MSDAN in solving EEG-based STS MI classification tasks. In total, there are $A_5^2 = 20$ STS MI classification tasks in the experiment. For the proposed MSDAN, the scale factors of the subdomain adaptation losses $\lambda^{<c>}$ and $\lambda^{<\tau>}$ are both updated from 0 to 1 through a progressive procedure during the training process where $\lambda^{<c>} = \lambda^{<\tau>} = 2/(1 + e^{-10*\xi}) - 1$ and ξ changes linearly from 0 to 1. At the beginning of the training, the optimization focused on classification loss. During the training, the adaption loss was increased to consider the influence of distribution differences more.

The proposed MSDAN is compared with the following methods: transfer component analysis (TCA) [39], joint distribution adaptation (JDA) [33], deep neural network without domain adaptation (DNN), deep correlation alignment (D_CORAL) [50], deep domain confusion (DDC) [54], deep subdomain adaptation network (DDAN) [16], and subdomain adaptation network (SDAN). The DNN is used as an ablation study, where the loss function only contains the classification loss. The D_CORAL and DDC contain the classification loss and global deep domain adaptation loss in their functions. The DDAN and SDAN are used as ablation studies containing classification loss and class-related subdomain adaptation loss in the loss function. The algorithms adopted for comparison in this paper are introduced as follows:

1) *TCA:* a traditional global domain adaptation method using MMD. The EEG features learned by DNN are used as inputs, and the k-nearest neighbor method (KNN) is used as the classifier.

2) *JDA:* a traditional class-related subdomain adaptation method using MMD. The EEG features learned by DNN are used as inputs, and KNN is used as the classifier.

3) *DNN:* a deep learning method modified from the proposed MSDAN by removing the domain adaptation parts. Only the classification loss is optimized in the loss function.

4) *D_CORAL:* a domain adaptation method that aligns the second-order statistics of the source and target domains. The scale factor λ is set to 100 to balance the classification and domain adaptation losses.

5) *DDC:* a domain adaptation method that uses MMD to minimize the global distribution difference between the source and target domains. The scale factor λ is set as the same as $\lambda^{<c>}$ in MSDAN.

6) *DDAN:* a subdomain adaptation method that uses MMD to minimize the global domain difference and uses the center-based discriminative feature learning method to maximize the inter-class discrepancy of the source domain. The scale factor of global domain adaptation loss λ_1 is set as the same as $\lambda^{<c>}$ in MSDAN. The scale factor of class-related subdomain loss λ_2 is set to 0.01 as recommended in [16].

7) *SDAN:* a subdomain adaptation method modified from the proposed MSDAN by removing the time-related subdomain adaptation part. The class-related subdomain adaptation loss is combined with the classification loss in the objective function. The scale factor of the class-related subdomain adaptation loss $\lambda^{<c>}$ is set the same as $\lambda^{<c>}$ in MSDAN.

In all deep learning methods, the learning rate is optimized by grid search and set to 0.001, the batch size is set to 32, and the number of epochs is set to 400. The SGD is selected as the optimizer. The L2 weight decay regularization is implemented to prevent the model from over-fitting with a decay coefficient of 0.01. The programming language adopted in this paper is Python. All methods are implemented based on the Pytorch framework. The hardware parameters are Intel Xeon CPU E5-2678 v3, Nvidia GeForce Titan RTX GPU 24GB, and 64GB RAM.

C. Experiment Results and Discussion

The experiment results of the proposed method and other comparison methods on the BCI competition III-IVa dataset

TABLE III
EXPERIMENTAL RESULTS ON BCI COMPETITION III-IVA DATASET

Subjects	TCA	JDA	DNN	D_CORAL	DDC	DDAN	SDAN	MSDAN
aa→al	85.22%	83.21%	87.14%	84.29%	91.07%	93.93%	93.21%	**95.71%**
aa→av	60.00%	64.29%	64.29%	63.57%	66.43%	63.21%	63.93%	**68.57%**
aa→aw	78.70%	81.43%	83.93%	84.29%	83.57%	84.29%	85.36%	**86.07%**
aa→ay	63.48%	63.93%	69.29%	67.14%	86.79%	86.79%	92.50%	**93.57%**
al→aa	75.65%	77.14%	76.43%	81.43%	82.50%	**84.29%**	79.64%	82.14%
al→av	59.13%	62.50%	61.79%	59.29%	66.43%	62.50%	61.43%	**73.57%**
al→aw	79.57%	79.29%	82.50%	81.07%	81.43%	83.57%	83.93%	**83.93%**
al→ay	82.61%	81.07%	81.43%	87.86%	90.71%	87.14%	92.50%	**94.29%**
av→aa	63.48%	63.93%	70.36%	67.14%	69.64%	68.93%	73.93%	**74.64%**
av→al	57.39%	59.29%	76.07%	74.29%	81.07%	83.93%	84.29%	**87.86%**
av→aw	59.57%	66.79%	72.50%	71.79%	68.21%	71.79%	77.14%	**80.36%**
av→ay	58.70%	73.21%	77.14%	81.43%	79.64%	82.14%	90.00%	**90.71%**
aw→aa	71.74%	76.07%	75.36%	75.36%	77.50%	76.43%	76.07%	**77.50%**
aw→al	76.96%	80.36%	81.07%	76.79%	88.57%	92.50%	92.50%	**92.86%**
aw→av	48.26%	44.64%	53.57%	55.36%	58.57%	57.86%	56.79%	**62.14%**
aw→ay	75.22%	75.00%	80.36%	82.50%	84.64%	85.00%	**90.71%**	89.64%
ay→aa	60.87%	61.79%	64.29%	**66.79%**	66.07%	66.07%	66.07%	66.43%
ay→al	78.26%	77.14%	82.50%	83.21%	86.43%	87.14%	88.21%	**93.21%**
ay→av	56.09%	70.71%	67.86%	**76.43%**	70.71%	70.36%	70.00%	70.00%
ay→aw	66.09%	66.79%	67.14%	74.29%	81.79%	80.71%	86.79%	**88.93%**
Average accuracy	67.85%	70.43%	73.75%	74.72%	78.09%	78.43%	80.25%	**82.61%**
Average time (s)	142.6	156.8	141.8	181.8	189.6	193.4	205.8	220.2
Significance level	2.43E-07	2.99E-06	5.23E-06	6.46E-04	1.56E-05	5.84E-05	1.82E-03	N.A.

are provided in Table III. The best classification accuracy of each transfer task is shown in boldface. In most tasks, the classification performance of the proposed MSDAN method outperforms other methods. The overall performance results in terms of classification accuracy and running time are shown in Fig. 4.

From the experimental results, the following observations can be obtained. The average accuracy of MSDAN is 82.61% which increases by 4.18% and 2.36% compared with the class-related subdomain adaptation methods DDAN and SDAN, respectively. Such results imply that the time-related data distribution problem needs to be considered, and the proposed MSDAN is able to handle the concerned problem in the STS MI classification task. Note that: 1) compared with global domain adaptation methods D_CORAL and DDC, the MSDAN method increases the classification accuracy by 7.89% and 4.52%, respectively; 2) compared with the DNN, the MSDAN method has an average increase in the classification accuracy of 8.86%. These comparison results indicate that the proposed MSDAN method can handle the time-related distribution shift problem and improve the overall classification accuracy.

As the proposed MSDAN is an end-to-end deep learning method designed to handle the distribution shift problem, it is possible to have minor outlier cases which the proposed methods may not achieve the best result among all the other method. In order to show the effectiveness of the proposed MSDAN method, the pairwise two-tailed t-test [53] is adopted to investigate whether the improvement of the proposed MSDAN has statistical significance over other methods. As all the significance levels between MSDAN and other methods in the last row of Table III are lower than 0.05, the proposed MSDAN method is significantly better than all the other methods in this paper.

To visualize the effectiveness of the proposed method, the data features extracted by DNN, D_CORAL, DDC, DDAN, SDAN, and MSDAN in the $aa \rightarrow al$ MI classification task are visualized by the t-SNE method [55]. TCA and JDA methods are not included as the feature extractions are the same as DNN. From the six subfigures in Fig. 5, the following observations can be obtained:

1) In Fig. 5(a), the data feature distribution difference between the two classes is not clear, showing the data feature extraction ability of DNN is not satisfactory;
2) In Fig. 5(b) and 5(c), the classification boundaries are more apparent than Fig. 5(a), showing the effectiveness of domain adaptation in the classification task by D_CORAL and DDC;
3) In Fig. 5(d) and 5(e), the classification boundaries are sharper than Fig. 5(b) and 5(c), showing the effectiveness of subdomain adaptation in the classification task by DDAN and SDAN;
4) In Fig. 5(f), the data features extracted by the proposed multi-subdomain adaptation method MSDAN achieve the

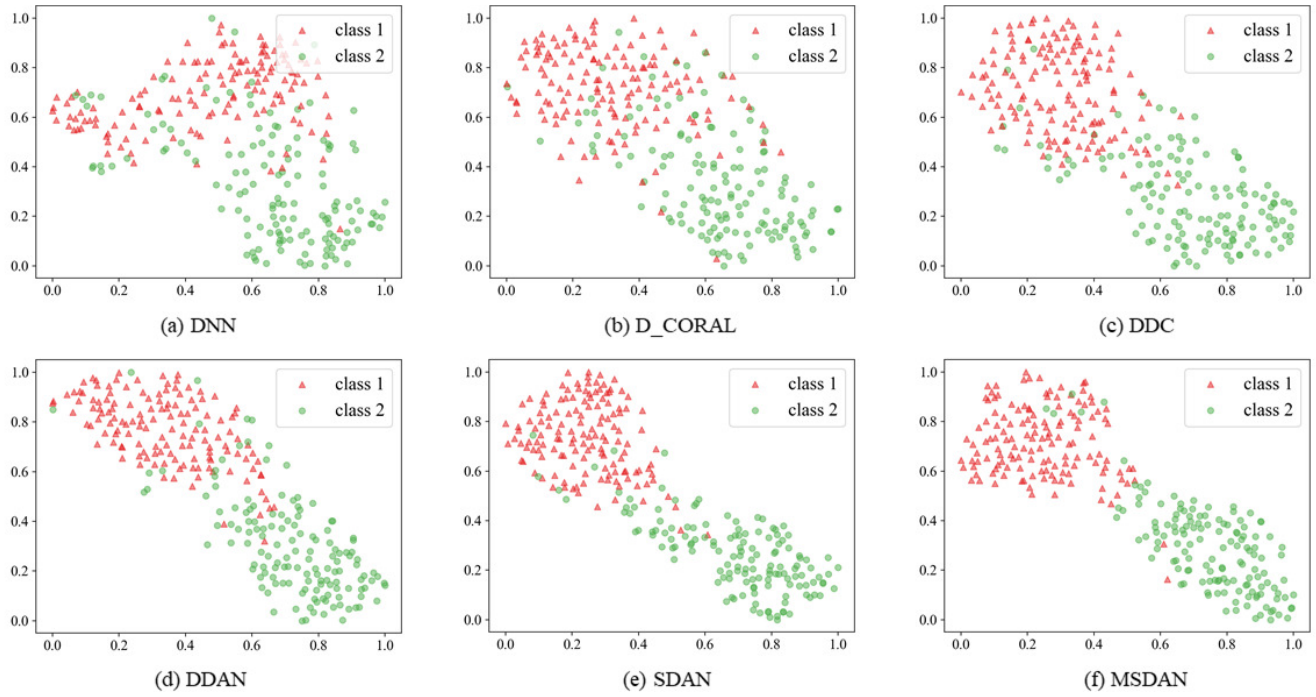


Fig. 5. Visualizations of features extracted by different methods in *aa* → *al* MI classification task.

most apparent classification boundary between different classes.

Therefore, Fig. 5 illustrates that MSDAN can extract better data features than all the other methods and the time-related subdomain adaptation helps learn discriminative data features in the EEG-based STS MI classification task.

V. Conclusion

In this paper, the time-related data distribution shift problem has been investigated in the EEG-based STS cross-subject MI classification task. As the data distribution shift increases the difference between the source and target domains and challenges the classification task, a novel MSDAN method has been proposed to solve the time-related distribution shift problem. The proposed method can minimize the classification loss, class-related adaptation loss, and time-related adaptation loss concurrently during the model training process. To illustrate the effectiveness of the proposed MSDAN method, STS MI classification experiments have been conducted on the BCI competition III-IVa dataset. The overall classification results in every STS MI classification task of the experiments have shown that the proposed method is capable of improving the classification performance compared with other well-known domain adaptation and deep learning methods. Based on the statistical significance analysis, the performances of the proposed MSDAN method have proven to be significantly improved than that of all the other methods in STS cross-subject MI classification. For the future works, the extension of the MSDAN method in solving other EEG classification problems such as the data fading and missing problems in the EEG measurement is a possible future research direction [14], [28], [31]. In addition, the MSDAN can also be updated and generalized to solve the classification problems in other BCI and biomedical applications, considering that time-related data distribution shift is a common problem in physiological signals [10], [20], [32], [34], [62], [63].

Acknowledgment

This research has been approved by University Ethics Committee in XJTLU.

References

[1] A. M. Abdelhameed and M. Bayoumi, "Semi-supervised deep learning system for epileptic seizures onset prediction," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Orlando, FL, USA, Dec. 2018, pp. 1186–1191.

[2] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, Dec. 2019.

[3] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1352–1359, Jul. 2019.

[4] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp. 432–441, Oct. 2008.

[5] N. Birbaumer and L. G. Cohen, "Brain–computer interfaces: Communication and restoration of movement in paralysis," *J. Physiol.*, vol. 579, no. 3, pp. 621–636, Mar. 2007.

[6] A. Buttfield, P. W. Ferrez, and J. R. Millan, "Towards a robust BCI: Error potentials and online learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 164–168, Jun. 2006.

[7] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nature Rev. Neurosci.*, vol. 13, no. 6, pp. 407–420, 2012.

[8] T. Cao, F. Wan, C. M. Wong, J. N. da Cruz, and Y. Hu, "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain–computer interfaces," *Biomed. Eng. Online*, vol. 13, no. 1, p. 28, 2014.

[9] S. Charbonnier, R. N. Roy, S. Bonnet, and A. Campagne, "EEG index for control operators' mental fatigue monitoring using interactions between brain regions," *Expert Syst. Appl.*, vol. 52, pp. 91–98, Jun. 2016.

[10] H. Cheng, Z. Wang, Z. Wei, L. Ma, and X. Liu, "On adaptive learning framework for deep weighted sparse autoencoder: A multiobjective evolutionary algorithm," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3221–3231, May 2022.

[11] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[12] T. Duan *et al.*, "Meta learn on constrained transfer learning for low resource cross subject EEG classification," *IEEE Access*, vol. 8, pp. 224791–224802, 2020.

[13] Z. Fang, W. Wang, and Z.-G. Hou, "Convolutional LSTM: A deep learning method for motion intention recognition based on spatiotemporal EEG data," in *Proc. Int. Conf. Neural Inf. Process.*, Sydney, NSW, Australia, 2019, pp. 216–224.

[14] H. Geng, H. Liu, L. Ma, and X. Yi, "Multi-sensor filtering fusion meets censored measurements under a constrained network environment: Advances, challenges and prospects," *Int. J. Syst. Sci.*, vol. 52, no. 16, pp. 3410–3436, Dec. 2021.

[15] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, Jan. 2012.

[16] W. Hang *et al.*, "Cross-subject EEG signal recognition using deep domain adaptation network," *IEEE Access*, vol. 7, pp. 128273–128282, 2019.

[17] H. He and D. Wu, "Channel and trials selection for reducing covariate shift in EEG-based brain–computer interfaces," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Bari, Italy, Oct. 2019, pp. 3635–3640.

[18] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.

[19] W. Heller, "Neuropsychological mechanisms of individual differences in emotion, personality, and arousal," *Neuropsychology*, vol. 7, no. 4, pp. 476–489, 1993.

[20] J. Hu, C. Jia, H. Liu, X. Yi, and Y. Liu, "A survey on state estimation of complex dynamical networks," *Int. J. Syst. Sci.*, vol. 52, no. 16, pp. 3351–3367, Dec. 2021.

[21] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, and K. E. Hild, "A framework for rapid visual image search using single-trial brain evoked responses," *Neurocomputing*, vol. 74, no. 12, pp. 2041–2051, Jun. 2011.

[22] R. Hu *et al.*, "Cross-subject federated transfer learning with quanvolutional layer for motor imagery classification," in *Proc. China Autom. Congr. (CAC)*, Oct. 2021, pp. 5736–5741.

[23] M. Jobert *et al.*, "Guidelines for the recording and evaluation of pharmaco-EEG data in man: The international pharmaco-EEG society (IPEG)," *Neuropsychobiology*, vol. 66, no. 4, pp. 201–220, 2012.

[24] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Netw.*, vol. 57, pp. 39–50, Sep. 2014.

[25] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[26] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1318–1324, Jun. 2010.

[27] G. Li *et al.*, "The impact of mental fatigue on brain activity: A comparative study both in resting state and task state using EEG," *BMC Neurosci.*, vol. 21, no. 1, pp. 1–9, Dec. 2020.

[28] W. Li, Y. Niu, and Z. Cao, "Event-triggered sliding mode control for multi-agent systems subject to channel fading," *Int. J. Syst. Sci.*, vol. 53, no. 6, pp. 1233–1244, Apr. 2022.

[29] Y. Liang and Y. Ma, "Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning," *Biomed. Signal Process. Control*, vol. 62, Sep. 2020, Art. no. 102101.

[30] Y. Liu, Z. Lan, J. Cui, O. Sourina, and W. Müller-Wittig, "EEG-based cross-subject mental fatigue recognition," in *Proc. Int. Conf. Cyberworlds (CW)*, Kyoto, Japan, Oct. 2019, pp. 247–252.

[31] L. Liu, L. Ma, J. Zhang, and Y. Bo, "Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks," *Int. J. Syst. Sci.*, vol. 52, no. 6, pp. 1192–1205, Apr. 2021.

[32] W. Liu, Z. Wang, Y. Yuan, N. Zeng, K. Hone, and X. Liu, "A novel sigmoid-function-based adaptive weighted particle swarm optimizer," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 1085–1093, Feb. 2021.

[33] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2200–2207.

[34] X. Luo, Y. Yuan, S. Chen, N. Zeng, and Z. Wang, "Position-transitional particle swarm optimization-incorporated latent factor analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3958–3970, Aug. 2022, doi: 10.1109/TKDE.2020.3033324.

[35] V. Mihajlovic and B. Grundlehner, "The effect of force and electrode material on electrode-to-skin impedance," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Hsinchu, Taiwan, Nov. 2012, pp. 57–60.

[36] V. Mihajlovic, B. Grundlehner, R. Vullers, and J. Penders, "Wearable, wireless EEG solutions in daily life applications: What are we missing?" *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 6–21, Jan. 2015.

[37] P. Nguyen, D. Tran, X. Huang, and D. Sharma, "A proposed feature extraction method for EEG-based person identification," in *Proc. Int. Conf. Artif. Intell.*, Las Vegas, NV, USA, 2012, pp. 826–831.

[38] O. Ozdenizci, Y. Wang, T. Koike-Akino, and D. Erdogmus, "Learning invariant representations from EEG via adversarial inference," *IEEE Access*, vol. 8, pp. 27074–27085, 2020.

[39] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[40] H. Raza, A. Chowdhury, and S. Bhattacharyya, "Deep learning based prediction of EEG motor imagery of stroke patients' for neuro-rehabilitation application," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Glasgow, U.K., Jul. 2020, pp. 1–8.

[41] M. Riyad, M. Khalil, and A. Adib, "Cross-subject EEG signal classification with deep neural networks applied to motor imagery," in *Proc. 5th Int. Conf. Mobile, Secure, Program. Netw.*, Mohammedia, Morocco, 2019, pp. 124–139.

[42] K. Roots, Y. Muhammad, and N. Muhammad, "Fusion convolutional neural network for cross-subject EEG motor imagery classification," *Computers*, vol. 9, no. 3, pp. 1–9, 2020.

[43] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 371–382, Feb. 2018.

[44] S. Saha *et al.*, "Wavelet entropy-based inter-subject associative cortical source localization for sensorimotor BCI," *Frontiers Neuroinform.*, vol. 13, pp. 1–11, Jul. 2019.

[45] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cogn. Emotion*, vol. 15, no. 4, pp. 487–500, Jul. 2001.

[46] D. J. Schutter, P. Putman, E. Hermans, and J. van Honk, "Parietal electroencephalogram beta asymmetry and selective attention to angry facial expressions in healthy human subjects," *Neurosci. Lett.*, vol. 314, no. 1, pp. 13–16, Nov. 2001.

[47] A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiol. Meas.*, vol. 21, no. 2, pp. 271–283, May 2000.

[48] A. Shalbaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," *Phys. Eng. Sci. Med.*, vol. 43, no. 4, pp. 1229–1239, Dec. 2020.

[49] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A Hilbert space embedding for distributions," in *Proc. 18th Int. Conf. Algorithmic Learn. Theory*, Sendai, Japan, vol. 2007, pp. 13–31.

[50] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 443–450.

[51] B. Sun, H. Zhang, Z. Wu, Y. Zhang, and T. Li, "Adaptive spatiotemporal graph convolutional networks for motor imagery classification," *IEEE Signal Process. Lett.*, vol. 28, pp. 219–223, 2021.

[52] S. Taghizadeh, S. Pirouzi, A. Zamani, A. Motealleh, and Z. Bagheri, "Does muscle fatigue alter EEG bands of brain hemispheres?" *J. Biomed. Phys. Eng.*, vol. 10, no. 2, pp. 187–196, Jul. 2017.

[53] H. W. Thompson, R. Mera, and C. Prasad, "A description of the appropriate use of student's *t*-test," *Nutritional Neurosci.*, vol. 1, no. 2, pp. 165–172, Jan. 1998.

[54] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.

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

[56] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021.

[57] Z. Wan, R. Yang, M. Huang, W. Liu, and N. Zeng, "EEG fading data classification based on improved manifold learning with adaptive neighborhood selection," *Neurocomputing*, vol. 482, pp. 186–196, Apr. 2022.

[58] X. Wang, R. Yang, M. Huang, Z. Yang, and Z. Wan, "A hybrid transfer learning approach for motor imagery classification in brain–computer interface," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Nara, Japan, Mar. 2021, pp. 496–500.

[59] M. Wei, R. Yang, and M. Huang, "Motor imagery EEG signal classification based on deep transfer learning," in *Proc. IEEE 34th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2021, pp. 85–90.

[60] K. Won, M. Kwon, M. Ahn, and S. C. Jun, "Selective subject pooling strategy to achieve subject-independent motor imagery BCI," *Sensors*, vol. 21, no. 16, pp. 3–6, 2021.

[61] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in EEG decoding with deep learning," *Frontiers Hum. Neurosci.*, vol. 14, p. 103, Apr. 2020.

[62] W. Yue, Z. Wang, B. Tian, M. Pook, and X. Liu, "A hybrid model- and memory-based collaborative filtering algorithm for baseline data prediction of Friedreich's ataxia patients," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1428–1437, Feb. 2021.

[63] W. Yue, Z. Wang, J. Zhang, and X. Liu, "An overview of recommendation techniques and their applications in healthcare," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 4, pp. 701–717, Apr. 2021.

[64] H. Zhao, Y. Wang, S. Sun, W. Pei, and H. Chen, "Obviating session-to-session variability in a rapid serial visual presentation-based brain–computer interface," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, San Francisco, CA, USA, Mar. 2019, pp. 171–174.

[65] X. Zhao, J. Zhao, C. Liu, and W. Cai, "Deep neural network with joint distribution matching for cross-subject motor imagery brain–computer interfaces," *BioMed Res. Int.*, vol. 2020, pp. 1–15, Feb. 2020.

[66] H. Zhao, Q. Zheng, K. Ma, H. Li, and Y. Zheng, "Deep representation-based domain adaptation for nonstationary EEG classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 535–545, Feb. 2021.

[67] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, "Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface," in *Proc. 32nd AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, pp. 1703–1710.

[68] X. Z. Zhang, W. L. Zheng, and B. L. Lu, "EEG-based sleep quality evaluation with deep transfer learning," in *Proc. 24th Int. Conf. Neural Inf. Process.*, Guangzhou, China, 2017, pp. 543–552.

[69] M. Zheng, B. Yang, and Y. Xie, "EEG classification across sessions and across subjects through transfer learning in motor imagery-based brain-machine interface system," *Med. Biol. Eng. Comput.*, vol. 58, no. 7, pp. 1515–1528, Jul. 2020.

[70] Y. Zhu *et al.*, "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.