Computers in Biology and Medicine 163 (2023) 107235

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A deep domain adaptation framework with correlation alignment for EEG-based motor imagery classification

Xiao-Cong Zhong, Qisong Wang *, Dan Liu, Jing-Xiao Liao, Runze Yang, Sanhe Duan, Guohua Ding, Jinwei Sun

School of Instrumentation Science and Engineering, Harbin Institute of Technology, Harbin, 150001, China

ARTICLE INFO

Keywords:
Brain-computer interface (BCI)
Motor imagery (MI)
Electroencephalography (EEG)
Domain adaptation (DA)
Correlation alignment

ABSTRACT

It is impractical to collect sufficient and well-labeled EEG data in Brain-computer interface because of the time-consuming data acquisition and costly annotation. Conventional classification methods reusing EEG data from different subjects and time periods (across domains) significantly decrease the classification accuracy of motor imagery. In this paper, we propose a deep domain adaptation framework with correlation alignment (DDAF-CORAL) to solve the problem of distribution divergence for motor imagery classification across domains. Specifically, a two-stage framework is adopted to extract deep features for raw EEG data. The distribution divergence caused by subjected-related and time-related variations is further minimized by aligning the covariance of the source and target EEG feature distributions. Finally, the classification loss and adaptation loss are optimized simultaneously to achieve sufficient discriminative classification performance and low feature distribution divergence. Extensive experiments on three EEG datasets demonstrate that our proposed method can effectively reduce the distribution divergence between the source and target EEG data. The results show that our proposed method delivers outperformance (an average classification accuracy of 92.9% for within-session, an average kappa value of 0.761 for cross-session, and an average classification accuracy of 83.3% for cross-subject) in two-class classification tasks compared to other state-of-the-art methods.

1. Introduction

Brain-computer interface (BCI) provides a direct way of information exchange between human beings and external devices [1,2], which attracts the great interest of a variety of researchers. Nowadays, BCI systems have important application prospects in the field of communication, transportation, and aerospace, especially in rehabilitation medicine [3]. For patients with hemiplegia caused by stroke, rehabilitation with BCI systems can achieve a synergistic training of injured brain and limbs [4], which can stimulate the remodeling process of damaged brain nerves and accelerate recovery [5]. To achieve an accurate, specific, and comprehensive monitor of brain activity, various measurements were invented one after another. Due to its simplicity, non-invasive and high temporal resolution [6], EEG is widely applied in BCI systems. EEG-based motor imagery has gained much attention in BCI research in the last 20 years because of its convenient control to external devices.

To better improve the performance of motor imagery classification, researchers made great efforts for EEG signal processing [7,8], feature extraction [9–11], and classification [12–15]. Despite a big stride that has been made in the past decade, collecting sufficient and well-labeled EEG data for motor imagery classification remains a great challenge, especially for patients. A simple way to solve the problem is to reuse EEG data from different subjects and time periods in the same motor imagery scenarios. Even if reusing EEG data can construct a large-scale dataset for training and testing, it still suffers from the scalability problem because EEG data from different subjects and time periods present different distributions.

So what makes EEG data distributions different? There are two main factors. (1) ***subject-related variations***: Many previous works suggested that EEG signal is a kind of highly subject-related characteristic [16–18] and people differ in the way of feeling and expressing emotions [19]. To be specific, the subject-related variations primarily come from attention [20], stress level [21], age [22], physiology [23], reaction time, handedness, and other variables, which lead to the EEG data distribution differences at the end. (2) ***time-related variations***: EEG signals change over time due to electrode impedance [24,25] and fatigue level [26]. In addition, because of the nonstationarity [27], EEG data of the same subject vary at different time periods, which makes EEG data unreusable in return. In general, the subject-related variations

* Corresponding author.
E-mail address: wangqisong@hit.edu.cn (Q. Wang).

https://doi.org/10.1016/j.compbiomed.2023.107235
Received 16 February 2023; Received in revised form 30 June 2023; Accepted 1 July 2023
Available online 4 July 2023

and time-related variations lead to a common problem: ***distribution divergence***.

Conventional classification methods usually assume that the training and testing data have the same distribution. The assumption does not hold in most motor imagery scenarios. Previous studies also pointed out that distribution divergence is a critical problem and it should be considered [28,29]. But most of them mainly focus on minimizing Maximum Mean Discrepancy (MMD [30]) to reduce the marginal distribution divergence [31,32]. Minimizing MMD is to align the first-order statistical features across domains (source domain and target domain) so that the nonlinear relationship of different distributions cannot be well exploited. To the best of our knowledge, most of the previous mainstream methods reduce the distribution divergence regardless of the nonlinear relationship. In this paper, different from MMD, we turn our eyes to seek a nonlinear distance to align the distributions and propose a deep domain adaptation framework with correlation alignment (DDAF-CORAL), which aligns the second-order statistics (covariance) between the source and target domains. The experiment results demonstrate that the DDAF-CORAL effectively addresses the distribution divergence caused by subject-related and time-related variations, and achieves outperformance against other state-of-the-art methods.

Main Contributions: (1) We investigate the reason for the challenging problem of EEG data distribution divergence. (2) A two-stage domain adaptation framework is designed to extract deep features automatically for the source and target EEG data. Combined with correlation alignment, the classification loss and adaptation loss are optimized simultaneously to minimize the distribution divergence between the source and target EEG data. (3) Extensive experiments on datasets suggest that the proposed method obtains great generalization performance on EEG data with various distributions and delivers competitive results compared to other well-known methods.

This paper is structured as follows. Some related works are briefly reviewed in Section 2. The proposed method is presented in Section 3. Section 4 introduces the datasets used in this context and analyzes the experiments in detail. Finally, conclusions are shown in Section 5.

2. Related works

Transfer Learning: Existing transfer learning methods can be mainly classified into three categories as follows. (1) *instance transfer*, which re-weights the labeled data of the source domain to increase the weight of similar samples in the source and target domains [33,34]. (2) *feature transfer*, which maps the samples of the source and target domains to a new feature space by feature transformation, to minimize the differences between the source and target domains in the new feature space [35–37]. (3) *parameter transfer*, which shares the parameters between the source and target domain models to improve the generalization performance on the target data [38–40].

DA for Motor Imagery Classification: To address the distribution divergence of EEG data, a variety of transfer learning methods, especially feature-transfer and parameter-transfer, were employed to motor imagery classification in recent years. Han et al. [31] proposed a deep domain adaptation network (DDAN) using MMD and discriminative feature learning (CDFL), which minimized the distribution discrepancy of the source and target EEG features and forced the features closer to their class centers. Chen et al. [32] measured the distribution differences between the source and target subdomains with MMD and optimized the adaptation and classification losses concurrently, realizing the best classification performance of cross-subject experiments on BCI competition III-IVa. As described above, the two domain adaptation methods mainly focus on minimizing MMD to reduce the marginal distribution differences between the source and target EEG data, and both achieve high classification performance. In this work, we borrow the idea of domain adaptation and propose a deep domain adaptation framework with correlation alignment to learn a classifier with strong generalization ability and transferability from the source and target domains.

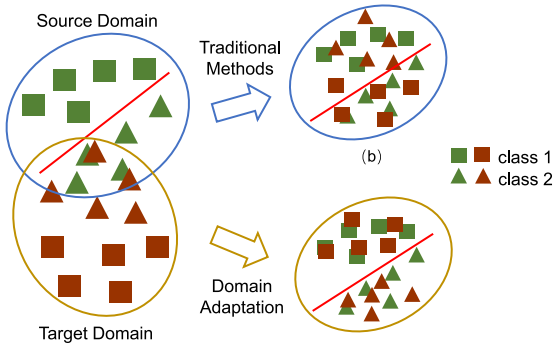

Fig. 1. Illustration of EEG data distribution divergence.

3. Proposed method

3.1. Problem definition and notations

Suppose we are given a well-labeled source domain $\mathcal{D}_s = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ and a unlabeled target domain $\mathcal{D}_t = \{\boldsymbol{x}_j^t\}_{j=1}^{n_t}$. $n \in \{1, \ldots, N\}$ is the class of label. $\boldsymbol{x}_i^s \in \mathbb{R}^{C\times T}$ and $\boldsymbol{x}_j^t \in \mathbb{R}^{C\times T}$ denote the ith and jth trials of the source and target domain EEG data respectively. Because of the subject-related variations and time-related variations, there is an obvious distribution divergence between the EEG data from different domains (shown in Fig. 1). So the goal of our work is to design a classifier $y = f(x)$ with the supervised source EEG data $\{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ and the unsupervised target EEG data $\{\boldsymbol{x}_j^t\}_{j=1}^{n_t}$ to predict labels $\{y_j^t\}_{j=1}^{n_t}$ for the target EEG data.

3.2. DDAF-CORAL

A deep domain adaptation framework with correlation alignment called DDAF-CORAL is proposed as described in Fig. 2. The proposed DDAF-CORAL consists of two main parts: feature extraction, and domain adaptation. On the one hand, a two-stage network is adopted to extract deep features for raw EEG data. The structure of input EEG data is a matrix with the size of (C, T), where C is the number of EEG channels and T is the time. The source domain and target domain EEG data are fed to the two-stage network sequentially to complete the extraction from general features to task-relevant specific features. Here, the network parameters in the purple part of Fig. 2 are transferred from the source network to the target network. On the other hand, we utilize a domain adaptation layer with correlation alignment to match the distributions between the source and target features, which minimizes the distance of the second-order statistics between source and target features. Since the source data is all labeled but the target data is not, the source data is trained under supervision with their labels to obtain the classification loss, while the target features and source features are domain adapted to compute the domain adaptation loss.

Therefore, the objective function of the proposed method comprises the classification loss and the domain adaptation loss. The loss function can be formalized as:

$$\mathcal{L} = \mathcal{L}_{clc} + \lambda \mathcal{L}_{coral} \tag{1}$$

where $\mathcal{L}_{clc}$ denotes the classification loss of the source data, $\mathcal{L}_{coral}$ denotes the domain adaptation loss measuring the distance of the covariance between the source and target features, and λ ($\lambda > 0$) is a weight of adjusting the adaptation loss.

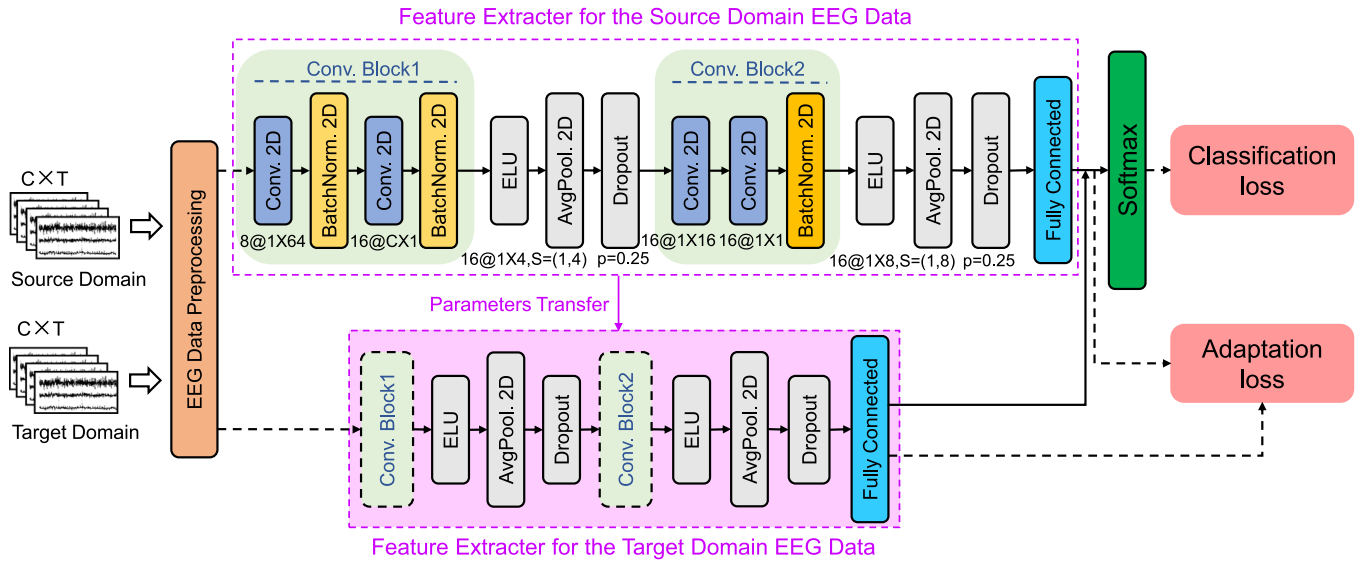

Fig. 2. Overall framework of our proposed DDAF-CORAL for EEG-based motor imagery classification.

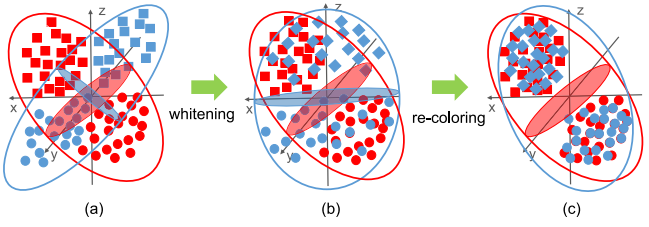

Fig. 3. Correlation alignment for domain adaptation. The blue color and red color denote the source domain and the target domain respectively. (a) Original source and target features. (b) Whitened features, *i.e.* removing the correlation of the source features. (c) Recolored features, *i.e.* adding the correlation of the target features to the source features.

Similar to other deep neural networks, the classification loss $\mathcal{L}_{clc}$ can be obtained with the labeled source EEG data as follows.

$$\begin{aligned}\mathcal{L}_{clc} &= \frac{1}{n_s}\sum_{i=1}^{n_s} J(f(\boldsymbol{x}_i^s), y_i^s) \\ &= -\frac{1}{n_s}\sum_{i=1}^{n_s}[y_i^s log(f(\boldsymbol{x}_i^s)) + (1-y_i^s) log(1-f(\boldsymbol{x}_i^s))] \\ &= -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{n=1}^{N} log(f(\boldsymbol{x}_i^s)) \cdot \mathbb{I}(y_i^s = n)\end{aligned} \tag{2}$$

where $J(\cdot,\cdot)$ denotes the cross entropy loss function, $\mathbb{I}(\cdot)$ is an indicator which equals to 1 if the condition true and 0 otherwise.

Although the classifier can learn discriminative features with the source EEG data, the distribution divergence between the source and target features still remains. It is notable that the learned features are normalized in each dimension by batch normalization layers. Even so, the source and target features also present different distributions as shown in Fig. 3a. In this case, minimizing the MMD of the marginal distribution cannot reduce the distribution divergence any longer. To address this problem, we apply correlation alignment to minimize the distance of the covariance between the source and target features. We formalize the distance using the Frobenius norm:

$$\begin{aligned}&min \|C_{\hat{s}} - C_t\|_F^2 \\ &= \min_M \|\frac{1}{n_s - 1}(D_s^\top M)(D_s M) - C_t\|_F^2 \\ &= \min_M \|M^\top C_s M - C_t\|_F^2\end{aligned} \tag{3}$$

where C_s (C_t) is the covariance of the source (target) features, D_s (D_t) denotes the learned source (target) features by the fully connected layer, M is the linear transformation applied to the source EEG features, and $C_{\hat{s}}$ is the covariance of the source EEG features after transformation.

According to the theorem [41,42], $C_{\hat{s}}^* = C_t$ is the optimal solution to the problem in Eq. (3). By using singular value decomposition (SVD) on C_s and C_t, we can obtain $C_s = U_s W_s V_s^\top$ and $C_t = U_t W_t V_t^\top$, where $U_s V_s^\top = \boldsymbol{I}$ (identity matrix). Since C_t and C_s are symmetric matrices, Eq. (3) can formulated as:

$$M^\top U_s W_s U_s^\top M = U_t W_t U_t^\top \tag{4}$$

Let $W_t = W_t^{\frac{1}{2}} U_t^\top U_s W_s^{-\frac{1}{2}} U_s^\top U_s W_s U_s^\top U_s W_s^{-\frac{1}{2}} U_s^\top U_t W_t^{\frac{1}{2}}$ and $X = U_s W_s^{-\frac{1}{2}} U_s^\top U_t W_t^{\frac{1}{2}}$. With the new notation X, the right side of Eq. (4) can be rewritten as:

$$\begin{aligned}U_t W_t U_t^\top &= U_t X^\top U_s W_s U_s^\top X U_t^\top \\ &= (X U_t^\top)^\top U_s W_s U_s^\top (X U_t^\top)\end{aligned} \tag{5}$$

Hence, replace the right side in Eq. (4), and the linear transformation M for source EEG features can be obtained: $M = X U_t^\top$. Specifically, M also can be formulated as follows.

$$M = (U_s W_s^{-\frac{1}{2}} U_s^\top) \cdot (U_t W_t^{\frac{1}{2}} U_t) = C_s^{-\frac{1}{2}} C_t^{\frac{1}{2}} \tag{6}$$

Finally, the source EEG features transformed by M can be calculated as Eq. (7). The process of correlation alignment for domain adaptation is shown in Fig. 3. $C_s^{-\frac{1}{2}}$ whitens the source EEG features by removing its correlation as shown in Fig. 3b, and $C_t^{\frac{1}{2}}$ recolors it with the covariance of the target EEG features as shown in Fig. 3c. In this way, we bring the covariance of the source and target features into alignment.

$$D_{\hat{s}} = D_s C_s^{-\frac{1}{2}} C_t^{\frac{1}{2}} \tag{7}$$

Motivated by the linear transformation, we conduct a deep neural network to seek a nonlinear transformation. The key advantage of using such a network to minimize the distribution divergence is that the nonlinear features can be explicitly obtained. Therefore, we make full use of the nonlinear characteristic of the covariance to align the different distributions. The distance of the covariance between the source and target features (the coral loss) can be calculated as:

$$\mathcal{L}_{coral} = \frac{1}{4d^2}\|C_s - C_t\|_F^2 \tag{8}$$

where d is the dimension of the output features of the fully connected layer. Through the chain rule, the gradient with respect to the output features can be calculated:

$$\frac{\partial \mathcal{L}_{coral}}{\partial D_s} = \frac{1}{d^2(n_s-1)}\left(\frac{D_s^{\top} D_s}{n_s-1} - \frac{D_t^{\top} D_t}{n_t-1}\right) D_s \tag{9}$$

$$\frac{\partial \mathcal{L}_{coral}}{\partial D_t} = \frac{1}{d^2(n_t-1)}\left(\frac{D_s^{\top} D_s}{n_s-1} - \frac{D_t^{\top} D_t}{n_t-1}\right) D_t \tag{10}$$

Combining the classification loss in Eq. (2) and the coral loss in Eq. (8), the gradient of the final objective function $\mathcal{L}$ can be expressed as ∇_p, which is optimized by the adaptive moment estimation (Adam) algorithm.

$$\begin{aligned}\nabla_p &= \frac{\partial \mathcal{L}_{clc}}{\partial p} + \lambda \frac{\partial \mathcal{L}_{coral}}{\partial p} \\ &= \frac{1}{b_s}\sum_{i=1}^{b_s} \frac{\partial J(f(\boldsymbol{x}_i^s), y_i^s)}{\partial p} + \lambda \frac{\partial \mathcal{L}_{coral}}{\partial D_s}\frac{\partial D_s}{\partial p} + \lambda \frac{\partial \mathcal{L}_{coral}}{\partial D_t}\frac{\partial D_t}{\partial p}\end{aligned} \tag{11}$$

where p is the network's parameter and b_s is the batch size.

The network parameters are shared between the two networks using batch covariances, accomplishing the purpose of domain adaptation. According to the learning objective of our proposed method, the optimized network can provide sufficient discriminative performance to classify the target EEG data and minimize the distribution divergence between the source and target EEG data after the training process. The complete learning process of the proposed framework is summarized in Algorithm 1.

Algorithm 1: DDAF-CORAL for EEG DA

Input: Source EEG data $\mathcal{D}_s = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{n_s}$, target EEG data $\{\boldsymbol{x}_j^t\}_{j=1}^{n_t}$ and λ.

Output: Classification results $\{y_j^t\}_{j=1}^{n_t}$.

```
Preprocess source and target EEG data D_s, D_t;
Initialize the model parameters p;
while not converge do
    Sample a couple-batch B = {B_s^n, B_t^n}_{n=1}^{batch_size} from domain D_s, D_t;
    Extract source EEG features D_s;
    Compute the classification loss L_clc by Eq. (2);
    Transfer parameters p to the target network;
    Extract target EEG features D_t ;
    Calculate the adaptation loss L_coral by Eq. (8);
    Combine L_coral and L_clc as L by Eq. (1);
    Compute gradient ∇_p and update parameters p by Eq. (11).
end
Predict labels on the target data {x_j^t}_{j=1}^{n_t}.
return Classification results {y_j^t}_{j=1}^{n_t} on target data.
```

4. Experiments and discussion

4.1. Datasets description

To evaluate the proposed method on motor imagery classification, systematic experiments are conducted on BCI competition II dataset III, BCI competition III dataset IVa and BCI competition IV dataset IIb. All of the datasets were recorded two-class EEG data of motor imagery experiments and measured at the international 10/20 systems. Details of the datasets are summarized in Table 1.

BCI competition II dataset III contains EEG data of left and right hand movements motor imagery from a normal subject. The experiment consists of 7 runs with 40 trials each (Fig. 4a) and all runs were conducted on the same day. The EEG signals were collected with 3 channels (C3, Cz, and C4) and filtered between [0.5, 30] Hz. Half of the trials were randomly selected as the training data and the remaining as the testing data.

Table 1
Summary of the datasets.

Dataset	Subject	Rate	Channels	Trials
Competition II-III	1	128 Hz	3	280
Competition III-IVa	5	1000 Hz	118	280
Competition IV-IIb	9	250 Hz	3	720

Table 2
Specific parameters of DDAF-CORAL.

Learning rate	Batch size	Epochs	Drop out	Early stop	N-fold
0.01	16	100	0.25	20	10

BCI competition III dataset IVa includes EEG data of right hand and foot movements motor imagery from 5 healthy subjects. The EEG signals were captured with 118 electrodes at a sampling rate of 1000 Hz and band-pass filtered between 0.05 and 200 Hz. It contained 288 EEG trials. The timing of the recording paradigm is shown in Fig. 4b.

BCI competition IV dataset IIb consists of EEG data from 9 healthy and right-handed subjects. All subjects were required to conduct five sessions at different times over a two-week period. The first two sessions include 120 trials each without feedback (Fig. 4c) and the last three sessions contain 160 trials each with smiley feedback (Fig. 4d). The EEG signals were recorded with 3 channels at 250 Hz and filtered by a band-pass filter from 0.5 Hz to 100 Hz and a notch filter at 50 Hz. During the EEG data acquisition, all subjects were requested to imagine left or right hand movements after the cue.

4.2. Experiment settings

In our experiments, within-session, cross-session, and cross-subject models were built to evaluate the generalization performance of the DDAF-CORAL respectively. And we separate the dataset into the training set and the testing set in the same way as the compared methods do. For within-session analysis, EEG data from BCI competition II dataset III were randomly divided into the training set and testing set, which corresponds to the competition's protocol of testing.

For cross-session analysis, EEG data of BCI Competition IV dataset IIb was used. For each subject, we used EEG data from the same subject but a different session as the source domain to train a model for the remaining sessions. Since session 3 follows the same experimental paradigms as sessions 4 and 5, which both involved feedback, the EEG data of session 3 was considered as the source domain, and sessions 4, 5 as the target domain. Time intervals of [4, 6]s, [4.25, 6.25]s, [4.5, 6.5]s, [4.75, 6.75]s, and [5, 7]s were extracted with moving windows for data augmentation as in [43] and trained to predict labels of [4, 6]s.

For cross-subject analysis, EEG data from BCI competition III dataset IVa was used. EEG data from one of the subjects was used as the source domain, and EEG data of the remaining 4 subjects were used as the target domain for label prediction respectively as in [32]. In other words, we performed the above transfer learning task four times on one subject and repeated it for $A_5^2 = 20$ times in total. To focus on the related information of motor imagery as well as reduce the size of the input EEG data, only 13 channels FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, Cpz, CP4 were used for assessment.

To obtain EEG data with a high signal-noise ratio (SNR) and improve the classification accuracy of motor imagery, we performed signal preprocessing as follows. First of all, we downsampled raw EEG data to 128 Hz to match the input size (C, 256) of the network. Following the settings in [44], EEG data downsampled was further band-pass filtered at [8, 35] Hz. Similar to the literature [45] for every

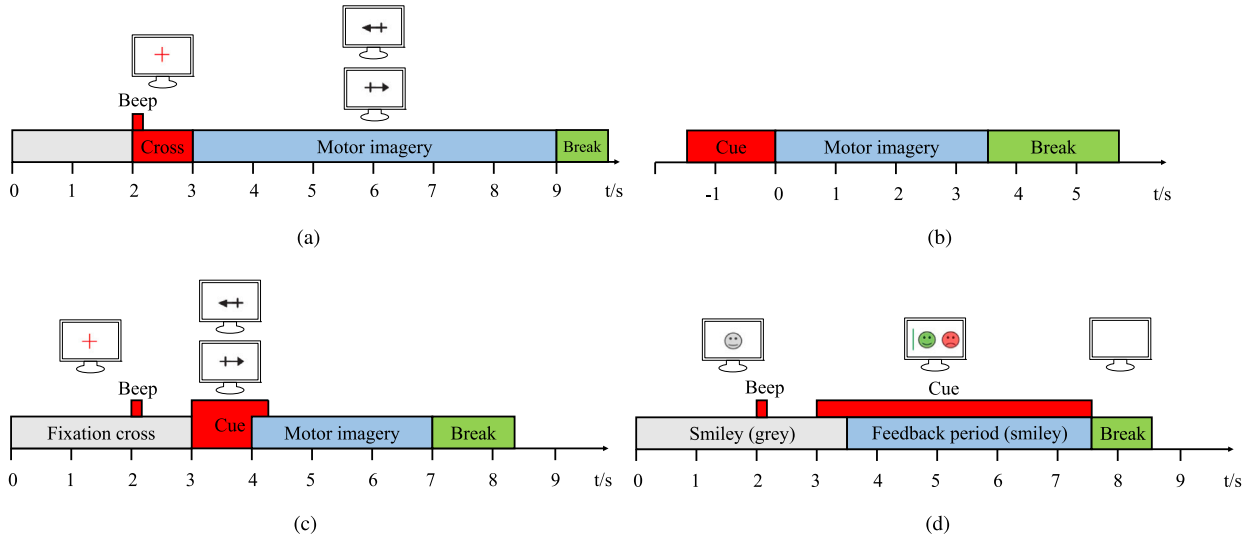

Fig. 4. Timing of recording paradigm for motor imagery experiments. (a) BCI competition II dataset III. (b) BCI competition III dataset IVa. (c) BCI competition IV dataset IIb without feedback. (d) BCI competition IV dataset IIb with smiley feedback.

EEG trial, we also used the time interval after the cue for analysis to remove the EEG data unrelated to motor imagery. Moreover, Independent Component Analysis (ICA) [46] was implemented to reduce the disturbances of EOG, ECG, EMG, and body movements. Finally, EEG data was standardized to [0, 1] by using the Min-Max strategy, which can improve the comparability of different EEG channels.

We used the cross-entropy loss function, and added an early stop to the model prediction. The remaining parameters are shown in Table 2. To access the performance of the proposed method, classification accuracy Acc and kappa coefficient κ are applied to describe the experimental results.

$$Acc = \frac{|\boldsymbol{x}_j^t : \boldsymbol{x}_j^t \in D_t \cap \hat{\boldsymbol{y}}_j^t = \boldsymbol{y}_j^t|}{|\boldsymbol{x}_j^t : \boldsymbol{x}_j^t \in D_t|} \tag{12}$$

$$\kappa = \frac{Acc - Acc_{rand}}{1 - Acc_{rand}} \tag{13}$$

In Eqs. (12) and (13), $\boldsymbol{x}_j^t$ is the target data, $\hat{\boldsymbol{y}}_j^t$ is the predicted label, $\boldsymbol{y}_j^t$ is the real label, and Acc_{rand} denotes the guess of random classification (0.5 for the binary classification). Additionally, the experiments were conducted based on Python 3.8 and PyTorch framework, and the hardware is Intel(R) Core(TM) i5-10400 2.90 GHz CPU with 16 GB RAM.

4.3. Overall classification performance

4.3.1. Within-session classification results on BCI competition II dataset III

Table 3 shows the classification performance of our proposed method on within-session analysis. As presented by the results, the proposed method achieves the highest classification accuracy of 92.9%. Compared with the winner's accuracy of the competition [47], our method outperforms theirs by 3.6%. We also compare it with the recently proposed methods and find that the CNN-SAE method is still inferior to our proposed method. Notably, the accuracy of DDAF-CORAL is increased by 7.2% than that of the EEGNet method, which greatly improves the generalization performance after domain adaptation with correlation alignment. In addition, based on convolutional neural networks (CNN), CNN-SAE, CBAM-CNN, and Dual-stream CNN all achieves great performance with an average accuracy of 90.0%, 90.7%, and 90.7% respectively.

Table 3
Within-session classification performance on BCI competition II dataset III.

Authors	Method	Accuracy	Kappa
Steven et al. [47]	Probabilistic model	0.893	0.786
Tabar et al. [48]	CNN-SAE	0.900	0.800
Lawhern et al. [49]	EEGNet	0.793	0.586
Chen et al. [50]	CBAM-CNN	0.907	0.814
Huang et al. [51]	Dual-stream CNN	0.907	0.814
Proposed	DDAF-CORAL	**0.929**	**0.858**

4.3.2. Cross-session classification results on BCI competition IV dataset IIb

We conducted comparison experiments for DDAF-CORAL and other state-of-the-art methods, including common spatial pattern (CSP) [52], filter bank common spatial pattern (FBCSP) [53], multi-scale convolutional neural network (MSCNN) [54], frequential deep belief network (FDBN) [55], parallel convolutional neural network (PCNN) [56], common spatial pattern and neighborhood component analysis (CSP-NCA) [57] and EEGNet [49]. For the cross-session analysis, the experimental setup follows that mentioned in Section 4.2. The kappa coefficient was used to evaluate the classification performance. The results are listed in Table 4 and the best performance of each subject is emphasized in boldface.

As it can be seen from Table 4, our proposed method is the one that achieves the best average kappa coefficient of 0.761, and the classification performance of our method is more effective than those comparison methods for subject 5-9. Although the PCNN and CSP-NCA obtain the best results for the other 4 subjects, both of the average kappa values are lower than that of our proposed, which are 0.659 and 0.680 respectively. Notably, the performance of PCNN is similar for all subjects, thus it reaches the lowest inter-subject standard deviation value of 0.048 among all methods, while the largest one of other comparison methods is 0.277. Furthermore, the traditional feature extraction algorithms CSP and FBCSP cannot perform well because of the EEG data distribution discrepancy of different time periods. Differently, CSP-NCA not only extracts spatial features using CSP, but also optimizes them based on neighborhood component analysis (NCA), which obtains similar features to decrease the domain discrepancy.

For MSCNN and FDBN, their kappa values are superior to CSP by 0.125 and 0.153 while inferior to our proposed method by 0.109 and 0.081, respectively. It indicates that even though deep learning

Table 4
Cross-session kappa value comparison with state-of-the-Art methods on BCI competition IV dataset IIb.

Subject	CSP	FBCSP	MSCNN	FDBN	PCNN	CSP-NCA	EEGNet	Proposed
B01	0.320	0.400	0.611	0.620	0.648	**0.712**	0.369	0.688
B02	0.240	0.220	0.309	0.300	**0.615**	0.326	0.157	0.463
B03	0.140	0.220	0.319	0.320	**0.602**	0.274	0.225	0.419
B04	0.940	0.960	0.986	0.960	0.646	**0.988**	0.813	0.963
B05	0.540	0.860	0.784	0.860	0.694	0.826	0.619	**0.931**
B06	0.500	0.620	0.722	0.760	0.621	0.584	0.263	**0.819**
B07	0.540	0.560	0.625	0.640	0.743	0.738	0.688	**0.744**
B08	0.860	0.860	0.776	0.880	0.648	0.888	0.863	**0.931**
B09	0.660	0.740	0.736	0.820	0.717	0.788	0.663	**0.894**
Avg	0.527	0.604	0.652	0.680	0.659	0.680	0.518	**0.761**
Std	0.268	0.277	0.220	0.239	**0.048**	0.244	0.267	0.203

Note: Avg indicates the average kappa value and Std indicates the standard deviation.

Table 5
Cross-subject classification accuracy comparison with state-of-the-Art methods on BCI competition III dataset IVa.

Task	TCA	JDA	DDC	RCSP	DDAN	MSDAN	EEGNet	Proposed
aa → al	0.852	0.832	0.911	0.768	0.939	**0.957**	0.800	0.946
aa → av	0.600	0.643	0.664	0.661	0.632	**0.686**	0.496	0.661
aa → aw	0.787	0.814	0.836	0.729	0.843	0.861	0.650	**0.921**
aa → ay	0.635	0.639	0.868	0.721	0.868	**0.936**	0.583	0.879
al → aa	0.757	0.771	0.825	0.707	0.843	0.821	0.689	**0.875**
al → av	0.591	0.625	0.664	0.657	0.625	**0.736**	0.621	0.700
al → aw	0.796	0.793	0.814	0.718	0.836	0.839	0.839	**0.975**
al → ay	0.826	0.811	0.907	0.846	0.871	**0.943**	0.768	0.904
av → aa	0.635	0.639	0.696	0.643	0.689	0.746	0.579	**0.782**
av → al	0.574	0.593	0.811	0.718	0.839	0.879	0.779	**0.900**
av → aw	0.596	0.668	0.682	0.686	0.718	0.804	0.789	**0.925**
av → ay	0.587	0.732	0.796	0.718	0.821	**0.907**	0.650	0.825
aw → aa	0.717	0.761	**0.775**	0.632	0.764	**0.775**	0.564	0.754
aw → al	0.770	0.804	0.886	0.757	0.925	0.929	0.818	**0.939**
aw → av	0.483	0.446	0.586	**0.657**	0.579	0.621	0.500	0.604
aw → ay	0.752	0.750	0.846	0.700	0.850	**0.896**	0.589	0.814
ay → aa	0.609	0.618	0.661	**0.704**	0.661	0.664	0.554	0.664
ay → al	0.783	0.771	0.864	0.904	0.871	0.932	0.821	**0.954**
ay → av	0.561	0.707	0.707	0.661	0.704	0.700	0.625	**0.732**
ay → aw	0.661	0.668	0.818	0.696	0.807	0.889	0.704	**0.907**
Avg	0.679	0.704	0.781	0.714	0.784	0.826	0.671	**0.833**
Std	0.105	0.098	0.096	0.067	0.128	0.104	0.112	0.119

Note: Avg indicates the average classification accuracy and Std indicates the standard deviation.

methods can extract the deep features from EEG data, they still cannot reduce the distribution divergence brought by EEG data during different time periods. Comparing the classification performance of EEGNet and the proposed, it is apparent to show that the distribution divergence of EEG data across sessions can be reduced by aligning the covariance of the source and target EEG data.

4.3.3. Cross-subject classification results on BCI competition III dataset IVa

To evaluate the cross-subject classification performance, the mainstream baseline methods were considered to compare with our proposed method. The experimental setup follows that mentioned in Section 4.2. Table 5 shows the results of classification accuracy on BCI Competition III dataset IVa. Overall, our proposed method obtains the best classification performance, achieving an average accuracy of 83.3%. And it can be clearly found that the proposed outperforms other state-of-the-art methods in most cross-subject tasks.

To be specific, the average classification accuracy of transfer component analysis (TCA) [35] and joint distribution adaptation (JDA) [36] on 20 tasks are 67.9% and 70.4% respectively. Compared with TCA, JDA achieved an improvement of 2.6% by minimizing the marginal distribution divergence and the joint distribution divergence simultaneously. Tzeng et al. [38] proposed deep domain confusion (DDC) by applying MMD to reduce the global distribution divergence between the source and target EEG data, which achieved an average accuracy of 78.1%. Regularized common spatial pattern (RCSP) [58] was employed to classify motor imagery EEG data with selected channels, and resulted in the best 65.7% and 70.4% classification accuracy on the tasks aw → av and ay → aa, respectively.

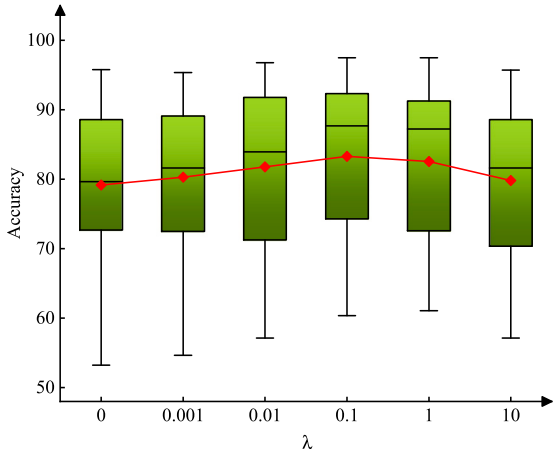

Fig. 5. Parameter sensitivity. The red line graph shows the variation of the average classification accuracy.

Similar to DDC, deep domain adaptation network (DDAN) [31] automatically extracts deep features from the raw EEG data based on ConvNet [59], and considered distribution matching and discriminant feature learning additionally. It is noteworthy that a multi-subdomain adaptation method (MSDAN) [32] was proposed to optimize the classification and adaptation losses in both class-related and time-related

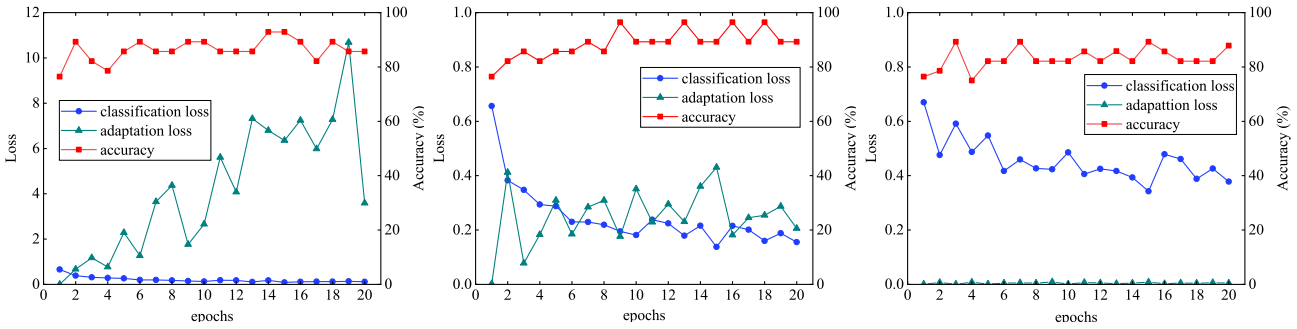

Fig. 6. Visualizations of the losses and the classification accuracy for cross-subject task aa→al.

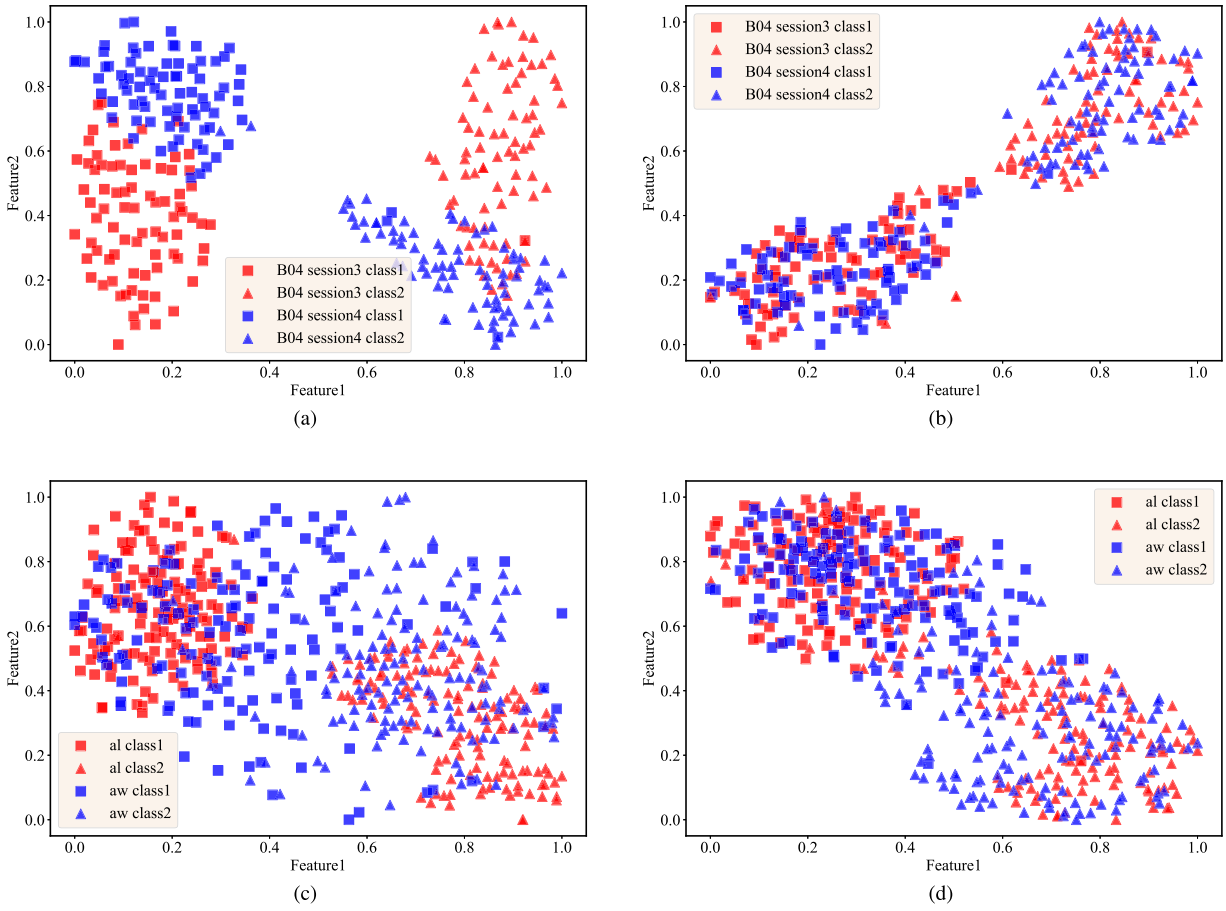

Fig. 7. Visualizations with t-SNE. (a) Feature distributions after EEGNet on cross-session task of B04 session 3 → session 4. (b) Feature distributions after DDAF-CORAL on cross-session task of B04 session 3 → session 4. (c) Feature distributions after EEGNet on cross-subject task of al → aw. (d) Feature distributions after DDAF-CORAL on the cross-subject task of al → aw.

subdomains, which reached up to the highest average classification accuracy in the published literature on dataset III-IVa. Different from the mentioned deep domain adaptation methods DDAN and MSDAN, our proposed DDAF-CORAL minimizes the feature covariance across subjects by correlation alignment after learning the deep features. Notably, our proposed method delivers an average improvement with a classification accuracy of 4.87% and 0.69%. Moreover, our proposed method obtains a dramatic improvement in the average accuracy of 16.21% against EEGNet, which indicates that DDAF-CORAL can effectively reduce the distribution divergence across subjects and improve the generalization ability of the model.

4.4. Parameter sensitivity

The DDAF-CORAL method has an essential parameter λ, which can adjust the classification loss and domain adaptation loss during the training process. To investigate the impact of λ on the classification performance, a series of parameter sensitivity experiments were conducted on the cross-subject model and the results are shown in Fig. 5. The red line graph in Fig. 5 shows the variation of the average classification accuracy for different values of λ. As the results show, the proposed method achieves the best average classification performance on all tasks with $\lambda = 0.1$. Besides, in Fig. 6, the classification loss, the

adaptation loss, and the classification accuracy are visualized for task aa → al when λ equals 0.001, 0.1, and 10, respectively. The comparison results suggest that the two losses are roughly the same and reach equilibrium when $\lambda = 0.1$. At the same time, the proposed model generalizes well on the target EEG data.

Additionally, we noticed that the best λ varies according to different tasks and the best λ cross-subject is generally larger than that of cross-session. Minimizing the classification loss alone might lead to overfitting on the source EEG data. While optimizing the coral loss itself is likely to lead to degenerated features, resulting in the poor performance of the classifier. Consequently, these two losses require to be balanced to achieve satisfactory classification performance.

4.5. Visualization

The feature distributions are visualized with t-SNE in Fig. 7. As can be seen from Fig. 7a, EEG features of B04 session 3 and session 4 after EEGNet are well separated, but the distribution divergence still remains. By aligning feature covariances across domains, our proposed method achieves an almost perfect feature adaptation between session 3 and session 4 as shown in Fig. 7b. Comparing Figs. 7c and 7d, we can definitely find that EEG features of subject al itself are well clustered, which can be classified easily. Although EEG features of the two classes of subject aw can also be distinguished from each other, it works tremendously poorly when testing EEG data of subject aw with the trained model of subject al. Meanwhile, this phenomenon confirms that the feature distribution divergence causes dramatic challenges on cross-subject tasks. Likewise, after features correlation alignment with our proposed, the performance of task subject al → aw is improved markedly. And our proposed method achieves an absolute average classification accuracy increase of 13.6% against EEGNet and MSDAN as shown in Table 5, reaching the largest average classification accuracy of 97.5% among all cross-subject tasks.

5. Conclusion

In this paper, we have proposed a deep domain adaptation framework with correlation alignment (DDAF-CORAL) for the EEG data distribution divergence caused by subject-related and time-related variations. A two-stage method shares the parameters of the source network with the target network and automatically extracts deep features for raw EEG data. Combined with the correlation alignment method, the classification loss and adaptation loss are optimized simultaneously in the training process, and the EEG distribution divergence is minimized by aligning the second-order statistics of the source and target features. Furthermore, we have conducted extensive experiments on BCI Competition II, III, and IV datasets for verification. The experimental results have shown that our proposed method exhibits the superiority in dealing with EEG data distribution divergence on cross-session and cross-subject classification, and outperforms other state-of-the-art methods. By visualizing the source and target feature distributions, the great generalization performance for EEG-based motor imagery classification is straightforwardly presented. Our feature works mainly focus on employing deep transfer learning methods to multi-class EEG-based motor imagery classification.

CRediT authorship contribution statement

Xiao-Cong Zhong: Conceptualization, Methodology. **Qisong Wang:** Writing – review & editing. **Dan Liu:** Writing – original draft. **Jing-Xiao Liao:** Visualization. **Runze Yang:** Supervision. **Sanhe Duan:** Data curation. **Guohua Ding:** Software, Validation. **Jinwei Sun:** Investigation.

Declaration of competing interest

All authors declare no conflicts of interest.

Acknowledgments

This work was sponsored by the Fundamental Research Funds for the Central Universities, China (Grant No. IR2021222), Future Science and Technology Innovation Team project of HIT (216506).

References

[1] Jacques J. Vidal, Toward direct brain-computer communication, Ann. Rev. Biophys. Bioeng. 2 (1) (1973) 157–180.
[2] Gert Pfurtscheller, Christa Neuper, Motor imagery and direct brain-computer communication, Proc. IEEE 89 (7) (2001) 1123–1134.
[3] Ravikiran Mane, Tushar Chouhan, Cuntai Guan, BCI for stroke rehabilitation: Motor and beyond, J. Neural Eng. 17 (4) (2020) 041001.
[4] Jingjing Luo, Qiying Cheng, Hongbo Wang, Youhao Wang, Qiang Du, Training therapy with BCI-based neurofeedback systems for motor rehabilitation, in: Recent Advances in AI-Enabled Automated Medical Diagnosis, CRC Press, 2022, pp. 268–280.
[5] Joseph N. Mak, Jonathan R. Wolpaw, Clinical applications of brain-computer interfaces: Current state and future prospects, IEEE Rev. Biomed. Eng. 2 (2009) 187–199.
[6] Lin Yang, Christopher Wilke, Benjamin Brinkmann, Gregory A. Worrell, Bin He, Dynamic imaging of ictal oscillations using non-invasive high-resolution EEG, Neuroimage 56 (4) (2011) 1908–1917.
[7] Saeid Sanei, Jonathon A. Chambers, EEG Signal Processing, John Wiley & Sons, 2013.
[8] Mohammad-Parsa Hosseini, Amin Hosseini, Kiarash Ahi, A review on machine learning for EEG signal processing in bioengineering, IEEE Rev. Biomed. Eng. 14 (2020) 204–218.
[9] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (1) (2007) 41–56.
[10] Amjed S. Al-Fahoum, Ausilah A. Al-Fraihat, Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains, Int. Sch. Res. Not. 2014 (2014).
[11] Abdulhamit Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (4) (2007) 1084–1093.
[12] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, Damien Coyle, Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (4) (2008) 317–326.
[13] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed Amine Mekhtiche, M. Shamim Hossain, Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, Future Gener. Comput. Syst. 101 (2019) 542–554.
[14] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, M. Shamim Hossain, Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification, Ieee Access 7 (2019) 18940–18950.
[15] Hamdi Altaheri, Ghulam Muhammad, Mansour Alsulaiman, Physics-inform attention temporal convolutional network for EEG-based motor imagery classification, IEEE Trans. Ind. Inform. (2022).
[16] Jan Berkhout, Donald O. Walter, Temporal stability and individual differences in the human EEG: An analysis of variance of spectral values, IEEE Trans. Biomed. Eng. (3) (1968) 165–168.
[17] Andreas Keil, Margarita Stolarova, Sabine Heim, Thomas Gruber, Matthias M. Müller, Temporal stability of high-frequency brain oscillations in the human EEG, Brain Topogr. 16 (2) (2003) 101–110.
[18] R.B. Paranjape, J. Mahovsky, L. Benedicenti, Z. Koles, The electroencephalogram as a biometric, in: Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No. 01TH8555), vol. 2, IEEE, 2001, pp. 1363–1366.
[19] James J. Gross, Oliver P. John, Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior, J. Personal. Soc. Psychol. 72 (2) (1997) 435.
[20] Tyler S. Lorig, Esther Huffman, Anthony DeMartino, Jay DeMarco, The effects of low concentration odors on EEG activity and behavior, J. Psychophysiol. (1991).
[21] Jun-Su Kang, Giljin Jang, Minho Lee, Stress status classification based on EEG signals, J. Inst. Internet, Broadcast. Commun. 16 (3) (2016) 103–108.
[22] Pasquale Marsella, Alessandro Scorpecci, Giulia Cartocci, Sara Giannantonio, Anton Giulio Maglione, Isotta Venuti, Ambra Brizi, Fabio Babiloni, EEG activity as an objective measure of cognitive load during effortful listening: A study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss, Int. J. Pediatric Otorhinolaryngol. 99 (2017) 1–7.
[23] Andreas Pedroni, Amirreza Bahreini, Nicolas Langer, Automagic: Standardized preprocessing of big EEG data, NeuroImage 200 (2019) 460–473.
[24] Vojkan Mihajlović, Bernard Grundlehner, The effect of force and electrode material on electrode-to-skin impedance, in: 2012 IEEE Biomedical Circuits and Systems Conference, BioCAS, IEEE, 2012, pp. 57–60.

[25] Kang Zhou, Lilong Cai, Study on effect of electrode force on resistance spot welding process, J. Appl. Phys. 116 (8) (2014) 084902.
[26] Budi Thomas Jap, Sara Lal, Peter Fischer, Evangelos Bekiaris, Using EEG spectral components to assess algorithms for detecting fatigue, Expert Syst. Appl. 36 (2) (2009) 2352–2359.
[27] Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti S. Hämäläinen, Matthieu Kowalski, Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations, NeuroImage 70 (2013) 410–422.
[28] Yan Li, Hiroyuki Kambara, Yasuharu Koike, Masashi Sugiyama, Application of covariate shift adaptation techniques in brain–computer interfaces, IEEE Trans. Biomed. Eng. 57 (6) (2010) 1318–1324.
[29] Simanto Saha, Khawza Iftekhar Uddin Ahmed, Raqibul Mostafa, Leontios Hadjileontiadis, Ahsan Khandoker, Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2) (2017) 371–382.
[30] Karsten M. Borgwardt, Arthur Gretton, Malte J. Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, Alex J. Smola, Integrating structured biological data by kernel maximum mean discrepancy, Bioinformatics 22 (14) (2006) e49–e57.
[31] Wenlong Hang, Wei Feng, Ruoyu Du, Shuang Liang, Yan Chen, Qiong Wang, Xuejun Liu, Cross-subject EEG signal recognition using deep domain adaptation network, IEEE Access 7 (2019) 128273–128282.
[32] Yi Chen, Rui Yang, Mengjie Huang, Zidong Wang, Xiaohui Liu, Single-source to single-target cross-subject motor imagery classification based on multisubdomain adaptation network, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1992–2002.
[33] Bianca Zadrozny, Learning and evaluating classifiers under sample selection bias, in: Proceedings of the Twenty-First International Conference on Machine Learning, 2004, p. 114.
[34] Corinna Cortes, Mehryar Mohri, Michael Riley, Afshin Rostamizadeh, Sample selection bias correction theory, in: International Conference on Algorithmic Learning Theory, Springer, 2008, pp. 38–53.
[35] Sinno Jialin Pan, Ivor W. Tsang, James T Kwok, Qiang Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Netw. 22 (2) (2010) 199–210.
[36] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, Philip S. Yu, Transfer feature learning with joint distribution adaptation, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2200–2207.
[37] Basura Fernando, Amaury Habrard, Marc Sebban, Tinne Tuytelaars, Unsupervised visual domain adaptation using subspace alignment, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2960–2967.
[38] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, Trevor Darrell, Deep domain confusion: Maximizing for domain invariance, 2014, arXiv preprint arXiv:1412.3474.
[39] Jindong Wang, Yiqiang Chen, Wenjie Feng, Han Yu, Meiyu Huang, Qiang Yang, Transfer learning with dynamic distribution adaptation, ACM Trans. Intell. Syst. Technol. 11 (1) (2020) 1–25.
[40] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, Qing He, Deep subdomain adaptation network for image classification, IEEE Trans. Neural Netw. Learn. Syst. 32 (4) (2020) 1713–1722.
[41] Jian-Feng Cai, Emmanuel J. Candès, Zuowei Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2010) 1956–1982.
[42] Baochen Sun, Jiashi Feng, Kate Saenko, Return of frustratingly easy domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, (no. 1) 2016.
[43] Mouad Riyad, Mohammed Khalil, Abdellah Adib, MI-EEGNET: A novel convolutional neural network for motor imagery classification, J. Neurosci. Methods 353 (2021) 109037.
[44] Smita Tiwari, Shivani Goel, Arpit Bhardwaj, MIDNN-A classification approach for the EEG based motor imagery tasks using deep neural network, Appl. Intell. 52 (5) (2022) 4824–4843.
[45] Ji-Seon Bang, Min-Ho Lee, Siamac Fazli, Cuntai Guan, Seong-Whan Lee, Spatio-spectral feature representation for motor imagery classification using convolutional neural networks, IEEE Trans. Neural Netw. Learn. Syst. (2021).
[46] Pierre Comon, Independent component analysis, a new concept? Signal Process. 36 (3) (1994) 287–314.
[47] S. Lemm, C. Schafer, G. Curio, BCI competition 2003-data set III: Probabilistic modeling of sensorimotor /spl mu/ rhythms for classification of imaginary hand movements, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1077–1080.
[48] Yousef Rezaei Tabar, Ugur Halici, A novel deep learning approach for classification of EEG motor imagery signals, J. Neural Eng. 14 (1) (2016) 016003.
[49] Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P Hung, Brent J. Lance, EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural Eng. 15 (5) (2018) 056013.
[50] Zhongye Chen, Yijun Wang, Zhongyan Song, Classification of motor imagery electroencephalography signals based on image processing method, Sensors 21 (14) (2021) 4646.
[51] E. Huang, X. Zheng, Y. Fang, Z. Zhang, Classification of motor imagery EEG based on time-domain and frequency-domain dual-stream convolutional neural network, IRBM 43 (2) (2022) 107–113.
[52] Haixian Wang, Qin Tang, Wenming Zheng, L1-norm-based common spatial patterns, IEEE Trans. Biomed. Eng. 59 (3) (2011) 653–662.
[53] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, Cuntai Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 2390–2397.
[54] Xianlun Tang, Wei Li, Xingchen Li, Weichang Ma, Xiaoyuan Dang, Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network, Expert Syst. Appl. 149 (2020) 113285.
[55] Na Lu, Tengfei Li, Xiaodong Ren, Hongyu Miao, A deep learning scheme for motor imagery classification based on restricted Boltzmann machines, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2016) 566–576.
[56] Yuexing Han, Bing Wang, Jie Luo, Long Li, Xiaolong Li, A classification method for EEG motor imagery signals based on parallel convolutional neural network, Biomed. Signal Process. Control 71 (2022) 103190.
[57] N.S. Malan, S. Sharma, Motor imagery EEG spectral-spatial feature optimization using dual-tree complex wavelet and neighbourhood component analysis, IRBM 43 (3) (2022) 198–209.
[58] Haiping Lu, How-Lung Eng, Cuntai Guan, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos, Regularized common spatial pattern with aggregation for EEG classification in small-sample setting, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2936–2946.
[59] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, Tonio Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Human Brain Mapp. 38 (11) (2017) 5391–5420.