Neurocomputing 421 (2021) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

A review on transfer learning in EEG signal analysis ☆

Zitong Wan [a,d], Rui Yang [b,c,*], Mengjie Huang [a,*], Nianyin Zeng [e], Xiaohui Liu [f]

[a] Design School, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[b] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[c] Research Institute of Big Data Analytics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[d] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, United Kingdom
[e] Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen 361005, China
[f] Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, United Kingdom

ARTICLE INFO

Article history:
Received 24 May 2020
Revised 16 July 2020
Accepted 7 September 2020
Available online 29 September 2020
Communicated by Ma Lifeng Ma

Keywords:
Electroencephalogram
Transfer learning

ABSTRACT

Electroencephalogram (EEG) signal analysis, which is widely used for human-computer interaction and neurological disease diagnosis, requires a large amount of labeled data for training. However, the collection of substantial EEG data could be difficult owing to its randomness and non-stationary. Moreover, there is notable individual difference in EEG data, which affects the reusability and generalization of models. For mitigating the adverse effects from the above factors, transfer learning is applied in this field to transfer the knowledge learnt in one domain into a different but related domain. Transfer learning adjusts models with small-scale data of the task, and also maintains the learning ability with individual difference. This paper describes four main methods of transfer learning and explores their practical applications in EEG signal analysis in recent years. Finally, we discuss challenges and opportunities of transfer learning and suggest areas for further study.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

With the maturity of technology and science, brain science becomes an important research field for humans to explore the mysteries of life. Brain science focuses on studying the structure and mechanism of the brain, which can be traced back to the 1960s, and has received consistent attention worldwide in the past decades [1]. To analyze the electrical activities of brain, electroencephalogram (EEG) is applied widely in different directions of brain science.

EEG signals can be measured by placing electrodes on the surface of the scalp and collected by brain-computer interface (BCI) systems, which are generated through the cortical nerve cell inhibitory and excitatory postsynaptic potentials [2]. BCI is a direct communication and control channel established between the human brain and computers or other electronic devices [3]. Fig. 1 summarizes the applications of EEG signals for human-computer interaction and neurological disease diagnosis. Commonly used EEG signals in BCI studies include P300, visual evoked potential (VEP), steadystate visually evoked potential (SSVEP) and slow cortical potential (SCP), event-related desynchronization (ERD) and event-related synchronization (ERS) [4]. EEG provides assistance in two main aspects: human-computer interaction and neurological disease diagnosis. EEG signals help the human-computer interaction systems in speech recognition [5], emotion detection [6], user experience evaluation [7] and rehabilitation [8]. It also serves as a useful tool for doctors and researchers to diagnose diseases associated with brain dysfunction, such as Alzheimer's disease [9], epilepsy [10], schizophrenia [11], cerebral palsy [12] and mild cognitive impairment [13].

To implement the applications mentioned above, the following characteristics of EEG signals must be considered in EEG signal feature extraction to ensure that the features closely related to the task are extracted. First, the brain signals are complex, which lead to non-stationary, non-linear and non-Gaussian in EEG signals [2]. Second, the characteristics of EEG signals are highly subject to various individual differences such as age and mentality and the EEG signals of different individuals may be different for the same event [14]. Last, EEG signals are highly noisy and susceptible to be distorted by artificial interference. Hence, it is difficult to extract effective information related to specific tasks from EEG signals [15].

☆ This document is partially supported by National Natural Science Foundation of China (61603223), Key Program Special Fund in XJTLU (KSF-E-34) and Research Development Fund of XJTLU (RDF-18-02-30).

* Corresponding authors at: School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (R.Yang); Design School, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (M. Huang).

E-mail addresses: R.Yang@xjtlu.edu.cn (R. Yang), Mengjie.Huang@xjtlu.edu.cn (M. Huang).

https://doi.org/10.1016/j.neucom.2020.09.017
0925-2312/© 2020 Elsevier B.V. All rights reserved.

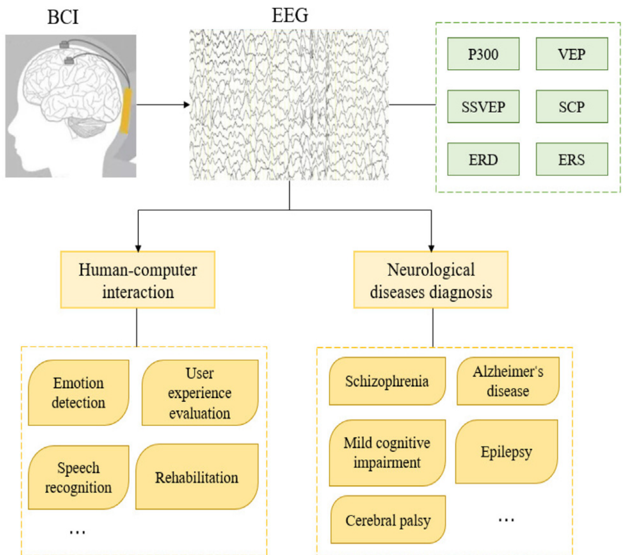


Fig. 1. Applications of EEG Signals.

According to the characteristics mentioned above, the analysis of EEG signal faces two main difficulties:

1) *Individual Difference*: In the process of feature extraction and classification, individual difference brings difficulties when the pre-trained model is applied in new tasks directly. Gianotti first demonstrated the neural signatures underlying individual differences in the feeling of being looked at through measuring the width of the cone of direct gaze in 137 healthy participants [16]. Matthews et al. found that there were large differences in five indicators reflecting EEG signal in psychometric measurements through an experiment on 150 patients, who completed psychological stress tests in a complex simulated operating environment [17].
2) *Insufficient Information*: Due to the different types of information in EEG signal and its non-stationary and low signal-to-noise ratio, it is difficult to filtrate the information sensitive to a specific task. Furthermore, the problem of insufficient data arises when applying EEG signals to new fields, especially for the emerging and interdisciplinary fields [18]. Dutta and Nandy reported that the lack of available EEG data sets made it difficult to estimate mental stats with deep learning models [19]. Luo et al. held the view that data deficiency is adverse to emotion recognition with high complexity [20].

In order to solve the above mentioned problems, traditional algorithms in EEG signal analysis are improved, and new methods are also developed by researchers. Generally, there are three main stages in traditional EEG signal analysis: first, the features of original signals are extracted; second, features related to the task are selected and reserved; third, a classifier is built to classify the selected feature set. In feature extraction, other types of features, such as connectivity features and high order statistics are also explored in addition to the traditional frequency band power features and time point features [21,22]. Moreover, several studies combined different types of features to extract more information [23,24]. In feature selection, apart from using filter to remove the features that are irrelevant to the task, wrapper and some sorting methods are developed to choose the feature subset [25,26]. Finally, in feature classification, deep learning, Riemann geometry and other methods are applied to improve the accuracy of classifier from different perspectives [27–30]. Some algorithms extract and classify the features jointly, such as neural networks and embedded approaches [31,32]. In most traditional methods, it is generally assumed that the training data and the testing data have the same distribution. However, this assumption usually cannot be established in practical applications. In this case, the expensive cost of model reconstruction and training data recollection obstructs the development of algorithms [33].

Transfer learning solves this problem by adjusting the model via priori knowledge to make it adaptable in new tasks. Apart from this, as a common situation in EEG signal analysis tasks, data scarcity makes model training difficult [34]. Transfer learning model still keeps learning ability without large amount of data based on priori knowledge learnt in related tasks. Transfer learning focuses on applying the knowledge learnt in one domain into a different but related domain, which was first proposed at the 1995 NIPS-95 seminar on 'Learning to Learn' [14,35,36]. In 2005, the Defense Advanced Research Projects Agency gave a new definition of transfer learning: the ability of the system to recognize and apply the knowledge and skills learnt in previous tasks to a new task [35]. Unlike traditional machine learning, the focus of transfer learning is on the target task with one or multiple source tasks which are trained to provide priori knowledge for the target task. Transfer learning has been applied in a variety of fields, such as data mining, image recognition, language translation, fault diagnosis and positioning system [37–40]. The advantages of transfer learning in EEG signal analysis are as follows:

1) *Match individual difference*: In EEG signal processing tasks, the difference between the training and testing data is huge, such as subjects, sampling time and task objectives [41], which increases the difficulty in analysis as explained before. Transfer learning can make the model flexibly matching different individuals and tasks through adjustments. Researchers started to propose algorithms that are adaptive to different subjects and individuals [42,43].
2) *Reduce data requirement*: In EEG signal analysis, problems of data scarcity and insufficient labeling hinder the learning of the target task [34,2]. Transfer learning learns the target task according to priori knowledge learnt in a similar domain by a small amount of data in the target domain to adjust the classifier, which reduces the requirements for available data. Different transfer learning models are proposed to solve the problem of small training trials in various applications [44,45].

With the persistent study of transfer learning, various transfer learning algorithms were developed in EEG signal analysis, which also obtained satisfactory results in experiments. This paper attempts to provide a brief summary of the literature relating to the applications of transfer learning algorithms in EEG signal analysis so far to the researchers in related fields. This paper generalizes the frequently used algorithms of transfer learning with the practical applications in EEG signal analysis. The challenges and development prospects of transfer learning in EEG signal analysis are also discussed.

The rest of the paper is organized as follows. In Section 2, the basics of transfer learning are introduced in detail. The transfer learning methods frequently used in EEG signal analysis are illustrated in Section 3. Section 4 discusses the status and development of transfer learning in EEG signal analysis. Finally, conclusion of this review is made in Section 5.

2. Preliminaries

Before explaining the transfer learning methods in EEG signal analysis, the basics of transfer learning are introduced in this section, including the definition and categories of transfer learning.

2.1. Definition of transfer learning

The concepts adopted in transfer learning are explained in this subsection and the notations are shown in Table 1.

1) *Domain*: A *domain* $\mathcal{D}$ consists of the feature space $\mathcal{X}$ of n dimensions and the probability distribution $P(X)$ of $\mathcal{X}$, where $X = \{x_1, x_2, \cdots, x_n\} \in \mathcal{X}$. In transfer learning, the domain containing known knowledge is called *source domain*, which is usually represented by $\mathcal{D}_S$; and the domain containing unknown knowledge to be learnt is called *target domain*, which is usually represented by $\mathcal{D}_T$.
2) *Task*: A *task* is a learning goal, which consists of the label space $\mathcal{Y}$ and the prediction function $f(\cdot)$ (also written as $P(Y|X)$ in probability theory, which means the probability of Y under condition X), where $Y = \{y_1, y_2, \cdots, y_n\} \in \mathcal{Y}$. According to the definition of task, the label space of the source domain and the target domain are represented as $\mathcal{Y}_S$ and $\mathcal{Y}_T$.
3) *Transfer learning*: When there is difference between domains or tasks, the knowledge learnt in the source domain can be transferred to the target domain through transfer learning. It is worth noting that the precondition of transfer learning is that the domains or tasks must be different but similar to some extent. In specific problems, the knowledge learnt is usually labeled as Y. Labels and their corresponding features make up *data* in this domain, which is denoted as $D = \{(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)\}$. With the above introduction, transfer learning can be described formally with the following definition:

Definition Given a source domain $\mathcal{D}_S$ and a target domain $\mathcal{D}_T$, $\mathcal{T}_S$ and $\mathcal{T}_T$ are the tasks of $\mathcal{D}_S$ and $\mathcal{D}_T$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$. The goal of transfer learning is to apply the knowledge in $\mathcal{D}_S$ to help learn the knowledge in $\mathcal{D}_T$, where $\mathcal{D}_S = \{X_S, P(X_S)\}, \mathcal{D}_T = \{X_T, P(X_T)\}, \mathcal{T}_S = \{Y_S, P(Y_S|X_S)\}$ and $\mathcal{T}_T = \{Y_T, P(Y_T|X_T)\}$.

In this definition, the condition $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$ is the problem that the transfer learning approaches mainly try to solve. The common idea of transfer learning is to reduce the difference between domains or tasks to ensure the corresponding feature or label space similar. The examples in Fig. 2 explain the possible conditions in detail.

The top half of Fig. 2 gives an example of $\mathcal{D}_S \neq \mathcal{D}_T$. The learning aim is to apply the knowledge of the emotion EEG signals of the left girl to detect the emotion of the right girl, which means the feature spaces of two domains are different. The bottom half of Fig. 2 gives an example of $\mathcal{T}_S \neq \mathcal{T}_T$. In this example, both of the feature spaces of two domains are the emotion EEG signals of the girl. However, the task in the source domain is to detect the emotion 'glad' (marked in red); the task in the target domain is to detect the emotion 'delightful' (marked in red), which is similar to 'glad'. Especially, when $\mathcal{D}_S = \mathcal{D}_T$ and $\mathcal{T}_S = \mathcal{T}_T$, the definition describes the situation of traditional machine learning.

Table 1
Notations in Transfer Learning in the Paper.

Notation	Description
$\mathcal{D}$	Domain
$\mathcal{T}$	Task
$\mathcal{X}$	Feature space
$\mathcal{Y}$	Label space
$P(X)$	Marginal distribution
$P(Y\|X)$	Conditional distribution
D_S	Source domain data
D_T	Target domain data

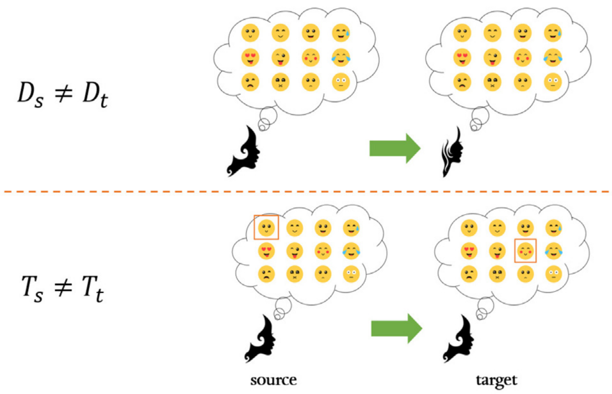


Fig. 2. Examples of Difference in Transfer Learning.

2.2. Categories of transfer learning

There is no standard and unified method for the categories of transfer learning so far. This subsection presents the categories of transfer learning by label, feature and learning style, which are all the commonly used metrics to help readers to understand transfer learning from multiple perspectives. The structure of categories of transfer learning is shown in Fig. 3.

2.2.1. Sort by label types

Transfer learning can be divided into the following three categories depending on the known data labels of the source and target domains: inductive transfer learning, transductive transfer learning and unsupervised transfer learning.

a) *Inductive transfer learning*: In inductive transfer learning, the labels of the target domain are known, and the target task and the source task are different ($\mathcal{D}_S \neq \mathcal{D}_T$). Most inductive transfer learning studies are under the situation that the source domain and the target domain labels are both known [46,47]. In terms of the situation that the source domain labels are unknown, Raina et al. proposed a self-taught learning framework for using unlabeled data in supervised tasks [48]. In this case, the system can be regarded as transfer learning in a state where the labels of the source domain is unavailable.
b) *Transductive transfer learning*: In transductive transfer learning, the labels of the target domain are unknown, and the source domain has a large amount of labeled training data available. Meanwhile, the tasks of the target and source are

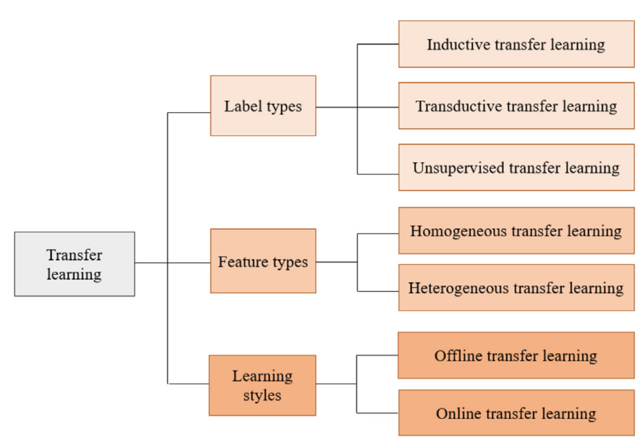


Fig. 3. The Structure of Categories of Transfer Learning.

the same but with difference in the domains ($\mathcal{T}_S = \mathcal{T}_T$ and $\mathcal{D}_S \neq \mathcal{D}_T$). As the difference between domains are reflected in feature spaces or marginal distributions, transductive transfer learning reduces this difference to predict labels in the target domain [49,50].

c) *Unsupervised transfer learning*: In unsupervised transfer learning, the data in the source domain and the target domain are both unlabeled, and the target and source tasks are different ($\mathcal{D}_S \neq \mathcal{D}_T$). Unsupervised transfer learning utilizes the correlation between the target domain and the source domain through unsupervised algorithms to learn the target task [51–53].

2.2.2. Sort by feature types

According to the meaning of the features in the source domain and the target domain, transfer learning can be divided into homogeneous transfer learning and heterogeneous transfer learning [35]. In homogeneous transfer learning, the semantics and dimensions of the feature space in the source domain and the target domain are the same. On the contrary, in heterogeneous transfer learning, the semantics and dimensions of the feature set in the source domain and the target domain are not exactly the same. For example, news categorization from different public service broadcasters is a typical multi-source homogeneous transfer learning task, but image categorization from truck head, container, wheel to truck is a heterogeneous transfer learning task [54].

2.2.3. Sort by learning styles

According to the learning styles, transfer learning can be divided into offline transfer learning and online transfer learning.

a) *Offline transfer learning:* In offline transfer learning, the source domain and the target domain are both settled, in which case the aim of learning is to complete only one time knowledge transfer to finish the model adjustment. For example, several literatures on EEG signal analysis used public P300 public database as experimental data [55,56]. After training with this historical data using offline transfer learning, the model may not perform well on other data sets due to lack of online updates.

b) *Online transfer learning:* Online transfer learning updates the model as new data are generated [57] and the model obtained by this learning style is more adaptable. The distribution difference is hard to measure, because the data in the target domain are transported dynamically and processed in real-time, which is more challenging than offline transfer learning [58]. Some specific problems are explored by online transfer learning algorithms, such as regression problem of driver drowsiness detection [59].

3. Transfer learning methods in EEG signal analysis

In the previous section, the definition and categories of transfer learning have been explained. This section introduces the transfer learning methods commonly used in EEG signal analysis in this survey. In the EEG signal processing, transfer learning is applied in feature extraction and classification. There are two main common approaches: 1) the classical algorithms in EEG signal analysis are improved based on transfer learning; 2) the algorithms in transfer learning are applied in EEG signal analysis. The algorithms and literature mentioned in this section are listed in Table 2.

3.1. Domain adaptation

The idea of domain adaptation in transfer learning aims to improve the models to adapt to the data distribution in the target domain. Domain adaptation is a broad concept with a variety of algorithms. Thus, this paper classifies them according to the difference of the distribution in the source domain and the target domain. The algorithms in domain adaptation can be divided to marginal distribution adaptation and conditional distribution adaptation.

3.1.1. Marginal distribution adaptation

The aim of marginal distribution adaptation is to transfer the knowledge when the marginal distributions of the source domain and the target ($P(X_S) \neq P(X_T)$) are different. The left of Fig. 4 shows

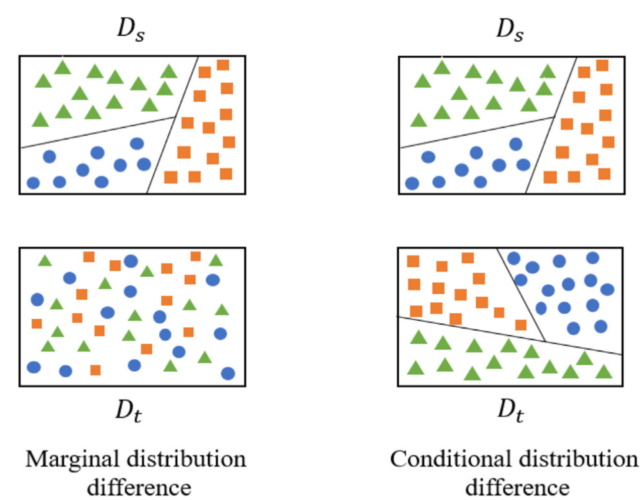


Fig. 4. Marginal Distribution Difference and Conditional Distribution Difference.

Table 2
Transfer Learning Methods in EEG Signal Analysis.

Transfer learning method	Algorithm		Literature
Domain adaptation	Marginal distribution adaptation	Multi-task learning	[68,70,66,69,72,71]
		Single-task learning	[65,61,63,64,60,62]
		Joint feature selection	[79,76,78,73,74,77,75]
	Conditional distribution adaptation		[82,81,80]
Improved CSP algorithms	Regularized CSP		[44,89,86,88,87]
	Other improved CSP algorithms		[91,90,15,92]
Deep neural networks	Fine-tuning		[97,42,115,116,95,102,98,96,99–101]
	Deep network adaptation		[103,107,105,106,104]
	GAN transfer		[110,113,114,112,111]
Subspace learning	Manifold learning	Riemannian manifold	[117–123]
		Locally linear embedding	[124–128]
		Other manifold	[129–131,98,132,133]
	Statistical feature alignment		[134–137]

an example of marginal distribution difference: the three classes (represented by blue, orange and green) in D_S are clearly divided, which is largely different from the randomly distributed classes in D_T. As a common situation in EEG signal analysis tasks, marginal distribution adaptation is the focus of most domain adaptation literature currently.

Marginal distribution adaptation made achievements in several cross-subject tasks. Santana et al. proposed a cross-subject classifier to predict the stimulus presented to a subject from the analysis of the brain activity [60]. Dai et al. proposed a framework called domain transfer multiple kernel boosting (DTMKB) through applying Adaboost in learning kernel classifier with the transfer of multiple kernels. DTMKB divided the kernel classifier into a number of base classifiers, then repeatedly learnt the base classifiers with multiple kernels. At each trial, the weights of the wrongly classified examples were increased while the weights of the correctly classified examples were decreased [61]. Zhang et al. built a cross-subject model to reduce the distribution difference through an inter-domain scatter matrix [62]. Jiang et al. added a label shift vector into a generalized linear model to adjust the weights in LSR classifier [63]. Liu et al. proposed a cross-subject framework based on transfer component analysis (TCA) to make comparison with logistic regression and deep learning classifiers, which indicated that the performance of TCA was better than the two classical algorithms in EEG signal analysis [64]. Chai et al. explored a fast domain adaptation strategy that learnt a linear transformation function to match the marginal distribution of the source and target subspaces without a regularization term [65].

Marginal distribution adaptation is suitable in multi-task learning with multiple source domains corresponding to one target domain. Lan et al. applied maximum independence domain adaptation (MIDA) into intra-dataset and inter-dataset multiple tasks on emotion classification [66]. MIDA measured the maximum independence between the target data and their features by the Hilbert-Schmidt independence criterion (HSIC) and covariance [67]. In intra-dataset experiment, they set one subject in the dataset as the target domain data and the other subjects as the source domain data for training. Lan assumed each subject as one domain, which resulted in multiple source domains. In inter-dataset experiment, the training data and the testing data were provided by different data sets. The experimental result showed that MIDA improved the accuracy on two public data sets compared to the baseline accuracy without domain adaptation [66]. He et al. also proposed a transfer learning framework based on boosting algorithm to learn the cross-subject task [68]. Li et al. proposed a multi-task learning model based on style transfer mapping for emotion recognition, and tested it into both supervised and semi-supervised learning [69]. Jayaram et al. explored a cross-subject multi-task classifier that decomposed the weights of features to components to learn the Gaussian distribution parameters for multi-task systems [70]. Zheng et al. proposed a cross-subject affective model by applying eye tracking data for calibration to transfer heterogeneous knowledge [71]. Xiao et al. built groups by comparing the emotion signals from different people and calculated the similarity of physiological signals between two individuals to divide groups for the robots to express emotions to different groups of people [72].

Joint feature selection is effective to decrease the marginal distribution difference between the source domain and the target domain. The basic assumption of joint feature selection is to build models based on the shared features in both source domain and target domain, in which the distributions of the two domains are the same. Joadder et al. proposed a feature fusion algorithm to determine the best set of features for motor imagery tasks [73]. Nakanishi et al. extracted shared responses from different devices to obtain the shared data of SSVEP [74]. Xie et al. connected the source domain and the target domain with common hyperplanes for different classes in the two domains [75]. Chen et al. combined several filtering algorithms to evaluate the joint features in different subjects to detect driver status [76]. Raza et al. employed an exponentially weighted moving average model to detect the covariate shifts in the features from motor imagery (MI) related brain responses and added new classifiers over time [77]. Lin and Jung explored a conditional transfer learning (cTL) framework for emotion classification to apply the data with similar feature space selectively by evaluating the transfer ability of every sample [43]. Fauzi et al. proposed a transfer learning framework by extracting a common matrix of the subjects and historical EEG data to MI tasks [78]. Banerjee et al. calculated three types of speech features and fused these features by a deep belief network model for post-traumatic stress disorder [79].

3.1.2. Conditional distribution adaptation

The aim of conditional distribution adaptation is to transfer the knowledge when the conditional distribution of the source domain and the target domain ($P(Y_S|X_S) \neq P(Y_T|X_T)$) are different. The right of Fig. 4 shows an example of conditional distribution difference: the three classes are distributed with no significant difference in both D_S and D_T, in which the learning task of the two domains can be different. The studies on conditional distribution adaptation are fewer than marginal distribution adaptation so far owing to its high learning difficulty.

Khalaf and Akcakaya introduced a transfer learning model by using Bhattacharyya distance to measure the similarity of the conditional distributions between new users and existing users [80]. García-Salinas et al. built a model to collect the features of some original vocabulary (codewords) by imagined speech and a new imagined word could be represented with these codewords via conditional distribution adaptation [81]. Dagois et al. applied the scores of regularized discriminant analysis of 10 individuals to learn the class conditional distribution across different subjects for session calibration in MI tasks[82].

3.2. Improved common spatial patterns algorithms

Common spatial patterns (CSP) is a successful algorithm in building spatial filter to extract frequency band power features [83]. In order to increase the adaptation of CSP to the problems of insufficient labels and large feature space difference, researchers improved CSP from different aspects.

In a given frequency band range, the variance of filtered EEG signals corresponds to the signal power [84]. The optimal set of spatial filters is leant by CSP to maximize the variance of filtered EEG signals from one channel, and minimize the variance from another channel [85,84]. In Fig. 5, the data from channel 1 and channel 2 belong to the multiple-channel input data and R_1 and R_2 are the covariance matrices of data from channel 1 and channel 2. To calculate the optimal spatial filter f, the inter-channel maximum covariance and the intra-channel minimum covariance can be calculated through common activity matrix R ($R = R_1 + R_2$). In

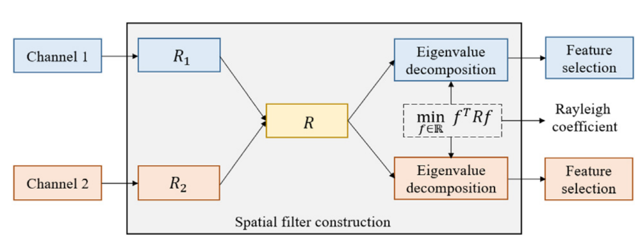


Fig. 5. Spatial Filter Construction through CSP.

order to facilitate calculation, the key is to compute into the minimum value of Rayleigh coefficient through eigenvalue decomposition. Through the optimal spatial filter, the features are extracted for further analysis.

3.2.1. Regularized common spatial patterns

Regularized common spatial patterns (RCSP) is an effective improvement on CSP model, which adjusts the objective function by adding a penalty for regularization in the cost function [86]. Xu et al. explored a type of RCSP independent of covariance matrix and compared the feature space of multiple source domains and one target domain to estimate the change of feature vectors of the target domain after transfer [87]. Cho et al. applied RCSP into cross-session data sets and found that RCSP performed well in a zero-training task [88]. Blankertz et al. proposed invariant CSP (iCSP) to reduce the adverse effect from individual factors and noise [89]. Azab et al. reported a dynamic time warping regularized CSP for small-scale data sets, which applied dynamic time warping in measuring the dissimilarity between two EEG segments after being aligned [44].

3.2.2. Other improved common spatial patterns algorithms

Apart from the RCSP family, CSP can be improved by several other techniques. Kang et al. applied two methods linearly to determine a composite CSP model, which increased the classification accuracy on small-scale data set [15]. Kang and Choi proposed a Bayesian CSP model for cross-subject learning to explore the relationship of subjects through assuming that spatial patterns across subjects share the same latent subspace [90]. Azab et al. proposed a improved CSP algorithm which assigned different weights to subjects by using KL divergence to measure the similarity of different subjects [91]. Samek et al. explored a stationary subspace CSP method to remove the subspace containing the principal non-stationary directions for most subjects before feature extraction [92].

3.3. Deep neural networks

Deep neural network (DNN) is a machine learning method that constructs a network model by simulating the structure of neural cells [93]. DNN consists of an input layer, an output layer, and several hidden layers, and can directly learn the data after normalization to get the features and classifier jointly. Different types of DNNs are applied to EEG signal analysis, such as convolutional neural network (CNN), recurrent neural network (RNN) and autoencoder [94,95].

Each layer of DNN consists several nodes, and the output value of the node in the upper layer is regarded as the input of the node in the next layer. Each node is given different weights to adjust their importance to the classification result. The normalized EEG signals are sent to the input layer to obtain the classification results by learning the parameters of each layer of nodes. This process is repeated for a certain number of trails until a satisfactory classification accuracy.

3.3.1. Fine-tuning

The most commonly used transfer learning method based on DNN in EEG signal analysis is to fine-tune a pre-trained DNN model when there is no significant difference in the source domain and the target domain. Fig. 6(a) gives examples of a CNN fine-tuning from two aspects: to fine-tune the whole network and to fine-tune some certain layers.

The CNN on the left in Fig. 6(a) fine-tunes the weights of whole network with target data. Wu et al. proposed a parallel multiscale filter bank CNN and fine-tuned it with 10, 20, 50 and 100 samples in target domain. The experimental results showed that the classification accuracy increased with the fine-tuning samples increased [96]. The similar applications of fine-tuning with different CNNs by other researchers are listed integrally in Table 2. Daoud and Bayoumi proposed a deep convolutional autoencoder architecture for epileptic seizure prediction and fine-tuned the model by pre-trained parameters [97]. Due to the time and data cost in pre-training a mature network, several proved successful network structures are used directly as pre-training models in fine-tuning, such as AlexNet, ResNet and VGG-16 [44,98–100].

The CNN on the right in Fig. 6(a) fine-tunes the weights of last three layers with other layers frozen. Yosinki et al. has proven that the performance of DNN could be improved by freezing some layers and fine-tuning certain layers via experiments [101]. Nejedly et al. constructed a CNN model for artifact identification and adjusted the parameters in the last fully connected layer with freezing other layers for transfer learning [95]. Raghu et al. also retrained the last layer of 10 pre-trained models to detect the variants of seizures and found that transfer learning based CNN outperformed conventional feature and clustering based methods [102].

3.3.2. Deep network adaptation

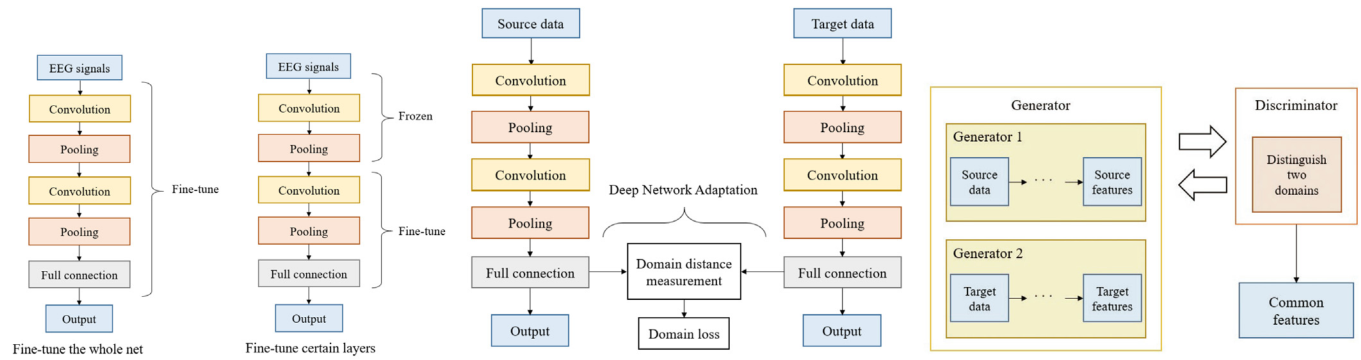
Although fine-tuning is easy to operate and understand, it is less effective when the distributions of source domain and target domain are different. In this case, researchers tried to consider distance measurement in transfer learning into the original networks, which is called deep neural network adaptation. This idea adjusts the cost function of the original network by adding a domain loss to measure the distribution of the source data and the target data. Fig. 6(b) displays a CNN example of deep network adaptation to adjust the distribution in full connection layer through domain distance measurement.

Jin et al. adopted domain adaptation network (DAN) with a domain classifier to maintain domain invariance [103]. Zhang et al. applied KL divergence into deep autoencoder to measure the difference between the source and target domains in feature extraction [104]. Tan et al. proposed a transfer learning model using AlexNet with an adversarial network to extract the general features and detect the difference and transferability of them [105]. Xie et al. applied a network structure called generalized hidden-mapping model (GHMM) into feature space to solve the problem of data scarcity, and tested GHMM in feedforward neural networks, fuzzy systems, and kernel regression models in epilepsy detection [106]. Svanera et al. established a multivariate link between the output and the fully connected layer of a CNN model through reduced rank regression and ridge regularization to transfer and decode the features to brain fMRI data [107]. Zhang et al. proposed a cross-subject transfer learning model by narrowing the difference of the feature distribution between the source and target domain via deep domain confusion [108].

3.3.3. Generative adversarial network transfer

Generative adversarial network (GAN) consists of generative network (generator) and discriminative network (discriminator), which was first proposed in 2014 [109]. The fake samples are generated by generators according to the given data and estimated by discriminators to distinguish the source of them. Researchers applied the principle of GAN to transfer learning in recent years: in transfer learning, the features of the source domain and the target domain learnt by generators are sent to discriminator, which verdicts the source of the features and feeds the result back to the generators until they cannot be distinguished. GAN obtains the common features of two domains in this process, as illustrated in Fig. 6(c).

Ma et al. proposed an adversarial domain generalization framework to reduce the influence of subject variability in BCI



(a) Structure of Fine-tuning (b) Structure of Deep Network Adaptation (c) Structure of GAN Transfer

Fig. 6. Structures of DNN Algorithms.

systems. Ma extracted the common features between the subjects and the features specific to subjects through adversarial training without using any target domain data [110]. Özdenizci et al. introduced an adversarial inference approach to decrease the variabilities cross subjects [111]. Özdenizci et al. performed an adversarial training to adapt the marginal distributions in the early layers of the DNN model and applied association reinforcement to adapt the conditional distributions in the last layers [112]. Ming et al. proposed a subject adaptation network inspired by GAN to align the distribution of different subjects [113]. Tang and Zhang presented a novel framework to add a conditional domain discriminator as an adversarial to learn the joint features in both domains for MI EEG signal decoding [114].

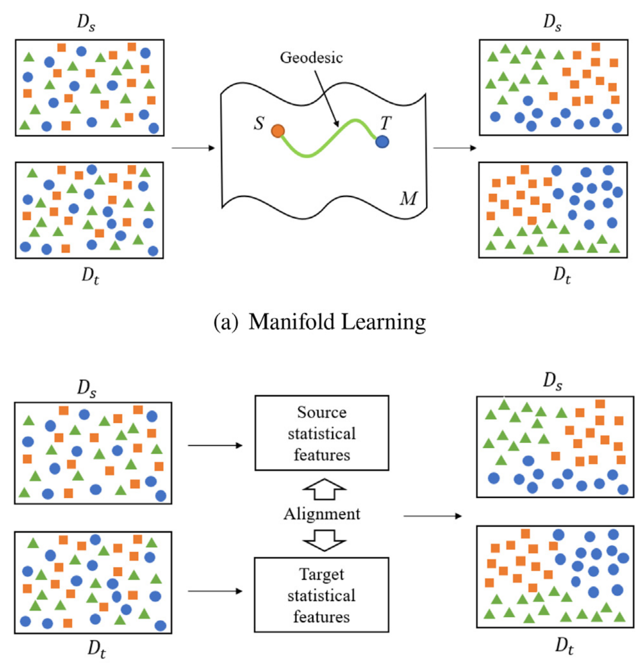
3.4. Subspace learning

Due to its effective optimization to complex data, subspace learning is gradually applied to EEG signal analysis as a dimensionality reduction technique. The basic assumption of subspace learning is to transform the source domain data and the target domain data into a subspace, in which their distributions are similar [85]. Traditional linear subspace learning algorithms have been successful in different data analysis, such as principal component analysis (PCA) and linear discriminant analysis (LDA) [138]. Although the relationship between the source and the target domain data is difficult to obtain, mapping them to another space which is compliant to certain structures with geometric or statistic operations can effectively solve this problem [139,140].

3.4.1. Manifold learning

Among subspace learning algorithms, manifold learning becomes popular in recent years. Manifold learning assumes that the original data are collected in a high-dimensional space, which contains low-dimensional manifold structure in this space [141]. A manifold is a space with a certain structure, which can reflect the distribution of high-dimensional data. The mapping in manifold is shown in Fig. 7(a): the data in the source domain and the target domain are mapped from the original space into the subspace manifold *M*, and the two domains are transformed into two points *S* and *T* on *M*. The distribution rules of three classes (represented by blue, orange and green) can be adjusted through the measurement of geodesic, representing by the green curve in the figure.

When symmetric positive definite (SPD) is chosen as the statistical feature to describe EEG signals, it is common to map the data into SPD manifold, which is a typical Riemannian manifold [119]. Rodrigues et al. proposed a framework based on SPD manifold to extract features through geometric operations and tested the algorithm into 8 public data sets, which obtained high classification



(b) Statistical Feature Alignment

Fig. 7. Structures of Subspace Learning Algorithms.

accuracy [121]. Zanini et al. applied SPD manifold and minimum distance to mean classifier for cross-subject motor imagery tasks [122]. Li et al. proposed a framework, which combines XDAWN spatial filter and SPD manifold to process cross-subject learning in P300 signals [118]. Nguyen et al. explored a method with Riemannian manifold based on covariance matrix on the manifold to measure the difference between source and target domains [120]. Bakhshali et al. evaluated the Riemannian distance of the correntropy spectral density (CSD) matrices for EEG signals from different channels for imagined speech recognition [117]. The research achievements of Riemannian manifold are listed integrally in Table 2.

Apart from the application of Riemannian manifold, there are also studies based on other manifolds. Li et al. proposed a feature extraction method based on locally linear embedding (LLE) to extracted the nonlinear features [124]. Gao et al. proposed a framework based on Isometric feature mapping (Isomap), a typical manifold learning algorithm to detect the connection of points [129]. Lin et al. explored a discriminative manifold learning based detection algorithm by using locality sensitive discriminant analysis

(LSDA), which mapped every class of features into a graph and measured the distance of the points in different graphs [131]. Zhang et al. successfully applied Laplacian eigenmap in extracting low-dimensional optimal EEG features [133]. Xia et al. mapped the data into a kernel space to calculate the predictive differences of the labeled data in the source domain by the pairwise constraint regularization term and explored the target domain distribution via the soft clustering regularization term using quadratic weights and Gini-Simpson diversity [132].

3.4.2. Statistical feature alignment

Statistical feature alignment aims to map the data into a subspace to align the statistical features, such as variance and median absolute deviation. Fig. 7(b) illustrates the process of statistical feature alignment: the statistical features of the three classes in two domains are adjusted to align through a transformation.

Chai et al. applied differential entropy feature to perform subspace alignment for non-stationary EEG signals into multi-subject unsupervised emotion recognition task [135]. Chai et al. also explored a new subspace alignment auto encoder, which combined subspace alignment and autoencoder into a framework through nonlinear transformation and a consistency constraint [134]. He and Wu proposed an unsupervised Euclidean space data alignment model to align the mean covariance matrix to Euclidean space, which was easy to implement rather than Riemannian subspace learning [137]. He and Wu also introduced a novel label alignment algorithm to decrease the conditional distribution difference when the label spaces of the source and target domains are different [136].

4. Challenges and opportunities

4.1. Challenges

In previous section, various applications of transfer learning in EEG signal analysis have been reviewed. Through the experiences of the above mentioned applications, the following is a brief description of the challenges that transfer learning methods face in the future study of EEG signal analysis, including the advantages and challenges of the four methods summarized in Table 3 and the problem of negative transfer.

4.1.1. Domain adaptation

Domain adaptation aims to allow the algorithm automatically adapting to the characteristics of data in different domains. Among domain adaptation algorithms, marginal distribution adaptation is applied effectively in multi-task learning. As a method receiving great attentions in the past few years, joint feature extraction should evaluate the effect of specific part of the domains on the outcome in future works, instead of analyzing the signals by joint features only.

Up to now, the focus of domain adaptation is the offline learning of marginal distribution adaptation. Conditional distribution adaptation algorithms discuss the problem of learning tasks in different target domains under similar distribution of source domains, which is more difficult than marginal distribution adaptation. In future study, it is worth noting that joint distribution adaptation can be explored when there are both marginal and conditional distribution difference in two domains, which should evaluate the effects of marginal and conditional distributions on the classification results from different points of view [142].

4.1.2. Improved CSP algorithms

CSP is a mature algorithm to build optimal spatial filter set for extracting frequency band features, which can be applied in many types of EEG signals. Therefore, researchers focused on improving CSP algorithms to make them adaptable to the non-stationarity and low signal-to-noise ratio of EEG signals. The improved CSP algorithms can increase the robustness of the model despite of the various interference factors in the EEG signal. Cheng et al. compared five types of improved CSP algorithms with traditional CSP, which indicated that in different experimental scenarios the improved CSP algorithms performed better than traditional CSP [86]. However, the room for CSP algorithms improvement is limited with the combination of frequency band features and other feature types, such as high order dynamics in EEG signal analysis [21,23,24]. For the new CSP algorithms that have emerged in recent years, such as CSP with kernel learning, it is still valuable to improve them by using transfer learning [143].

4.1.3. DNN methods

DNN methods use the original EEG data as input directly to extract and classify features by imitating the structure and connections of neurons. DNN methods are available to process all types of EEG signals resulting from its self-learning ability and have made achievements in offline and small-scale data set tasks so far. Jadhav et al. reported that fine-tuning is suitable when the distributions of the two domains are similar in small-scale data set tasks [144]. Deep network adaptation and GAN can reduce the large distribution difference of the two domains. CNN models are the most applied and explored, followed with autoencoder according to the literature.

Table 3
Advantages and Challenges of Transfer Learning Methods.

Methods	Advantages	Challenges
Domain adaptation	1. Effectively application of Marginal distribution adaptation in multi-task learning 2. Success in offline learning tasks	1. Domain privacy evaluation of Joint feature extraction 2. Joint distribution adaptation
Improved CSP Algorithms	1. The robustness of the model increase 2. Adaptability to the non-stationarity and low signal-to-noise ratio of EEG signals	1. Limited improvement room for CSP algorithms 2. Possibility of transfer learning with other CSP algorithms
DNN Methods	1. Ability to process all types of EEG signals 2. Effectively application of fine-tuning in similar distributions of two domains and small-scale data set tasks 3. Effectively application of GAN in large distribution difference of the two domains tasks	1. Deficiency of interpretability and universality 2. Deficiency of stability. 3. Similar development directions of current DNN methods
Subspace Learning	1. Effectively application on high-dimensional complex data 2. Suitable in semi-supervised learning tasks.	1. Complicated computation of manifold learning 2. Less effective as other algorithms on low-dimensional data 3. Easy to ignore the relationship of different sources in multi-task learning

However, DNN methods have a number of drawbacks. First, DNN methods lack interpretability and universality, and the network structure and parameters may greatly affect the learning ability of DNN models. Although Yosinki et al. has proven the transferability of DNN by experiments, it is still unfeasible to explain DNN theoretically at present [101]. Second, the results of DNN models lack stability, which can be deviated with different parameter values in the cost function. Third, the current DNN based transfer learning approaches for EEG signal analysis are mainly limited in fine-tuning the network because training a DNN requires a huge amount of data, which is difficult for EEG analysis tasks.

According to the literature, the current development of the transfer learning based DNN is slower than the other three methods because the improved transfer learning approaches based on CNN are similar although DNN models with different structures are proposed by researchers. On the basis of the existing achievements, the possible improvement can be focused on new DNN models, such as transfer dynamical autoencoder and Kohonen networks [145,146]. Future study should explore more possibilities in the cooperation of DNN and transfer learning, such as data enhancement and convergence rate improvement [147,148]. A DNN model without adjusting parameters should also be exploited to improve its robustness and reusability.

4.1.4. Subspace learning

Subspace learning maps the data in the source and target domain to another space to execute geometric or statistical operations. Hidden data structures and features that are not available in the original space, especially for high-dimensional complex data, can be discovered via subspace learning. In semi-supervised transfer learning, subspace learning method predicts labels of the target domain through mapping, which can be extended to many unsupervised applications [135].

The idea of subspace algorithms has already been applied in various scenarios as a dimensionality reduction approach. However, complicated computation is one concern in improving these algorithms, especially for manifold learning. As a nonlinear dimension reduction method, the calculation of geodesics in manifold learning is complex and unstable, which is more difficult than statistical feature alignment [137]. Apart from this, subspace learning is not as effective as other algorithms on low-dimensional data set because the features of these data can be extracted by low-rank matrix in the original feature space [104]. Subspace learning may ignore the relationship of different sources in multi-task learning, which also should be prudently improved in further study.

4.1.5. Negative transfer

Due to the characteristics and processing difficulties of EEG signals mentioned in Section 1, transfer learning is introduced to transfer the knowledge learnt in the source domain to the target domain based on the relationship between the data. However, few previous studies discussed the problem of *negative transfer* in EEG signal analysis. Negative transfer refers to the knowledge learnt in source domain supports a negative effect on the task in the target domain owing to data dissimilarity or unreasonable transfer method [149]. Negative transfer appears when the data in the source domain is dissimilar with the data in the target domain, or the transfer method fails to find the transferable components. In order to avoid negative transfer, the transferability between the source task and the target task and the similarity between domains or tasks must be properly analyzed before building effective models to guarantee the proper selection of the data sources and algorithms [43]. Negative transfer is also an underlying factor in adversely affecting the analysis result under the scenario which all the source data are utilized to predict the labels of the target domain, but actually only partial data in the source domain are related or similar to the data in the target domain. In this case, the transferability of the data must be considered before measuring the distance between them to avoid the adverse influence by irrelevant information.

4.2. Opportunities

So far, this paper has focused on the review of application and the challenges of transfer learning methods in EEG signal analysis. The following subsection discusses the opportunities and possible study directions of transfer learning in EEG signal analysis.

a) *Application forms*: Transfer learning can be used as an adjustment step after the whole model has been built through adjusting the model by mixing source domain data with all or partial amount of target domain data. This method is easy to operate and is applicable to situations where the difference between the source domain and the target domain is inconspicuous. The models which have been proven effective and successful can be applied to make direct adjustment to keep the interpretability of the whole algorithm. Transfer learning can also be used as a part of the whole proposed model in feature extraction and classification. In this case, the connectivity and distance of the source domain and the target domain are measured and minimized to decrease the difference between the two domains.

b) *Application Scenarios*: The application scenarios of transfer learning methods discussed in this review paper is summarized in Fig. 8. Domain adaptation algorithms are effective in multi-task and cross-subject learning problems. Improved CSP methods are suitable to reduce the adverse effects of individual factors and noise in non-stationary signal and to maintain the robustness under data scarcity condition. DNN is capable for all types of EEG signals and is helpful with fine-tuning in data scarcity tasks. Subspace learning methods are effective in cross-subject tasks and dimensionality reduction.

c) *Algorithms integration*: All algorithms do have own strengths and weaknesses for different problems under different situations. A reasonable combination of different algorithms may yield better results, such as joint feature selection with GAN and subspace learning with CSP [118,114,150]. Algorithms integration allows the model to overcome the weakness and limitation in their application scenarios, and is also possible to derive new algorithms, which is a potential research direction in this field.

d) *Algorithm choices*: Up to now, there are some transfer learning algorithms that have not been fully utilized in EEG signal analysis. For example, few transfer learning studies discuss the heterogeneous feature space tasks in EEG signal analysis. Existing machine learning methods which have been proved to be effective in EEG signal analysis, such as tensor classifier, are potential candidates to combine with transfer learning techniques [151]. Another example is joint distribution adaptation, which has been investigated in transfer learning, has not been discussed in EEG signal analysis [152]. Exploring the potential possibility of these algorithms is a direction worth pursuing.

e) *Performance evaluation*: It is difficult to compare transfer learning methods uniformly due to various criteria to measure the distance between the source domain and the target domain in different data sets, tasks and parameter settings, such as MMD, HSIC and OT. Cook proposed an interesting assumption in their review about transfer learning: to measure the difference between two domains using a universal distance independent of the domain in further studies [153]. A domain-independent distance helps researchers to intuitively select rea-

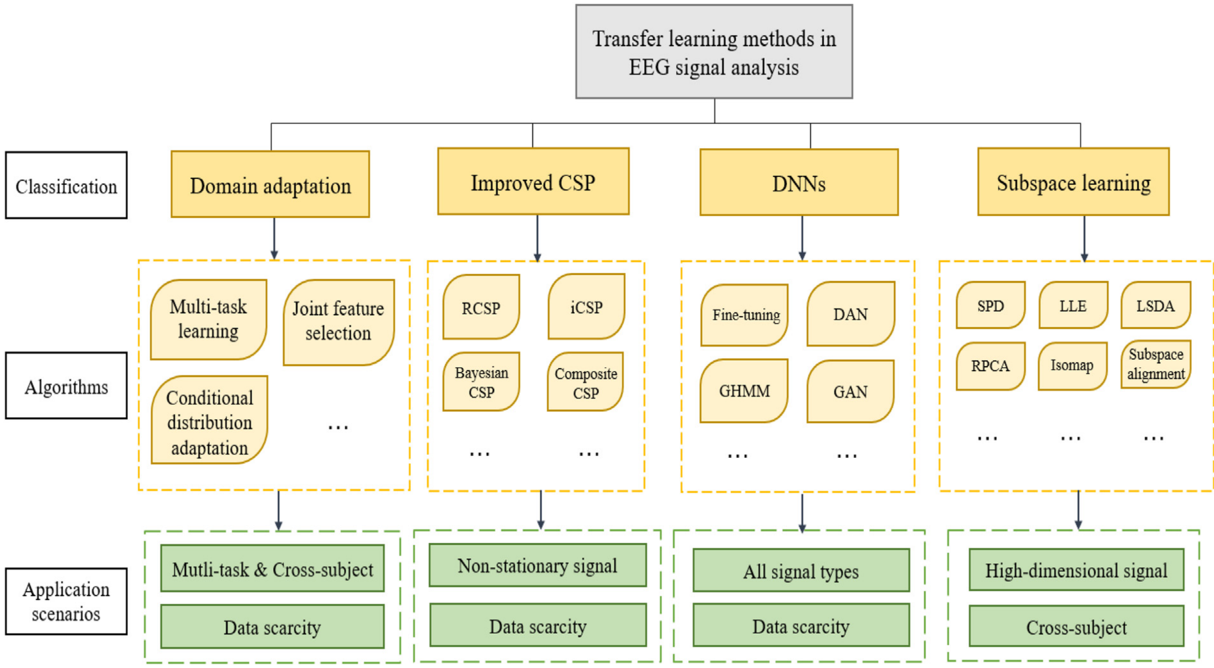


Fig. 8. Application Scenarios of Transfer Learning Methods in EEG Signal Analysis.

sonable algorithms and evaluate the performance of transfer learning algorithms.

f) *Cross-category learning*: The result of fusing different categories of EEG signal analysis may outperform single type of EEG signal. For example, Zou proposed a novel hybrid BCI paradigm based on both MI and P300 for EEG signal classification and the proposed hybrid paradigm yielded significantly better performance than the single-modality paradigm with less data size [154]. Therefore, cross-category learning is possible to be a direction of heterogeneous transfer learning in EEG signal analysis.

g) *Online transfer learning*: Most transfer learning studies focus on offline learning, which is not appropriate for many practical applications [14,36]. Online learning based on a small amount of real-time EEG data is inevitable, such as drowsy driving detection and wearable devices [155,156]. Online transfer learning methods should be the target in the long-term study of EEG signal analysis.

5. Conclusion

This paper reviews the applications of four frequently used transfer learning methods in EEG signal analysis in recent years: domain adaptation, improved CSP algorithms, DNN algorithms and subspace learning. The paper also discusses strengths and weaknesses of the above four methods as well as challenges transfer learning faces in the future development of EEG signal analysis. In the future, the studies can evaluate the transferability of data before building transfer learning models to avoid negative transfer. Furthermore, online learning needs to be developed to ensure the performance of transfer learning in practical applications.

CRediT authorship contribution statement

Zitong Wan: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Rui Yang:** Conceptualization, Formal analysis, Methodology, Project administration, Resources, Supervision, Writing - review & editing. **Mengjie Huang:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing - review & editing. **Nianyin Zeng:** Formal analysis, Writing - review & editing. **Xiaohui Liu:** Formal analysis, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. McClelland, Interpersonal processes and brain sciences – a new anthropology, Eur. Psychiatry 11 (1996) 224.
[2] D.P. Subha, P.K. Joseph, R.A. U, C.M. Lim, Eeg signal analysis: a survey, J. Med. Syst. 34 (2010) 195–212.
[3] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain-computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[4] G. Giacopelli, M. Migliore, D. Tegolo, Graph-theoretical derivation of brain structural connectivity, Appl. Math. Comput. 377 (2020), 125150.
[5] Y.-N. Chen, K.-M. Chang, J. Mostow, Towards using EEG to improve ASR accuracy, in: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012, pp. 382–385.
[6] P. Petrantonakis, L. Hadjileontiadis, Emotion recognition from EEG using higher order crossings, Int. Conf. IEEE Eng. Med. Biol. Soc. 14 (2010) 186–197.
[7] S.M. Slobounov, W. Ray, B. Johnson, E. Slobounov, K.M. Newell, Modulation of cortical activity in 2D versus 3D virtual reality environments: an EEG study, Int. J. Psychophysiol. 95 (2015) 254–260.
[8] K. Yoon, K. Kim, Multiple kernel learning based on three discriminant features for a P300 speller BCI, Neurocomputing 237 (2017) 133–144.
[9] L. Cai, X. Wei, J. Wang, H. Yu, B. Deng, R. Wang, Reconstruction of functional brain network in Alzheimer's disease via cross-frequency phase synchronization, Neurocomputing 314 (2018) 490–500.
[10] N. Kannathal, M.L. Choo, U.R. Acharya, P.K. Sadasivan, Entropies for detection of epilepsy in EEG, Comput. Methods Programs Biomed. 80 (2005) 187–194.
[11] S.L. Oh, J. Vicnesh, E.J. Ciaccio, R. Yuvaraj, U.R. Acharya, Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals, Appl. Sci. 9 (2019) 2870.

[12] B. Zineddin, Z. Wang, X. Liu, Cellular neural networks, the navier–stokes equation, and microarray image reconstruction, IEEE Trans. Image Process. 20 (2011) 3296–3301.

[13] C.A. Frantzidis, A.B. Vivas, A. Tsolaki, M.A. Klados, M. Tsolaki, P.D. Bamidis, Functional disorganization of small-world brain networks in mild Alzheimer's disease and amnestic mild cognitive impairment: an EEG study using relative wavelet entropy (RWE), Front. Aging Neurosci. 6 (2014), 224–224.

[14] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: a 10-year update, J. Neural Eng. 15 (2018) 31005.

[15] H. Kang, Y. Nam, S. Choi, Composite common spatial pattern for subject-to-subject transfer, IEEE Signal Process. Lett. 16 (2009) 683–686.

[16] L.R.R. Gianotti, J.S. Lobmaier, C. Calluso, F.M. Dahinden, D. Knoch, Theta resting EEG in TPJ/pSTS is associated with individual differences in the feeling of being looked at, Social Cogn. Affect. Neurosci. 13 (2018) 216–223.

[17] G. Matthews, L. Reinerman-Jones, J. Abich, A. Kustubayeva, Metrics for individual differences in EEG response to cognitive workload: optimizing performance prediction, Personality Individ. Differ. 118 (2017) 22–28.

[18] D. Hubner, T. Verhoeven, K.-R. Muller, P.-J. Kindermans, M. Tangermann, Unsupervised learning for brain-computer interfaces based on event-related potentials: review and online comparison [research frontier], IEEE Comput. Intell. Mag. 13 (2018) 66–77.

[19] S. Dutta, A. Nandy, Data augmentation for ambulatory EEG based cognitive state taxonomy system with RNN-LSTM, Int. Conf. Innov. Tech. Appl. Artif. Intell. (2019) 468–473.

[20] Y. Luo, EEG data augmentation for emotion recognition using a conditional wasserstein GAN, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2018, 2018, pp. 2535–2538.

[21] N. Brodu, F. Lotte, A. LéCuyer, Exploring two novel features for EEG-based brain-computer interfaces: Multifractal cumulants and predictive complexity, Neurocomputing 79 (2012) 87–94.

[22] H. Zhang, R. Chavarriaga, J. del R. Millán, Discriminant brain connectivity patterns of performance monitoring at average and single-trial levels, NeuroImage 120 (2015) 64–74.

[23] J. Frey, A. Appriou, F. Lotte, M. Hachet, Classifying EEG signals during stereoscopic visualization to estimate visual comfort, Comput. Intell. Neurosci. 2016 (2016), 2758103–2758103.

[24] R.N. Roy, S. Charbonnier, A. Campagne, S. Bonnet, Efficient mental workload estimation using task-independent EEG features, J. Neural Eng. 13 (2016) 26019.

[25] D. Delisle-Rodriguez, A.C. Villa-Parra, T. Bastos-Filho, A. López-Delis, A. Frizera-Neto, S. Krishnan, E. Rocon, Adaptive spatial filter based on similarity indices to preserve the neural information on EEG signals during on-line processing, Sensors 17 (2017) 2725.

[26] J. Hasan, J.-M. Kim, A hybrid feature pool-based emotional stress state detection algorithm using EEG signals, Brain Sci. 9 (2019) 376.

[27] A. Abu-Rmileh, E. Zakkay, L. Shmuelof, O. Shriki, Co-adaptive training improves efficacy of a multi-day EEG-based motor imagery BCI training, Front. Human Neurosci. 13 (2019) 362.

[28] T. Zhang, W. Chen, LMD based features for the automatic seizure detection of EEG signals using SVM, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (2017) 1100–1108.

[29] X. Navarro-Sune, A.L. Hudson, F.D.V. Fallani, J. Martinerie, A. Witon, P. Pouget, M. Raux, T. Similowski, M. Chavez, Riemannian geometry applied to detection of respiratory states from eeg signals: the basis for a brain–ventilator interface, IEEE Trans. Biomed. Eng. 64 (2017) 1138–1148.

[30] H. Dose, J.S. Møller, H.K. Iversen, S. Puthusserypady, An end-to-end deep learning approach to MI-EEG signal classification for BCIs, Expert Syst. Appl. 114 (2018) 532–542.

[31] M. Dai, D. Zheng, R. Na, S. Wang, S. Zhang, EEG classification of motor imagery using a novel deep learning framework, Sensors 19 (2019) 551.

[32] D. Du, Experimental study on neural feedback in embedded system teaching processing based on ERP signal analysis, Int. J. Emerg. Technol. Learn. 14 (2019) 109–120.

[33] K.R. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (2016) 9.

[34] J.C. Shaw, Correlation and coherence analysis of the EEG: a selective tutorial review, Int. J. Psychophysiol. 1 (1984) 255–266.

[35] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359.

[36] S. Thrun, L. Pratt, Learning to Learn: Introduction and Overview, Springer, 1998, pp. 3–17.

[37] Y. Chen, J. Wang, M. Huang, H. Yu, Cross-position activity recognition with stratified transfer learning, Pervasive Mobile Comput. 57 (2019) 1–13.

[38] M. Johnson, M. Schuster, Q.V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F.B. Viégas, M. Wattenberg, G. Corrado, M. Hughes, J. Dean, Google's multilingual neural machine translation system: enabling zero-shot translation, Trans. Assoc. Comput. Linguist. 5 (2017) 339–351.

[39] Y. Peng, B.-L. Lu, Discriminative manifold extreme learning machine and applications to image and EEG signal classification, Neurocomputing 174 (2016) 265–277.

[40] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Industr. Inf. 15 (2019) 2416–2425.

[41] R.A. Ramadan, A.V. Vasilakos, Brain computer interface: control signals review, Neurocomputing 223 (2017) 26–44.

[42] H. Dose, J.S. Møller, H.K. Iversen, S. Puthusserypady, An end-to-end deep learning approach to MI-EEG signal classification for BCIs, Expert Syst. Appl. 114 (2018) 532–542.

[43] Y.-P. Lin, T.-P. Jung, Improving EEG-based emotion classification using conditional transfer learning, Front. Human Neurosci. 11 (2017) 334.

[44] A.M. Azab, H. Ahmadi, L. Mihaylova, M. Arvaneh, Dynamic time warping-based transfer learning for improving common spatial patterns in brain-computer interface, J. Neural Eng. 17 (2020) 16061.

[45] H. Phan, O.Y. Chen, P. Koch, A. Mertins, M.D. Vos, Deep transfer learning for single-channel automatic sleep staging with channel mismatch, in: 2019 27th European Signal Processing Conference (EUSIPCO), 2019, pp. 1–5.

[46] Z. Deng, Y. Jiang, H. Ishibuchi, K.-S. Choi, S. Wang, Enhanced knowledge-leverage-based TSK fuzzy system modeling for inductive transfer learning, ACM Trans. Intell. Syst. Technol. 8 (2016) 11.

[47] I. Hossain, A. Khosravi, I. T. Hettiarachchi, S. Nahavandhi, Informative instance transfer learning with subject specific frequency responses for motor imagery brain computer interface, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 252–257.

[48] R. Raina, A. Battle, H. Lee, B. Packer, A.Y. Ng, Self-taught learning: transfer learning from unlabeled data, in, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 759–766.

[49] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F.-L. Chung, K.-S. Choi, S. Wang, Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (2017) 2270–2284.

[50] Q. Yao, H. Yang, A. Yu, J. Zhang, Transductive transfer learning-based spectrum optimization for resource reservation in seven-core elastic optical networks, J. Lightwave Technol. 37 (2019) 4164–4172.

[51] H. Chang, J. Han, C. Zhong, A.M. Snijders, J.-H. Mao, Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 1182–1194.

[52] W. Liu, R. Qin, A multikernel domain adaptation method for unsupervised transfer learning on cross-source and cross-region remote sensing data classification, IEEE Trans. Geosci. Remote Sens. 58 (2020) 4279–4289.

[53] S. Lu, C. Guan, H. Zhang, Unsupervised brain computer interface based on intersubject information and online adaptation, IEEE Trans. Neural Syst. Rehabil. Eng. 17 (2009) 135–145.

[54] Q. Wu, H. Wu, X. Zhou, M. Tan, Y. Xu, Y. Yan, T. Hao, Online transfer learning with multiple homogeneous or heterogeneous sources, IEEE Trans. Knowl. Data Eng. 29 (2017) 1494–1507.

[55] K. S. Barsim, W. Zheng, B. Yang, Ensemble learning to EEG-based brain computer interfaces with applications on P300-spellers, in: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, pp. 631–638.

[56] R. Ramele, A.J. Villar, J.M. Santos, EEG waveform analysis of P300 ERP with applications to brain computer interfaces, Brain Sci. 8 (2018) 199.

[57] S. Bhattacharyya, A. Konar, D.N. Tibarewala, M. Hayashibe, A generic transferable EEG decoder for online detection of error potential in target selection, Front. Neurosci. 11 (2017) 226.

[58] P. Zhao, S.C. Hoi, J. Wang, B. Li, Online transfer learning, Artif. Intell. 216 (2014) 76–102.

[59] D. Wu, V.J. Lawhern, S. Gordon, B.J. Lance, C.-T. Lin, Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR), IEEE Trans. Fuzzy Syst. 25 (2017) 1522–1535.

[60] R. Santana, L. Marti, M. Zhang, GP-based methods for domain adaptation: using brain decoding across subjects as a test-case, Genet. Programm. Evol. Mach. 20 (2019) 385–411.

[61] M. Dai, S. Wang, D. Zheng, R. Na, S. Zhang, Domain transfer multiple kernel boosting for classification of EEG motor imagery signals, IEEE Access 7 (2019) 49951–49960.

[62] J. Zhang, Y. Wang, S. Li, Cross-subject mental workload classification using kernel spectral regression and transfer learning techniques, Cogn. Technol. Work 19 (2017) 587–605.

[63] Z. Jiang, F.-L. Chung, S. Wang, Recognition of multiclass epileptic EEG signals based on knowledge and label space inductive transfer, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (2019) 630–642.

[64] Y. Liu, Z. Lan, J. Cui, O. Sourina, W. Muller-Wittig, EEG-based cross-subject mental fatigue recognition, in: 2019 International Conference on Cyberworlds (CW), 2019, pp. 247–252.

[65] X. Chai, Q. Wang, Y. Zhao, Y. Li, D. Liu, X. Liu, O. Bai, A fast, efficient domain adaptation technique for cross-domain electroencephalography(EEG)-based emotion recognition, Sensors 17 (2017) 1014.

[66] Z. Lan, O. Sourina, L. Wang, R. Scherer, G.R. Muller-Putz, Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets, IEEE Trans. Cogn. Develop. Syst. 11 (2019) 85–94.

[67] K. Yan, L. Kou, D. Zhang, Learning domain-invariant subspace using domain features and independence maximization, IEEE Trans. Syst. Man Cybern. 48 (2018) 288–299.

[68] J. He, G. Zhou, H. Wang, E. Sigalas, N. Thakor, A. Bezerianos, J. Li, Boosting transfer learning improves performance of driving drowsiness classification using EEG, in: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2018, pp. 1–4.

[69] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, H. He, Multisource transfer learning for cross-subject EEG emotion recognition, IEEE Trans. Syst. Man Cybern. 50 (2019) 3281–3293.
[70] V. Jayaram, M. Alamgir, Y. Altun, B. Schölkopf, M. Grosse-Wentrup, Transfer learning in brain-computer interfaces, arXiv preprint arXiv:1512.00296 (2015).
[71] W. Zheng, Z. Shi, B. Lv, Building cross-subject EEG-based effective models using heterogeneous transfer learning, Chin. J. Comput. 101 (2019) 177–189.
[72] G. Xiao, Y. Ma, C. Liu, D. Jiang, A machine emotion transfer model for intelligent human-machine interaction based on group division, Mech. Syst. Signal Process. 142 (2020), 106736.
[73] M.A.M. Joadder, J.J. Myszewski, M.H. Rahman, I. Wang, A performance based feature selection technique for subject independent MI based BCI, Health Inf. Sci. Syst. 7 (2019) 15.
[74] M. Nakanishi, Y.-T. Wang, C.-S. Wei, K.-J. Chiang, T.-P. Jung, Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses, IEEE Trans. Biomed. Eng. 67 (2020) 1105–1113.
[75] X. Xie, S. Sun, H. Chen, J. Qian, Domain adaptation with twin support vector machines, Neural Process. Lett. 48 (2018) 1213–1226.
[76] L. lan Chen, A. Zhang, X. guang Lou, Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning, Expert Syst. Appl. 137 (2019) 266–280.
[77] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, G. Prasad, Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface, Neurocomputing 343 (2019) 154–166.
[78] H. Fauzi, M.I. Shapiai, U. Khairuddin, Transfer learning of BCI using CUR algorithm, J. Signal Process. Syst. 92 (2020) 109–121.
[79] D. Banerjee, K. Islam, G. Mei, L. Xiao, G. Zhang, R. Xu, S. Ji, J. Li, A deep transfer learning approach for improved post-traumatic stress disorder diagnosis, in: 2017 IEEE International Conference on Data Mining (ICDM), 2017, pp. 11–20.
[80] A. Khalaf, M. Akcakaya, A probabilistic approach for calibration time reduction in hybrid EEG-fTCD brain-computer interfaces, Biomed. Eng. Online 19 (2020) 23.
[81] J.S. García-Salinas, L. Villaseñor-Pineda, C.A. Reyes-García, A.A. Torres-García, Transfer learning in imagined speech EEG-based BCIs, Biomed. Signal Process. Control 50 (2019) 151–157.
[82] E. Dagois, A. Khalaf, E. Sejdic, M. Akcakaya, Transfer learning for a multimodal hybrid EEG-fTCD brain–computer interface, IEEE Sens. Lett. 3 (2019) 1–4.
[83] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, G. Curio, The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects, NeuroImage 37 (2007) 539–550.
[84] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, Int. Conf. IEEE Eng. Med. Biol. Soc. 8 (2000) 441–446.
[85] K.K. Ang, Z.Y. Chin, H. Zhang, C. 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.
[86] M. Cheng, Z. Lu, H. Wang, Regularized common spatial patterns with subject-to-subject transfer of EEG signals, Cogn. Neurodyn. 11 (2017) 173–181.
[87] Y. Xu, Q. Wei, H. Zhang, R. Hu, J. Liu, J. Hua, F. Guo, Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI, J. Circuits Syst. Comput. 28 (2019) 1950123.
[88] H. Cho, M. Ahn, K. Kim, S.C. Jun, Increasing session-to-session transfer in a brain-computer interface with on-site background noise acquisition, J. Neural Eng. 12 (2015), 66009–66009.
[89] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, K.-R. Müller, V.V. Nikulin, Invariant common spatial patterns: alleviating nonstationarities in brain-computer interfacing, in: Advances in Neural Information Processing Systems 20, 2007, pp. 113–120.
[90] H. Kang, S. Choi, Bayesian common spatial patterns for multi-subject EEG classification, Neural Networks 57 (2014) 39–50.
[91] A.M. Azab, L. Mihaylova, K.K. Ang, M. Arvaneh, Weighted transfer learning for improving motor imagery-based brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (2019) 1352–1359.
[92] W. Samek, F.C. Meinecke, K.-R. Muller, Transferring subspaces between subjects in brain–computer interfacing, IEEE Trans. Biomed. Eng. 60 (2013) 2289–2298.
[93] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing 234 (2017) 11–26.
[94] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (2008) 41–56.
[95] P. Nejedly, J. Cimbalnik, P. Klimes, F. Plesinger, J. Halamek, V. Kremen, I. Viscor, B.H. Brinkmann, M. Pail, M. Brazdil, G.A. Worrell, P. Jurak, Intracerebral EEG artifact identification using convolutional neural networks, Neuroinformatics 17 (2018) 225–234.
[96] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, M. Dong, A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification, Front. Neurosci. 13 (2019) 1275.
[97] H. Daoud, M.A. Bayoumi, Efficient epileptic seizure prediction based on deep learning, IEEE Trans. Biomed. Circuits Syst. 13 (2019) 804–813.
[98] S. Siddharth, T.-P. Jung, T.J. Sejnowski, Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing, IEEE Transactions on Affective Computing (2019) 1, 1.
[99] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, W. Che, A deep transfer convolutional neural network framework for EEG signal classification, IEEE Access 7 (2019) 112767–112776.
[100] Özal Yildirim, P. Plawiak, R.S. Tan, U.R. Acharya, Arrhythmia detection using deep convolutional neural network with long duration ECG signals, Comput. Biol. Med. 102 (2018) 411–420.
[101] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks, in: Advances in Neural Information Processing Systems 27, 2014, pp. 3320–3328.
[102] S. Raghu, N. Sriraam, Y. Temel, S.V. Rao, P.L. Kubben, EEG based multi-class seizure type classification using convolutional neural network and transfer learning, Neural Networks 124 (2020) 202–212.
[103] Y.-M. Jin, Y.-D. Luo, W.-L. Zheng, B.-L. Lu, EEG-based emotion recognition using domain adaptation network, in: 2017 International Conference on Orange Technologies (ICOT), 2017.
[104] X.-Z. Zhang, W.-L. Zheng, B.-L. Lu, EEG-based sleep quality evaluation with deep transfer learning, in: International Conference on Neural Information Processing, 2017, pp. 543–552.
[105] C. Tan, F. Sun, W. Zhang, Deep transfer learning for EEG-based brain computer interface, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 916–920.
[106] L. Xie, Z. Deng, P. Xu, K.-S. Choi, S. Wang, Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals, IEEE Trans. Syst. Man Cybern. 49 (2019) 2200–2214.
[107] M. Svanera, M. Savardi, S. Benini, A. Signoroni, G. Raz, T. Hendler, L. Muckli, R. Goebel, G. Valente, Transfer learning of deep neural network representations for fMRI decoding, J. Neurosci. Methods 328 (2019), 108319.
[108] W. Zhang, F. Wang, Y. Jiang, Z. Xu, S. Wu, Y. Zhang, Cross-subject EEG-based emotion recognition with deep domain confusion, in: International Conference on Intelligent Robotics and Applications, 2019, pp. 558–570.
[109] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems 27, 2014, pp. 2672–2680.
[110] B.-Q. Ma, H. Li, W.-L. Zheng, B.-L. Lu, Reducing the subject variability of EEG signals with adversarial domain generalization, in: International Conference on Neural Information Processing, 2019, pp. 30–42.
[111] O. Ozdenizci, Y. Wang, T. Koike-Akino, D. Erdogmus, Learning invariant representations from EEG via adversarial inference, IEEE Access 8 (2020) 27074–27085.
[112] J. Li, S. Qiu, C. Du, Y. Wang, H. He, Domain adaptation for EEG emotion recognition based on latent representation similarity, IEEE Trans. Cogn. Develop. Syst. 12 (2020) 344–353.
[113] Y. Ming, W. Ding, D. Pelusi, D. Wu, Y.-K. Wang, M. Prasad, C.-T. Lin, Subject adaptation network for EEG data analysis, App. Soft Comput. 84 (2019), 105689.
[114] X. Tang, X. Zhang, Conditional adversarial domain adaptation neural network for motor imagery EEG decoding, Entropy 22 (2020) 96.
[115] F. Fahimi, Z. Zhang, W.B. Goh, T.-S. Lee, K.K. Ang, C. Guan, Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI, J. Neural Eng. 16 (2019) 26007.
[116] S. Kundu, S. Ari, MsCNN: a deep learning framework for P300-based brain–computer interface speller, IEEE Trans. Med. Robot. Bionics 2 (2020) 86–93.
[117] M.A. Bakhshali, M. Khademi, A. Ebrahimi-Moghadam, S. Moghimi, EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density, Biomed. Signal Process. Control 59 (2020), 101899.
[118] F. Li, Y. Xia, F. Wang, D. Zhang, X. Li, F. He, Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and Riemannian geometry classifier, Appl. Sci. 10 (2020) 1804.
[119] C.H. Nguyen, P.K. Artemiadis, EEG feature descriptors and discriminant analysis under riemannian manifold perspective, Neurocomputing 275 (2018) 1871–1883.
[120] C.H. Nguyen, G.K. Karavas, P. Artemiadis, Inferring imagined speech using EEG signals: a new approach using riemannian manifold features, J. Neural Eng. 15 (2018) 16002.
[121] P.L.C. Rodrigues, C. Jutten, M. Congedo, Riemannian procrustes analysis: Transfer learning for brain–computer interfaces, IEEE Trans. Biomed. Eng. 66 (2019) 2390–2401.
[122] P. Zanini, M. Congedo, C. Jutten, S. Said, Y. Berthoumieu, Transfer learning: a riemannian geometry framework with applications to brain–computer interfaces, IEEE Trans. Biomed. Eng. 65 (2018) 1107–1116.
[123] W. Zhang, D. Wu, Manifold embedded knowledge transfer for brain-computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 1117–1127.
[124] M. Li, X. Luo, J. Yang, Y. Sun, Applying a locally linear embedding algorithm for feature extraction and visualization of MI-EEG, J. Sens. 2016 (2016) 1–9.
[125] Y. Pan, S.S. Ge, A.A. Mamun, Weighted locally linear embedding for dimension reduction, Pattern Recogn. 42 (2009) 798–811.
[126] Y. Pan, S.S. Ge, A.A. Mamun, F.R. Tang, Detection of seizures in EEG signal using weighted locally linear embedding and SVM classifier, in: 2008 IEEE Conference on Cybernetics and Intelligent Systems, 2008, pp. 358–363.

[127] H. Rajaguru, S. K. Prabhakar, K. Saravanan, M. Kumar, Visualizing local linear embedding and fast ICA with linear neural networks for epilepsy classification, in: 2017 2nd International Conference on Communication and Electronics Systems (ICCES), 2017.
[128] Z. Yin, J. Zhang, Identification of temporal variations in mental workload using locally-linear-embedding-based EEG feature reduction and support-vector-machine-based clustering and classification techniques, Comput. Methods Programs Biomed. 115 (2014) 119–134.
[129] J. Gao, P. Lin, Y. Yang, P. Wang, C. Zheng, Real-time removal of ocular artifacts from EEG based on independent component analysis and manifold learning, Neural Comput. Appl. 19 (2010) 1217–1226.
[130] Y.-P. Lin, Constructing a personalized cross-day EEG-based emotion-classification model using transfer learning, IEEE J. Biomed. Health Inf. 24 (2020) 1255–1264.
[131] C. Lin, B.-H. Wang, N. Jiang, R. Xu, N. Mrachacz-Kersting, D. Farina, Discriminative manifold learning based detection of movement-related cortical potentials, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (2016) 921–927.
[132] K. Xia, T. Ni, H. Yin, B. Chen, Cross-domain classification model with knowledge utilization maximization for recognition of epileptic EEG signals, IEEE/ACM Trans. Comput. Biol. Bioinf. (2020), 1–1.
[133] J. Zhang, Z. Yin, R. Wang, Pattern classification of instantaneous cognitive task-load through GMM clustering, Laplacian eigenmap, and ensemble SVMs, IEEE/ACM Trans. Comput. Biol. Bioinf. 14 (2017) 947–965.
[134] X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, Y. Li, Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition, Comput. Biol. Med. 79 (2016) 205–214.
[135] X. Chai, Q. Wang, Y. Zhao, X. Liu, D. Liu, O. Bai, Multi-subject subspace alignment for non-stationary EEG-based emotion recognition, Technol. Health Care 26 (2018) 327–335.
[136] H. He, D. Wu, Different set domain adaptation for brain-computer interfaces: a label alignment approach, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 1091–1108.
[137] H. He, D. Wu, Transfer learning for brain–computer interfaces: a Euclidean space data alignment approach, IEEE Trans. Biomed. Eng. 67 (2020) 399–410.
[138] L. Qing, A survey: Subspace analysis for face recognition, Acta Automat. Sin. (2003).
[139] C. Anderson, J. Knight, T. O'Connor, M. Kirby, A. Sokolov, Geometric subspace methods and time-delay embedding for EEG artifact removal and classification, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2006) 142–146.
[140] M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review, Brain-Comput. Interfaces 4 (2017) 155–174.
[141] T. Lin, H. Zha, Riemannian manifold learning, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2008) 796–809.
[142] J. Wang, Y. Chen, S. Hao, W. Feng, Z. Shen, Balanced distribution adaptation for transfer learning, in: 2017 IEEE International Conference on Data Mining (ICDM), 2017, pp. 1129–1134.
[143] W. Samek, A. Binder, K.-R. Muller, Multiple kernel learning for brain-computer interfacing, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2013, 2013, pp. 7048–7051.
[144] P. Jadhav, G. Rajguru, D. Datta, S. Mukhopadhyay, Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network, Biocybern. Biomed. Eng. 40 (2020) 494–504.
[145] Z. Yin, M. Zhao, W. Zhang, Y. Wang, Y. Wang, J. Zhang, Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework, Neurocomputing 347 (2019) 212–229.
[146] Z. Wang, B. Zineddin, J. Liang, N. Zeng, Y. Li, M. Du, J. Cao, X. Liu, A novel neural network approach to cDNA microarray image segmentation, Comput. Methods Programs Biomed. 111 (2013) 189–198.
[147] Y. Jiao, Y. Deng, Y. Luo, B.-L. Lu, Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks, Neurocomputing 408 (2020) 100–111.
[148] W. Liu, Z. Wang, Y. Yuan, N. Zeng, K. Hone, X. Liu, A novel sigmoid-function-based adaptive weighted particle swarm optimizer, IEEE Trans. Syst. Man Cybern. (2019) 1–10.
[149] L.R. Novick, Analogical transfer, problem similarity, and expertise, J. Exp. Psychol.: Learn. Memory Cogn. 14 (1988) 510–520.
[150] W. Yue, Z. Wang, B. Tian, M. Pook, X. Liu, A hybrid model- and memory-based collaborative filtering algorithm for baseline data prediction of friedreich's ataxia patients, IEEE Trans. Industr. Inf. (2020), 1–1.
[151] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, T. Ristaniemi, Tensor decomposition of EEG signals: a brief review, J. Neurosci. Methods 248 (2015) 59–69.
[152] M. Long, J. Wang, G. Ding, J. Sun, P. S. Yu, Transfer feature learning with joint distribution adaptation, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 2200–2207.
[153] D. Cook, K.D. Feuz, N.C. Krishnan, Transfer learning for activity recognition: a survey, Knowl. Inf. Syst. 36 (2013) 537–556.
[154] C. Zuo, J. Jin, E. Yin, R. Saab, Y. Miao, X. Wang, D. Hu, A. Cichocki, Novel hybrid brain-computer interface system based on motor imagery and P300, Cogn. Neurodyn. 14 (2020) 253–265.
[155] A. Casson, D. Yates, S. Smith, J. Duncan, E. Rodriguez-Villegas, Wearable electroencephalography, IEEE Eng. Med. Biol. Mag. (2010).
[156] Y. Jiang, Y. Zhang, C. Lin, D. Wu, C.-T. Lin, EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system, IEEE Trans. Intell. Transp. Syst. (2020) 1–13.

Zitong Wan received the M.S. degree in Applied Informatics from University of Liverpool, Liverpool, United Kingdom, in 2019. She is currently working towards the Ph.D. degree in University of Liverpool, Liverpool, United Kingdom. Her research interests include transfer learning and biomedical signal analysis.

Rui Yang received the B.Eng. degree in Computer Engineering and the Ph.D. degree in Electrical and Computer Engineering from National University of Singapore in 2008 and 2013 respectively. He is currently an Assistant Professor in the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, and an Honorary Lecturer in the Department of Computer Science, University of Liverpool, Liverpool, United Kingdom. His research interests include machine learning based data analysis and applications.

Mengjie Huang received the Ph.D. degree from National University of Singapore in 2014, and the B.Eng degree from Sichuan University in 2009. She is now an Assistant Professor in the Design School, Xi'an Jiaotong-Liverpool University, Suzhou, China. Dr. Huang's current research interests include human–computer interaction and applications.

Nianyin Zeng was born in Fujian Province, China, in 1986. He received the B.Eng. degree in Electrical Engineering and Automation in 2008 and the Ph. D. degree in Electrical Engineering in 2013, both from Fuzhou University. From October 2012 to March 2013, he was a RA in the Department of Electrical and Electronic Engineering, the University of Hong Kong. From September 2017 to August 2018, he as an ISEF Fellow founded by the Korea Foundation for Advance Studies and also a Visiting Professor at the Korea Advanced Institute of Science and Technology. Currently, he is an Associate Professor with the Department of Instrumental & Electrical Engineering of Xiamen University. His current research interests include intelligent data analysis, computational intelligent, time-series modeling and applications. He is the author or co-author of several technical papers and also a very active reviewer for many international journals and conferences. Dr. Zeng is currently serving as an Associate Editor for Neurocomputing, and also Editorial Board members for Computers in Biology and Medicine and Biomedical Engineering Online. He also serves as a technical program committee member for ICBEB 2014, an Invited Session Chair of ICCSE 2017.

Xiaohui Liu received the B.Eng. degree in Computing from Hohai University, Nanjing, China, in 1982 and the Ph.D. degree in Computer Science from Heriot-Watt University, Edinburg, U.K., in 1988. He is a Professor of Computing with Brunel University London, Uxbridge, U. K., where he directs the Centre for Intelligent Data Analysis. He has over 100 journal publications in computational intelligence and data science. Prof. Liu was a recipient of the Highly Cited Researchers Award by Thomson Reuters.