Prediction Consistency and Confidence-based Proxy Domain Construction for Privacy-Preserving in Cross-subject EEG Classification

Yong Peng, Jiangchuan Liu, Honggang Liu, Natasha Padfield, Junhua Li, Wanzeng Kong, Bao-Liang Lu, *Fellow, IEEE* and Andrzej Cichocki, *Fellow, IEEE* .

***Abstract*—Domain adaptation has proven effective for suppressing the inter-subject variability problem in cross-subject EEG classification tasks in which labeled data is available for source subjects while only unlabeled data is provided for target subjects. Existing domain adaptation methods typically reduced the distribution discrepancy between source and target domains by directly utilizing source domain samples or features. To safeguard the privacy of source domain data, we propose to construct a Proxy Domain by simultaneously considering the prediction Consistency and Confidence (PDCC) of locally trained source models on target EEG samples, serving as the substitute to the source domain. The framework commences with the augmentation and alignment of the source domain data to enhance feature generalizability, after which source models are trained independently on each source subject's data in a decentralized manner. Knowledge transfer from source to target domains is achieved exclusively through accessing to the source domain model, enabling the PDCC-based proxy domain construction that encapsulates the source knowledge. Finally, domain adaptation is performed using the proxy domain and target domain. As a result, PDCC eliminates the need to access source domain data while effectively leveraging source knowledge. Experimental results on four benchmark EEG datasets demonstrate that PDCC consistently outperforms eleven existing methods, including several advanced transfer learning and source-free methods. Especially, the effectiveness of the proxy domain is extensively investigated. The source code for reproducing the experimental results is available from https://github.com/SunseaIU/PDCC.**

***Index Terms*—Brain computer interfaces, EEG, prediction consistency and confidence, privacy preserving, proxy domain.**

I. Introduction

BRAIN-computer interfaces (BCIs) give their users communication and control channels that do not depend on the brain's normal output channels of peripheral nerves and muscles [1]. Among the widely used neuroimaging techniques such as functional Magnetic Tomography, functional Near Infra-Red Spectroscopy and Magnetoencephalography, EEG is the most popular one due to its desirable properties such as objective, non-invasive, cost-efficient, and high temporal resolution. By detecting and recording electrical potential changes produced by neural activities inside the brain through electrodes placed on the scalp, EEG has occupied the largest proportion in representative BCI paradigms such as motor imagery, P300, steady-state visually evoked potentials and affective BCI [2]. Our main emphasis in this research is the EEG-based intention decoding of BCI users.

Decoding user intentions from collected EEG data using machine learning models constitutes a core component of many BCI systems. This process generally involves several key stages, including data preprocessing, feature extraction, model training, and prediction of mental intentions. Commonly extracted features include the temporal domain ones such as event related potential and some statistical metrics, frequency domain features such as power spectral density and differential entropy, and spatial domain features such as common spatial patterns and brain connectivity (functional) network. In contrast, deep learning models often circumvent the hand-crafted feature engineering stage, as they are capable of automatically learning feature representations from raw EEG data during the model training process though the interpretability of obtained results is limited [3].

Due to the non-stationary property and high inter-subject variability of EEG signals, transferring knowledge learned from source to target subjects is a critical challenge that needs to be addressed. Transfer learning, which leverages data or models from one or more source domains to enhance the learning in target domain, offers a promising approach to tackling these problems in cross-subject EEG decoding [4], [5]. Generally, two primary solutions were employed by transfer learning models for EEG classification; to be specific, one is learning domain-invariant EEG features shared by both source and target subjects from the feature perspective such as domain adaptation [6], and the other is first learning a pre-trained model on source data and then fine-tuning it by target data from the model perspective [7]. Then, the learned model can be effectively adapted to the target subject to shorten the

This work was supported by the National Key Research and Development Program of China under Grant 2023YFE0114900, the 'Pioneer' and 'Leading Goose' R&D Program of Zhejiang under Grant 2025C04001, Ministry for Science and Technology of PRC through the Sino-Malta Fund 2023 under Grant Sino-Malta-2023-18 (*Corresponding author: Wanzeng Kong*).

Y. Peng, H. Liu, and W. Kong are with School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China; J. Liu is with the HDU-ITMO Joint Institute, Hangzhou Dianzi University, Hangzhou 310018, China (email: kongwanzeng@hdu.edu.cn).

N. Padfield is with Centre for Biomedical Cybernetics, University of Malta, Misda 2080, Malta.

J. Li is with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, UK.

B.-L. Lu is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

A. Cichocki is with the Systems Research Institute of Polish Academy of Sciences, Warszawa, 01-447, Poland and the Warsaw University of Technology, Warsaw 00-661, Poland

model calibration time in BCI systems.

It is evident the majority of existing transfer learning models are instance-based and feature-based ones, meaning that we need to access the EEG data or features of the source subjects during the transfer process. However, EEG signals contain physiological information about the subjects such as the health status, sleep states, emotions, cognitive conditions, and even pathological information. Due to the laws of various countries and the high regard for privacy, EEG data from source subjects cannot be directly accessed or used publicly. To ensure data privacy during the transfer learning process, a two-stage paradigm is commonly adopted. In this framework, the model training phase—utilizing data from source subjects—is performed locally, and during the transfer learning stage, only the predictions generated by the source models on target subject samples are shared for knowledge transfer. This approach not only prevents privacy leakage of source data but also significantly reduces both time and transmission costs.

To enhance the generalization capability of transfer learning model in cross-subject EEG classification while ensuring privacy preserving, we propose to construct a Proxy Domain to serve as a substitute of the source domain, which is composed of the target samples with high Consistency and Confidence predicted by the locally learned source models (PDCC). The proposed approach begins with data augmentation and alignment in the source domain, followed by training source domain models locally. Knowledge transfer is facilitated through exclusive access to the source domain models rather than source domain EEG data, enabling the construction of a proxy domain that encapsulates the knowledge from the source domain. Finally, domain adaptation is conducted based on the proxy domain and target domain data. Compared to traditional EEG classification methods and deep learning approaches, PDCC effectively leverages source domain knowledge without requiring access to raw source domain data, ensuring both privacy preserving and better classification performance.

The contributions of this paper are summarized as follows.

- A three-stage source-free domain adaptation framework is proposed for cross-subject EEG classification by primarily constructing a proxy domain as a substitute to the source domain, based on which data privacy is preserved and knowledge transfer is achieved. Besides such core stage, source models are trained locally on augmented and preliminarily aligned EEG data in the first stage, and joint domain invariant feature learning and label estimation is completed in the last stage.
- A comprehensive score is built for selecting samples from target domain to form the proxy domain, which takes both the target domain sample prediction consistency and confidence obtained by the locally trained source models into consideration. Then, the selected samples are not only representative for source domain samples but also accurate enough for modeling the class prototypes.
- Comprehensive experiments are conducted on four publicly available EEG datasets, and the results depict the superior performance of the proposed PDCC model to several state-of-the-art non-source free and source-free models. Besides, some intermediate processes including the effectiveness of data augmentation and proxy domain are evaluated in detail.

The remainder of this paper is organized as follows. Section II introduces some background knowledge. Section III presents the PDCC model from four aspects, *i.e.*, data augmentation and alignment, local source model training, proxy domain construction and domain adaptation. Section IV conducts experiments to evaluate the effectiveness of PDCC in cross-subject EEG classification. Section V concludes the paper.

II. Background

A. Data Augmentation and Alignment

Data augmentation refers to the method of constructing iterative optimization or sampling algorithms by introducing unobserved data or latent variables [8]. This technology is particularly beneficial in machine learning and deep learning, when the given dataset is in moderate size. Its main goal is to increase the volume, quality and diversity of training data [9]. Then, the parameter space of learning models can be more sufficiently fitted; or equivalently, the learning models can more accurately and comprehensively capture the underlying data properties to enhance their generalization ability.

Data alignment refers to minimize the distribution divergence between different-sourced data, i.e., different sessions and (or) subjects. In [10], He *et al*. proposed to align EEG trials from different subjects in the Euclidean space to make them more similar, and hence improve the classification performance for a new subject. When considering more practical settings that source and target domains have different label spaces, a label alignment method was proposed to independently move the per-class covariance of each source subject, to re-center them at the corresponding class center of the target subject [11]. As a preprocessing step to the following domain adaptation process, Zhang *et al*. proposed to align the centroids of the covariance matrices of source and target domain data to reduce the marginal distribution shifts [12]. In [6], [13], domain invariant feature learning was tightly unified together with the estimation of target domain label indicator matrix for cross-subject EEG classification.

B. Privacy-preserving Learning

The extensive data collection required for machine learning and deep learning models raises significant privacy concerns, which have been gaining increasing attention within many communities. Many data collectors, such as those aiming to leverage the analysis of clinical records in medical institutions, face challenges in sharing and utilizing sensitive data due to privacy restrictions, often encountering risks of data leakage during the process. Therefore, a common dilemma is that abundant data is maintained locally instead of centralized sharing for model training. In BCI research, input data such as EEG contains rich privacy information and the developed machine learning model is usually proprietary. Therefore, data and model transmission among different parties may incur significant privacy threats [14], especially when performing

EEG decoding in cross-subject task by transferring auxiliary knowledge from source to target subjects. Therefore, it is of great necessity to develop advanced learning models with guaranteed privacy preserving ability.

C. Source-Free Domain Adaptation

Source-Free Domain Adaptation (SFDA) is an emerging technique that aims to address the limitations of conventional domain adaptation methods. Unlike traditional approaches, which often assume predefined label relationships between the source and target domains—such as closed-set or open-set conditions—and require simultaneous access to both source and target data during training, SFDA operates without direct access to the source domain data [15]. However, these assumptions are often impractical in real-world scenarios, particularly when protecting the privacy of source domain data is a critical concern. SFDA addresses these challenges by eliminating the need for direct access to source domain data, making it a more feasible and privacy-preserving alternative for domain adaptation tasks.

In recent years, researchers have developed various methods to tackle the challenges of data privacy preserving and domain adaptation in cross-domain tasks across diverse scenarios [16]. Among different SFDA models, SHOT (Source HypOthesis Transfer) is a fundamental framework that freezes the classifier module (hypothesis) of the source model and learns the target-specific feature extraction module [17]. In the fine-tuning stage of SHOT, both information maximization and self-supervised pseudo-labeling are exploited to implicitly align data representations from the target domain to the source hypothesis. Based on this framework, Zhang *et al.* proposed a multi-source decentralized transfer (MSDT) method by leveraging source model parameters in gray-box setting or prediction results in black-box setting to achieve effective knowledge transfer from multi-source models to the target domain [18]. To achieve more accurate label estimation of target samples, Gaussian mixture model was introduced to complete the self-supervised pseudo-labeling [19]. In [20], some labeled samples in target domain are available to adaptively refine the different contributions of source models. In the case that EEG samples in target domain arrive sequentially, Wu *et al.* proposed an online privacy-preserving transfer learning model based on the passive aggressive algorithm, which exhibited satisfactory performance in cross-subject motor imagery classification [7].

Different from the common paradigm of 'pre-training on source data locally and then fine-tuning on target data', Zhang *et al*. constructed a virtual intermediate domain to serve as the institute of source domain, which facilitates knowledge transfer and safeguards the privacy of source data [21]. This work inspires us a lot to construct a more representative and effective proxy domain to achieve privacy-preserving knowledge transfer for cross-subject EEG classification.

III. Methodology

A. Problem Definitions and Model Overview

Suppose we have K source subjects and the k-th subject contains $n_{s,k}$ labeled EEG trials (samples). The source data for each subject is represented as $D_s = \{\mathbb{X}_{s,i}, y_{s,i}\}_{i=1}^{n_{s,k}}$ where $\mathbb{X}_{s,i} \in \mathbb{R}^{ch\times l}$ denotes the EEG data matrix of the i-th trial, and $y_{s,i} \in \{1, \ldots, C\}$ represents the corresponding class label. Here, ch and l respectively denote the number of EEG channels and time-domain sampling points, and C is the number of classes. Similarly, the target domain contains n_t unlabeled trials (samples), represented as $D_t = \{\mathbb{X}_{t,i}\}_{i=1}^{n_t}$, where $\mathbb{X}_{t,i} \in \mathbb{R}^{ch\times l}$ is the EEG data matrix of the i-th trial in the target domain. The unlabeled target domain data is used for domain adaptation and to evaluate the model performance. The source data are only accessible during the stages of data alignment, data augmentation, and source model training. The goal is to aggregate data from all K source subjects to train M source models and obtain a robust source domain model. All the symbols used in this paper are defined in Table I.

TABLE I
NOTATIONS AND DESCRIPTIONS USED IN THIS PAPER.

Notation	Description	Notation	Description
D_s	source domain	$\mathbb{X}$	EEG trials
D_t	target domain	$\mathbf{X}$	feature matrix
D_v	proxy domain	$\mathbf{x}$	feature vector
$\mathbf{Y}$	label matrix	$\mathbf{y}$	label vector of $\mathbf{x}$
$\mathbf{F}_t$	prediction matrix	δ, β, γ	model parameters
M	# source models	$n_{s/v/t}$	# samples
K	# source subjects	$\mathbf{W}$	projection matrix
L_{noise}	noise injection ratio	$\mathbf{P}_{v/t}$	coupled projections
$\mathbf{Q}$	covariance matrix	$\mathbf{M}$	reference matrix
F_{shift}	frequency shift ratio	PM	confidence metric
L_{mult}	scaling parameter	SD	consistency metric
$rand$	random seed	CS	comprehensive score
$std()$	standard deviation	t	threshold

The proposed PDCC framework is shown in Figure 1, which includes the following four steps. To be specific, they are 1) *data alignment and augmentation*, which focuses on minimizing discrepancies in source data while enhancing the model's generalization capability through data alignment and augmentation techniques; 2) *local source model training*, which trains source models locally to reduce both computational time and data transmission costs; 3) *proxy domain construction*, which involves creating an intermediate domain to facilitate knowledge transfer from the source domain to the target domain while ensuring data privacy; and 4) *domain adaptation*, which aims to bridge the gap between the proxy domain and the target domain, thereby enhancing the model performance in cross-subject EEG classification. As shown in this figure, the first two steps are combined to form the first stage, making PDCC a three-stage framework.

B. Data Augmentation and Alignment

Data augmentation in each subject aims to improve both the size and diversity of EEG data, and then the source models can be sufficiently fitted to capture the underlying semantics. Following the study in [18], four data augmentation strategies including the noise injection, data flipping, data scaling, and frequency shift are used. Noise injection involves adding Gaussian noise to the data, which helps the model generalize better by making it less sensitive to small variations in the input data. Data flipping directly reverses the amplitude

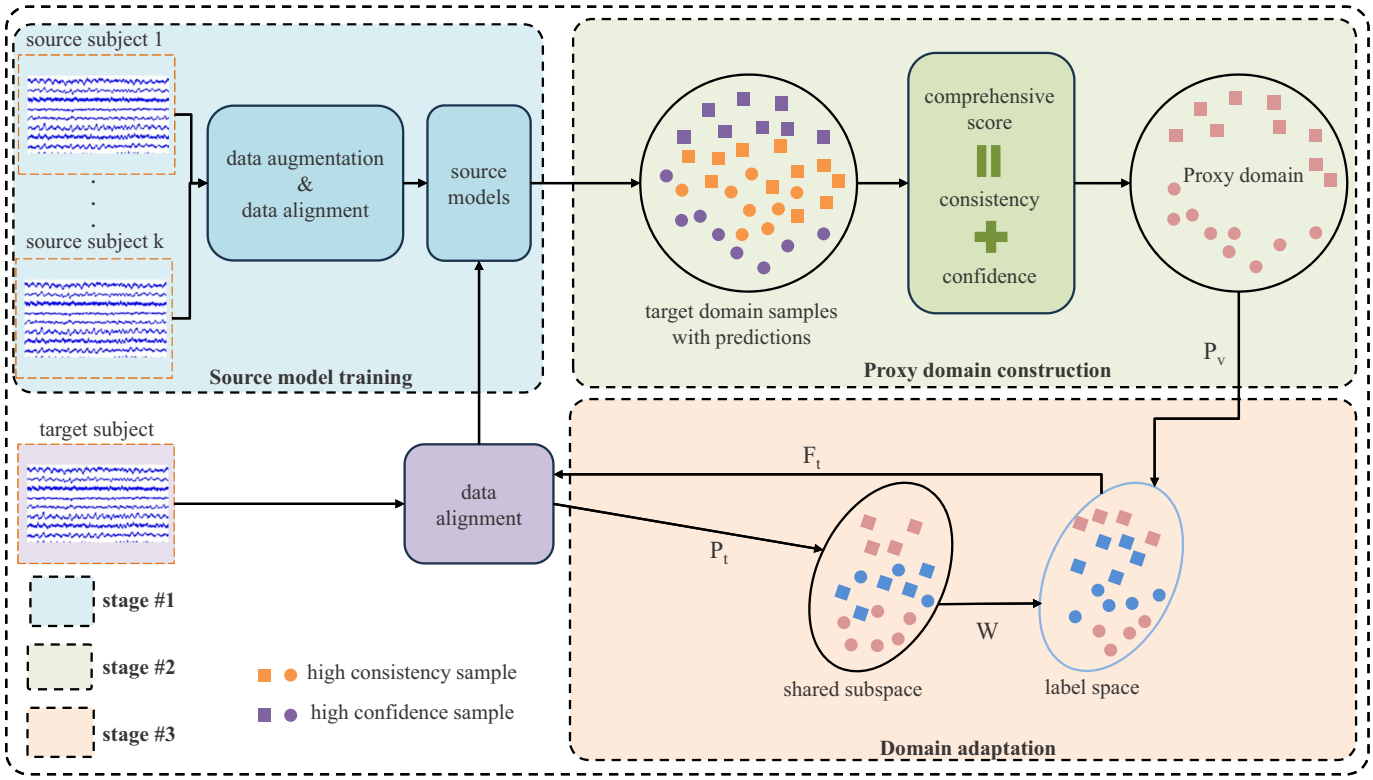

Fig. 1. The general framework of our proposed PDCC model for cross-subject EEG classification.

of the temporal represented EEG data, which is beneficial for learning the temporal patterns that are invariant to the direction of time. Data scaling is implemented by multiplying the data with a random scalar, in order to generalize the model to different amplitudes of EEG signals. Frequency shift involves shifting the frequency components of the EEG signals, which aims to improve the model robustness to slight variations in the frequency domain. By applying a combination of these four data augmentation strategies, we aim to create a more diverse and robust dataset, which in turn helps the model to generalize better and be more resilient to variations in the input data. The four strategies are summarized in Table II.

TABLE II
DATA AUGMENTATION STRATEGIES USED IN THIS WORK.

Augmentation Strategy	Mathematical formula
Noise Injection	$\mathbb{X}_{s,i} + rand * std(\mathbb{X}_{s,i})/L_{noise}$.
Data Flipping	$\max(\mathbb{X}_{s,i}) - \mathbb{X}_{s,i}$.
Data Scaling	$\mathbb{X}_{s,i} * (1 \pm L_{mult})$.
Frequency Shift	$F_{shift}(\mathbb{X}_{s,i} \pm L_{freq})$.

The raw EEG data consists of three dimensions, *i.e.*, trials, channels, and sampling points. To transform this data into a suitable format for decoding model training, feature extraction techniques including centroid alignment (CA) and tangent space mapping (TSM) are employed. CA aims to align trial data to a common reference point (*i.e.*, centroid) in order to achieve spatial consistency. This alignment reduces inter-trial variability and enhances the stability and accuracy of subsequent classification tasks. In CA, the covariance matrix is used as the feature representation of EEG data. By computing the covariance matrix for each trial and aligning these matrices to a centroid, we usually choose either the Euclidean centroid or the Riemannian centroid. The resultant aligned covariance matrices form a consistent feature space, facilitating more reliable analysis and classification. For each trial $\mathbb{X}_i \in \mathbb{R}^{m \times l}$, the key step in CA is to compute the mean covariance matrix for all trials using a chosen centering method. We can choose Euclidean mean, Riemannian mean, etc. The aligned covariance matrix $\mathbf{Q}_i'$ of the i-th trial is computed as

$$\mathbf{Q}_i' = \mathbf{M}_{ref}\mathbf{Q}_i\mathbf{M}_{ref}^T, \tag{1}$$

where $\mathbf{Q}_i'$ is the aligned covariance matrix, $\mathbf{M}$ is the mean covariance matrix, $\mathbf{M}_{ref} = \mathbf{M}^{-\frac{1}{2}}$ is the reference matrix. This process maps each trial's data to a common reference frame, thereby reducing inter-trial variability. After CA, we map each covariance matrix into tangent space by

$$\mathbf{x}_i = \text{upper}(\log(\mathbf{Q}_i')), \tag{2}$$

where $\mathbf{x}_i$ is the projection of the i-th aligned covariance matrix, upper$(\cdot)$ is an operator to extract the upper triangular part of the matrix. There are K source subjects in total. Therefore, we apply the transformation in (2) to all EEG samples from each source subject to obtain the corresponding feature vectors, which are then combined to form the corresponding source domain feature matrix. Finally, we merge all the source domain feature matrices into a single domain.

C. Local Source Model Training

In this step, we utilize ensemble learning methods to construct source models, which has proven an effective strategy

to enhance model performance and generalization capabilities. For simplicity, we selected three classic models, *i.e.*, support vector machine (SVM), logistic regression (LR), and linear discriminant analysis (LDA), and trained them on source domain data $\{\mathbf{X}_s, \mathbf{Y}_s\}$. This allows us not only to evaluate the performance of different algorithms but also to exploit their individual strengths, thereby achieving more stable and effective results in subsequent transfer learning tasks. Once each model is trained, it is saved locally, ensuring ease of retrieval and tuning in future steps, while also protecting the privacy of the source subjects.

To avoid the prediction bias of a single classification model in making decision, we adopted an ensemble learning strategy to combine the predictions of multiple source domain models using a soft voting mechanism. The underlying principle of this approach is that combining predictions from multiple models can effectively mitigate both bias and variance caused by individual models, thereby enhancing the robustness and accuracy of the overall prediction.

D. Proxy Domain Construction

The process of constructing a proxy domain to serve as the substitute of the source domain is a core step in ensuring privacy preserving and to facilitate the subsequent knowledge transfer. The underlying rationality to construct such proxy domain is that if certain target samples can be consistently and confidently predicted by the trained source models, they share similar properties with those in the source domain and should be selected to form the proxy domain. Accordingly, two complementary metrics, *i.e.*, prediction consistency and prediction confidence, are introduced as described below.

For the *prediction consistency*, we propose to use the standard deviation (SD) to quantitatively measure the divergence across all the learned source domain models on a certain target sample. Specifically, for $\mathbf{x}_{t,i}|_{i=1}^{n_t} \in D_t$, if its prediction probability belonging to the $c|_{c=1}^{C}$-th class by the $m|_{m=1}^{M}$-th source model is $y_{m,c}^{(i)}$, then SD_i is defined as

$$SD_i = \sqrt{\frac{1}{C}\frac{1}{M}\sum_{c=1}^{C}\sum_{m=1}^{M}\left(y_{m,c}^{(i)} - \frac{1}{M}\sum_{m'=1}^{M} y_{m',c}^{(i)}\right)^2}. \quad (3)$$

It is evident that even if all source models have fairly consistent predictions for a target domain EEG sample but these predictions are pretty fuzzy, then such sample is still not suitable to be selected in the proxy domain. Equivalently, such sample is not general enough to represent the properties of a certain class. Therefore, we need to additionally take the prediction confidence into consideration. To this end, we propose to use the power mean (PM) metric to measure the prediction confidence of source models made on a certain target sample $\mathbf{x}_{t,i}$; that is,

$$PM_i = \frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{M}\sum_{m=1}^{M}(y_{m,c}^{(i)})^p\right)^{\frac{1}{p}}, \quad (4)$$

where p is a free parameter and we set it to three in the following experiments.

As a result, we obtain the comprehensive score (CS) for each sample by weighting the standard deviation and power mean metrics. Assuming that α is a hyper-parameter to control the weights of both metrics, the formula for calculating the comprehensive score is

$$CS_i = \alpha \cdot PM_i + (1-\alpha)\cdot(1 - SD_i). \quad (5)$$

By adjusting their relative importance, the PDCC model can find a balance between prediction consistency and confidence. To determine the number of proxy domain samples, we use such combined metric by comparing each $CS_i|_{i=1}^{n_t}$ with a predefined threshold (*i.e.*, t). Target domain samples with comprehensive score values higher than the threshold are selected as the proxy domain samples. This approach ensures that the selected samples are both consistently and confidently predicted across different source models. Then, the constructed proxy domain (*i.e.*, $D_v \triangleq \{\mathbf{X}_v, \mathbf{Y}_v\}$) is high-fidelity to the source domain, which on one hand satisfies the privacy-preserving requirement and on the other hand facilitates the subsequent domain adaptation task. It is worth mentioning that the pseudo-label indicator matrix $\mathbf{Y}_v$ of proxy domain samples is formed by the soft voting among different source models.

E. Domain Adaptation

In this stage, we utilize the joint EEG feature transfer and semi-supervised cross-subject recognition (JTSR) model to achieve knowledge transfer due to its promising performance [13]. In JTSR, the domain-invariant feature representation subspaces and target labels are jointly optimized to achieve the optimum, which has exhibited excellent performance in enhancing the transferability of EEG features across subjects. JTSR assumes that there exists a latent common subspace shared by both proxy and target EEG data by two coupled projections $\mathbf{P}_v/\mathbf{P}_t \in \mathbb{R}^{d\times p}$, where p is the subspace dimensionality ($p \ll d$). Then, JTSR completes the domain invariant feature learning by considering both the marginal and conditional distributions of proxy and target domains; that is

$$\begin{aligned}\mathcal{D}(\mathbf{P}_v, \mathbf{P}_t, \mathbf{F}_t) = &\left\|\mathbf{P}_v^T\mathbf{X}_v\overline{\mathbf{Y}}_v\overline{\mathbf{N}}_v - \mathbf{P}_t^T\mathbf{X}_t\overline{\mathbf{F}}_t\overline{\mathbf{N}}_t\right\|_2^2 \\ &+ \delta\left\|\mathbf{P}_v - \mathbf{P}_t\right\|_2^2,\end{aligned} \quad (6)$$

where the second term ensures that there is no significant deviation between the two projection matrices. $\overline{\mathbf{Y}}_v = [\mathbf{1}_{n_v}, \mathbf{Y}_v]$, $\overline{\mathbf{F}}_t = [\mathbf{1}_{n_t}, \mathbf{F}_t]$, and $\overline{\mathbf{N}}_{v/t} = \mathrm{diag}(\frac{1}{n_{v/t}}, \mathbf{N}_{v/t})$. $\mathbf{F}_t \in \mathbb{R}^{n_t\times C}$ is an unknown label indicator matrix of target samples, in which each row is enforced to satisfy the probabilistic constraints. n_v (n_t) is the number of EEG samples in the proxy (target) domain. Due to the absence of labels on the target samples, the exact sample size per class cannot be precisely obtained. Instead, we estimate the the c-th class size by $n_t^c = \sum_{j=1}^{n_t} f_t^{(c,j)}$. $\mathbf{N}_v$ ($\mathbf{N}_t$) is a diagonal matrix whose c-th diagonal element is $\frac{1}{n_v^c}$ ($\frac{1}{n_t^c}$).

By connecting the data in subspace representation with label information by a projection matrix $\mathbf{W} \in \mathbb{R}^{p\times C}$, we have

$$\begin{aligned}\mathcal{L}(\mathbf{P}_{v/t}, \mathbf{W}, \mathbf{b}, \mathbf{F}_t) = &\beta\left\|\begin{bmatrix}\mathbf{X}_v^T\mathbf{P}_v \\ \mathbf{X}_t^T\mathbf{P}_t\end{bmatrix}\mathbf{W} + \mathbf{1}\mathbf{b}^T - \begin{bmatrix}\mathbf{Y}_v \\ \mathbf{F}_t\end{bmatrix}\right\|_2^2 \\ &+ \lambda\left\|\mathbf{W}\right\|_2^2, \text{s.t. } \mathbf{F}_t \geq 0, \mathbf{F}_t\mathbf{1} = \mathbf{1}.\end{aligned} \quad (7)$$

As a whole, the objective function of JTSR is formulated by combining (6) and (7) together as

$$\mathcal{O} = \mathcal{D}(\mathbf{P}_v, \mathbf{P}_t, \mathbf{F}_t) + \mathcal{L}(\mathbf{P}_v, \mathbf{P}_t, \mathbf{W}, \mathbf{F}_t). \tag{8}$$

By directly calculating the partial derivatives w.r.t. variables $\mathbf{W}$ and $\mathbf{b}$ and setting them to zero, the updating rules to them are respectively obtained as

$$\mathbf{b} = \frac{1}{n}\left(\mathbf{Y}^T\mathbf{1} - \mathbf{W}^T\mathbf{P}^T\mathbf{X}\mathbf{1}\right), \tag{9}$$

and

$$\mathbf{W} = \left(\mathbf{P}^T\mathbf{X}\mathbf{H}\mathbf{X}^T\mathbf{P} + \gamma\mathbf{I}\right)\mathbf{P}^T\mathbf{X}\mathbf{H}\mathbf{Y}, \tag{10}$$

where $\mathbf{H}$ is a data centering matrix and $\mathbf{P} = [\mathbf{P}_v; \mathbf{P}_t]$. By rewriting $\mathcal{D}(\mathbf{P}_v, \mathbf{P}_t, \mathbf{F}_t)$ as

$$\mathrm{Tr}\left(\mathbf{P}^T\mathbf{T}\mathbf{P}\right) + \alpha\mathrm{Tr}\left(\mathbf{P}^T\mathbf{G}\mathbf{P}\right), \tag{11}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{X}_v\bar{\mathbf{Y}}_v\bar{\mathbf{N}}_v\bar{\mathbf{N}}_v\bar{\mathbf{Y}}_v^T\mathbf{X}_v^T & -\mathbf{X}_v\bar{\mathbf{Y}}_v\bar{\mathbf{N}}_v\bar{\mathbf{N}}_t\bar{\mathbf{F}}_t^T\mathbf{X}_t^T \\ -\mathbf{X}_t\bar{\mathbf{F}}_t\bar{\mathbf{N}}_t\bar{\mathbf{N}}_v\bar{\mathbf{Y}}_v^T\mathbf{X}_v^T & \mathbf{X}_t\bar{\mathbf{F}}_t\bar{\mathbf{N}}_t\bar{\mathbf{N}}_t\bar{\mathbf{F}}_t^T\mathbf{X}_t^T \end{bmatrix},$$

and $\mathbf{G} = [\mathbf{I}_d, -\mathbf{I}_d; -\mathbf{I}_d, \mathbf{I}_d]$; therefore, with respect to $\mathbf{P}$, the objective function can be specified as

$$\min_{\mathbf{P}} \mathrm{Tr}\left(\mathbf{P}^T\left(\mathbf{T} + \alpha\mathbf{G}\right)\mathbf{P}\right) + \beta\|\mathbf{H}\mathbf{X}^T\mathbf{P}\mathbf{W} - \mathbf{H}\mathbf{Y}\|_2^2. \tag{12}$$

Taking the derivative of (12) w.r.t. $\mathbf{P}$ and setting it to zero, we have

$$(\mathbf{T} + \delta\mathbf{G})\mathbf{P} + \beta\mathbf{X}\mathbf{H}\mathbf{X}^T\mathbf{P}\mathbf{W}\mathbf{W}^T = \beta\mathbf{X}^T\mathbf{H}\mathbf{Y}\mathbf{W}^T, \tag{13}$$

which has the form $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{B} = \mathbf{C}$ and this is a standard Sylvester equation. Since both equality and inequality constraints are defined on $\mathbf{F}_t$, based on the Lagrangian multiplier method with KKT conditions, we get its updating rule as

$$\mathbf{F}_t = \frac{[\mathbf{Z}_l]^+ + [\mathbf{Z}_v]^- + \beta[\mathbf{Q}]^+ + \eta\mathbf{F}_{t1}\mathbf{1}\mathbf{1}^T}{[\mathbf{Z}_l]^- + [\mathbf{Z}_v]^+ + \beta[\mathbf{Q}]^- + \eta\mathbf{1}\mathbf{1}^T} \circ \mathbf{F}_t, \tag{14}$$

where

$$\begin{cases} \mathbf{Q} = \mathbf{H}_t\mathbf{F}_t - \mathbf{H}_t\mathbf{X}_t^T\mathbf{P}_t\mathbf{W}, \\ \mathbf{Z}_t = \mathbf{X}_t^T\mathbf{P}_t\left(\mathbf{P}_t^T\mathbf{X}_t\mathbf{F}_t\mathbf{N}_t\right)\mathbf{N}_t, \\ \mathbf{Z}_v = \mathbf{X}_t^T\mathbf{P}_t\left(\mathbf{P}_v^T\mathbf{X}_v\mathbf{Y}_v\mathbf{N}_v\right)\mathbf{N}_t. \end{cases} \tag{15}$$

The pseudo-code of our proposed PDCC model is shown in Algorithm 1.

Algorithm 1: Procedure of the proposed PDCC model.

Input: K labeled source subjects $\mathcal{S} = \{\mathcal{D}_{s,k}\}_{k=1}^K$; unlabeled target domain $\mathcal{D}_t$; balance parameter α; threshold t for proxy domain sample selection; domain adaptation model parameters δ, β, γ; number of iterations N; number of source models M

; **Output:** $\mathbf{F}_t$, estimated target domain labels.

`/* Data Augmentation */`

Using noise injection, data flipping, data scaling and frequency shift to augment source data.

`/* Data Alignment */`

Using (1) and (2) to align source data and unlabeled target data.

`/* Local Source Model Training */`

Combine all feature matrix $\mathbf{X}_s^k$ as well as label matrix $\mathbf{Y}_s^k$ from $\mathcal{S}$ to get one source domain $(\mathbf{X}_s, \mathbf{Y}_s)$, and generate M source models $\{\theta_m = f_m(\mathbf{X}_s, \mathbf{Y}_s)\}_{m=1}^M$.

`/* Proxy Domain Construction */`

Compute Consistency Score for each target sample by (5) and Select those target samples whose comprehensive scores exceed the threshold t to construct domain $\mathcal{D}_v = \{\mathbf{X}_v, \mathbf{Y}_v\}$ (The detailed processes are shown in Algorithm 2).

`/* Domain Adaptation */`

Initialize $\mathbf{F}_t = \frac{1}{C}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{n_t \times C}$.

while *not converged* **do**

- update $\mathbf{b}$ by rule (9);
- update $\mathbf{W}$ by solving (10);
- update $\mathbf{P}_{v/t}$ by rule (13);
- update $\mathbf{F}_t$ by rule (14);

Algorithm 2: Proxy domain construction

Input: M source models trained on source domain data, target domain $D_t = \{\mathbf{x}_{t,i}|_{i=1}^{n_t}\}$, number of classes C, power parameter $p = 3$, weight parameter α, selection threshold t;

Output: Proxy domain $D_v = \{\mathbf{X}_v, \mathbf{Y}_v\}$;

Initialize $D_v.\mathbf{X}_v = \emptyset$ and $D_v.\mathbf{Y}_v = \emptyset$;

for *each target sample* $\mathbf{x}_{t,i}$ *in* D_t **do**

- Calculate predictions based on the source models;
- Calculate prediction consistency SD_i by (3);
- Calculate prediction confidence PM_i by (4);
- Calculate comprehensive score CS_i by (5);
- **if** $CS_i > t$ **then**
  - Obtain the pseudo-label $\hat{\mathbf{y}}_{t,i}$ of $\mathbf{x}_{t,i}$ by soft voting among the M source models;
  - Add $(\mathbf{x}_{t,i}, \hat{\mathbf{y}}_{t,i})$ to proxy domain D_v;
- **else**
  - Skip $\mathbf{x}_{t,i}$;

return D_v;

IV. Experiments

A. Data Preparation

Four motor imagery EEG datasets were used in the following experiments as described in Table III. The first two datasets are from the BCI Competition IV [22]. The signals were sampled with 250 Hz and bandpass-filtered between 0.5 Hz and 100 Hz. BNCI2014002 was measured with a biosignal amplifier and active Ag/AgCl electrodes at a sampling rate of 512 Hz [23]. For BNCI2015001, the system sampling rate is also 512 Hz, with a bandpass filter between 0.5 and 100 Hz and a notch filter at 50 Hz [24]. The tasks in the BNCI2014001-2 dataset include the motor imagery of left hand, right hand while they are the left hand, right hand, feet and tongue in BNCI2014001-4; for the other two datasets, the imagined movements are right hand and feet. The basic experimental procedure is as follows. At the beginning of each trial, a fixed intersection point is displayed on the screen. After

two seconds, an arrow prompt (left, right, down, up) appears to indicate the motor imagery task that the subject needs to perform. The subject then continues the motor imagery task until the fixed intersection point disappears at the sixth second.

TABLE III
DESCRIPTIONS ON THE MAIN PROPERTIES OF USED DATASETS

Dataset	# subject	# channel	# trial	# class
BNCI2014001-2	9	22	144	2
BNCI2014001-4	9	22	288	4
BNCI2014002	14	15	120	2
BNCI2015001	12	13	400	2

In Table III, the first dataset can be downloaded from the BCI competition[1]. For the rest ones, we use the moabb library[2] to download and process the raw EEG data using the motor imagery paradigm, which helps extract relevant data from each subject's dataset.

B. Baseline Models and Experimental Settings

We conducted cross-subject motor imagery classification experiments by comparing the proposed PDCC model with some classic, deep learning, and source free domain adaptation models. Below are brief introductions to each of them.

- *Classic models* (Type-I). 1) CSP-LDA (Common Spatial Pattern-Linear Discriminant Analysis). CSP completes feature extraction by identifying spatial filters that best distinguish between different classes [25]. LDA is a supervised linear classifier to classify the extracted EEG features. 2) EA-CSP-LDA (Euclidean Alignment-CSP-LDA). This method aligns EEG data from different subjects in Euclidean space to reduce the inter-subject differences [10]. 3) CA-TSM-LDA (Centroid Alignment-Tangent Space Mapping-LDA). It combines centroid alignment [12] and tangent space mapping [26] to reduce the distribution differences of EEG data across subjects.
- *Deep learning models* (Type-II). 1) Deep Convolutional Network (DCN). It leverages convolutional neural networks to automatically extract spatio-temporal EEG features [27]. 2) Deep Adversarial Network (DAN). In DAN, a generator and a discriminator interact with each other and train each other through an adversarial process [28]. 3) Domain Adversarial Neural Network (DANN). Similar to DAN, DANN reduces the distribution differences between source and target domains by introducing domain discriminators [29].
- *Source-free knowledge transfer models* (Type-III). 1) EEG-DG. It is a multi-source domain generalization framework for motor imagery EEG classification that learns domain-invariant features with strong representation by optimizing both the marginal distribution and the conditional distribution to minimize the discrepancy across a variety of source domains, without accessing the target domain during training [30]. 2) FedBS. It is a privacy-preserving federated learning framework for EEG-based motor imagery classification that handles data heterogeneity across subjects through local batch-specific normalization and enhances the generalization ability via sharpness-aware optimization [31]. 3) SHOT. It is a source-free domain adaptation model that freezes the classifier module of a pretrained source model and adapts the target-specific feature extractor using information maximization and self-supervised pseudo-labeling to align target representations with the source hypothesis [17]. 4) Augmentation-based Source-Free Adaptation (ASFA). ASFA employs the data augmentation techniques in the source model training stage and considers both the uncertainty reduction for domain adaptation and consistency regularization for robustness in target model training [32]. 5) Lightweight Source-Free Transfer (LSFT). In this work, an intermediate virtual domain was constructed based on the prediction consistency of some trained source models on target domain samples to serve as the substitute of the source domain, achieving privacy preserving while performing knowledge transfer [21].

We employ the commonly used leave-one-subject-out cross-validation paradigm, where one subject is designated as the target while the remaining subjects within the same dataset serve as the source. This experiment focuses exclusively on unsupervised transfer learning, meaning that all EEG samples from the target subject are unlabeled. To safeguard the privacy of source subjects, data augmentation and alignment, source model training are conducted locally. Model performance is evaluated by using classification accuracy as the metric.

The DCN model comprises three main blocks, with each block consisting of a convolutional layer, a batch normalization layer, an activation layer, a pooling layer, and a dropout layer. DAN consists of a feature extractor and a domain discriminator, employing the MK-MMD (multi-kernel maximum mean discrepancy) loss to reduce the distribution discrepancy between the source and target domains. DANN integrates a feature alignment layer, a domain discriminator, and a gradient reversal layer to achieve domain adaptation. For these neural networks, the batch sizes are 32, 128, and 128 respectively. The learning rates for the source model training are set to 0.002, 0.01, and 0.01. The bottleneck layer dimensions are configured as 288, 50, and 50 respectively. EEG-DG employs a multi-branch convolutional architecture with domain-invariant feature learning and feature weighting for EEG domain generalization. This model consists of four parallel temporal convolution branches, each producing four feature maps, followed by a depthwise convolution block. A domain classifier and a feature weighting module enable domain-invariant representation learning. FedBS utilizes a compact convolutional architecture for EEG signal processing, featuring a temporal convolution block, a depthwise spatial convolution block, and a separable convolution block. This model uses a dropout rate of 0.25. This architecture progressively reduces feature dimensions through temporal filtering, spatial filtering across EEG channels, and separable convolutions, achieving efficient EEG feature extraction. For ASFA and LSFT, all the settings align with those described in their respective

[1]https://www.bbci.de/competition/iv/#dataset1
[2]https://github.com/NeuroTechX/moabb

papers. To be specific, in ASFA, the batch size is 128, and the bottleneck layer dimension is 50. The training epochs for the source and target domain models are 20 and 300, respectively, while the discriminator is trained for 100 epochs. The learning rate for all models is 0.01. Primary parameters include a weakening probability of p=0.1, weakening bound λ=0.5, Tsallis entropy parameter a=2, and trade-off parameter β=0.1. Only one auxiliary classifier is used throughout the training process. In LSFT, during the virtual intermediate domain construction and feature adaptation, a maximum inconsistency threshold 0.1 and μ=0.1 are applied in all experiments. Feature adaptation in LSFT uses a subspace dimensionality of p=20 and T=10 iterations. In PDCC, pseudo labels for target domain samples are generated through a soft voting mechanism among source model predictions. The parameter α is set to 0.7 to ensure the balance between prediction consistency and confidence. The thresholds (*i.e.*, CSs) for the four datasets were set to 0.79, 0.60, 0.68, and 0.65 respectively.

C. Cross-subject Classification Results

The cross-subject EEG classification accuracies for the four datasets are presented in Tables IV-VII, which include the classification accuracy for each target subject and also the average accuracy and standard deviation across all the target subjects. The best performance in each task is marked in bold and the second-best is underlined. On average, our proposed PDCC model achieved the best average performance on all the datasets.

The experimental results show the proposed PDCC model outperforms all classic methods. Compared to deep learning-based baseline models, PDCC achieves much higher accuracy than DCN, DAN, and DANN. While these models benefit from feature extraction and domain adaptation capabilities, they are less effective in unsupervised domain adaptation settings. PDCC demonstrates significant improvements over the compared source-free domain adaptation methods. The six source-free knowledge transfer models achieve privacy-preserving from four different perspectives. EEG-DG is a domain generalization model, while FedBS works under the federated learning framework. ASFA is an improved model from SHOT; therefore, both employ the pre-training and fine-tuning paradigm. PDCC and LSFT achieve source data privacy preserving by constructing a proxy domain to serve as the substitute of the source domain. Notably, our proposed method achieves consistent improvements across most of the subjects; for example, the challenging cases such as the second and the fifth subjects in BNCI2014001-2, where the accuracies are significantly higher. Overall, PDCC achieves the highest overall average accuracy on all the four datasets. Moreover, we find that PDCC usually has a smaller standard deviation, indicating the stability of the model's performance.

To rigorously assess the significance of performance improvements, especially for comparisons with marginal gains, we conducted pair-wised students t-test between PDCC and each of the other compared methods, by pooling the results across all the four datasets together (aggregated 44 participants). The statistical test results in Table VIII confirm that PDCC's improvements over all the baseline models are statistically significant ($p < 0.05$ for every comparison). These results validate the effectiveness of our proposed PDCC model, equivalently, the effectiveness of its building blocks, i.e., the proxy domain construction, the domain adaptation and the data augmentation methods.

D. Impact of Data Augmentation

To intuitively depict the effectiveness of data augmentation, Figure 2 takes the BNCI2014001-2 dataset as an example and shows the two-dimensional visualization of the EEG samples before and after data augmentation by t-SNE [33]. The augmented dataset strengthens the model's robustness by incorporating greater diversity into existing EEG samples. This augmentation allows the model to better deal with various disturbances when encountering unknown data patterns, thereby mitigating the risk of overfitting and improving its generalization capacity. As the diversity of the dataset increases, the model is able to more effectively capture the distinct features of different samples during training, leading to an improvement in prediction accuracy.

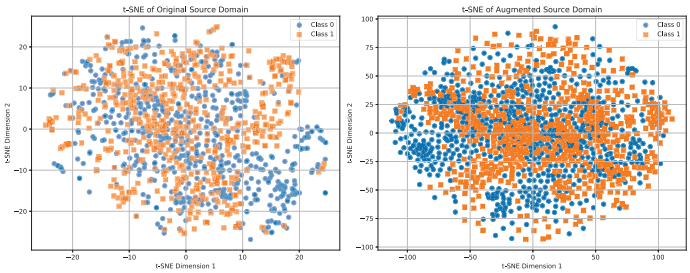

Fig. 2. Visualization of EEG samples before and after augmentation in BNCI2014001-2.

Generally, different data augmentation methods may have different impacts on the EEG classification performance. By taking the subject 4 from the BNCI2014001-2 dataset as an example, the experimental results in Figure 3 demonstrate that PDCC with data augmentation always obtained better performance than that without data augmentation. It is worth noting that the strategy of combining multiple data augmentation methods achieved the best results, corresponding to the highest average classification accuracy. This indicates that the combination enhancement method further improves the model generalization ability by capturing multiple feature variations. Similar results were also found on some other subjects.

E. Impact of the Proxy Domain Construction

According to the comprehensive score defined in (5) for the proxy domain construction, the selected samples from target domain are enforced to be similar to source domain samples to guarantee high prediction consistency as well as close to class prototypes to ensure high prediction confidence, which are also expected to be representative enough to capture the primary characteristics of target domain samples. Only in this way, the subsequent domain adaptation stage will accordingly achieve desirable performance. As shown in Figure 4, it corresponds to the two-dimensional visualization of EEG samples from

TABLE IV
CROSS-SUBJECT EEG CLASSIFICATION ACCURACIES (%) ON THE BNCI2014001-2 DATASET.

	Algorithm	ID_subject									Avg.±std
		1	2	3	4	5	6	7	8	9	
Type-I	CSP-LDA	66.67	54.86	90.97	65.28	44.44	54.17	61.81	89.58	75.00	66.98±15.81
	EA-CSP	86.11	53.47	91.67	70.14	55.56	67.36	63.89	90.97	77.92	73.01±14.47
	CA-TSM	72.22	51.39	77.78	**75.69**	56.94	67.36	56.94	89.58	73.61	69.06±12.15
Type-II	DCN	76.39	56.94	87.50	71.53	56.94	68.75	54.86	91.67	71.53	70.68±13.18
	DAN	65.28	55.56	86.81	65.28	52.08	65.97	52.78	78.47	70.14	65.82±11.64
	DANN	77.78	48.61	84.03	74.31	56.94	66.67	63.19	87.50	71.53	70.06±12.60
Type-III	EEG-DG	75.02	51.58	88.28	72.12	61.34	67.50	64.78	89.26	73.61	71.50±12.12
	FedBS	82.51	59.32	92.36	71.53	65.28	71.53	65.28	84.03	77.78	74.40±10.62
	SHOT	79.17	52.78	**96.53**	70.14	52.08	70.14	64.58	92.36	72.22	72.22±15.39
	ASFA	76.22	56.25	74.31	69.21	**70.33**	69.40	**69.10**	90.51	74.13	72.16±08.97
	LSFT	**87.50**	49.31	94.44	72.92	55.56	68.75	65.28	**96.53**	**86.11**	75.16±16.94
	PDCC	84.03	**65.28**	89.58	74.31	69.28	**75.01**	68.75	86.81	79.17	**76.91**±08.55

TABLE V
CROSS-SUBJECT EEG CLASSIFICATION ACCURACIES (%) ON THE BNCI2014001-4 DATASET.

	Algorithm	ID_subject									Avg.±std
		1	2	3	4	5	6	7	8	9	
Type-I	CSP-LDA	54.17	24.31	56.94	36.11	28.12	24.65	34.38	69.79	50.35	42.09±16.25
	EA-CSP	68.40	24.65	73.61	45.14	33.33	39.24	60.42	72.57	55.56	52.55±17.84
	CA-TSM	70.14	32.29	77.78	45.49	36.11	38.19	48.96	70.14	63.54	53.63±17.01
Type-II	DCN	65.90	32.64	74.34	41.94	25.00	30.73	40.90	67.64	59.13	48.69±18.27
	DAN	57.64	32.29	62.50	40.28	35.07	41.32	39.93	60.42	51.74	46.80±11.41
	DANN	62.50	29.17	68.75	44.10	31.25	39.93	50.00	65.97	59.03	50.08±14.86
Type-III	EEG-DG	67.22	24.13	70.97	45.00	35.19	36.74	58.12	61.77	62.12	51.25±16.44
	FedBS	71.32	27.16	78.25	45.78	38.91	47.58	49.76	75.12	64.05	55.33±17.66
	SHOT	71.53	27.08	81.25	46.18	33.68	35.76	64.58	77.78	60.42	55.36±20.26
	ASFA	72.99	28.19	**86.25**	42.95	36.01	41.60	**69.51**	78.78	59.90	57.35±20.78
	LSFT	**77.08**	26.74	85.76	45.49	36.46	38.54	63.89	80.21	**67.71**	57.99±21.60
	PDCC	60.71	**50.00**	81.25	**46.53**	**47.92**	**52.08**	45.49	**83.33**	60.00	**58.59**±14.49

TABLE VI
CROSS-SUBJECT EEG CLASSIFICATION ACCURACIES (%) ON THE BNCI2014002 DATASET.

	Alg.	ID_subject														Avg.±std
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Type-I	CSP-LDA	55.63	60.00	91.88	81.25	56.88	57.50	83.12	56.88	91.25	65.00	50.63	57.50	50.00	48.13	64.69±15.38
	EA-CSP	56.25	61.88	93.13	82.50	57.50	66.25	69.38	61.88	90.00	61.88	50.00	56.25	50.00	49.38	64.73±14.34
	CA-TSM	50.00	60.00	95.63	50.00	53.75	58.75	71.25	56.25	80.00	64.38	58.75	58.75	50.63	50.00	61.30±13.12
Type-II	DCN	53.13	78.75	85.00	50.00	60.63	50.00	50.00	54.38	50.00	63.13	50.00	62.50	50.00	50.00	57.68±11.42
	DAN	60.63	81.25	88.75	73.75	71.25	74.38	79.38	66.88	77.50	59.38	63.75	65.00	56.25	47.50	68.98±11.14
	DANN	61.88	81.88	86.88	78.13	66.25	65.00	78.13	60.63	80.63	58.13	66.25	67.50	61.25	48.13	68.62±10.93
Type-III	EEG-DG	60.00	81.42	89.85	79.41	70.68	68.39	80.83	65.62	85.31	61.56	56.44	69.38	57.38	49.63	69.71±12.14
	FedBS	66.25	80.00	88.00	78.13	74.90	72.10	81.30	72.20	93.00	61.25	**74.38**	75.62	59.50	61.88	74.18±09.82
	SHOT	67.50	83.75	99.38	**85.00**	65.00	61.25	86.88	50.63	93.75	60.00	51.88	61.88	50.63	50.00	69.11±17.25
	ASFA	65.20	67.80	70.90	57.90	51.80	49.20	80.70	**99.00**	69.90	**74.90**	57.20	65.00	**84.90**	**93.10**	70.54±14.83
	LSFT	60.62	83.75	**100.0**	83.13	72.50	71.88	**91.25**	65.00	**94.37**	66.25	62.50	80.00	54.37	52.50	74.15±14.96
	PDCC	**70.00**	**84.38**	97.50	80.63	**78.75**	**75.62**	80.00	74.38	93.13	66.25	71.88	**81.25**	65.00	66.87	**77.55**±09.69

the subject 1 in the BNCI2014001-2 dataset, from which we observe that the samples in the constructed proxy domain are more dispersed compared to the overall data distribution, corresponding to higher representativeness.

The proxy domain serves as the substitute to the source domain knowledge. Therefore, domain adaptation is performed between proxy domain and target domain. In Figure 5, we show the EEG samples before and after domain adaptation, from which it is observed that both the marginal and conditional distributions of EEG samples in proxy and target domains are well aligned, indicating the effectiveness of the JTSR method in domain invariant feature learning.

To explicitly show the superiority of the comprehensive score to the prediction consistency in proxy domain con-

TABLE VII
CROSS-SUBJECT EEG CLASSIFICATION ACCURACIES (%) ON THE BNCI2015001 DATASET.

	Algorithm	ID_subject												Avg.±std
		1	2	3	4	5	6	7	8	9	10	11	12	
Type-I	CSP-LDA	55.63	60.00	91.88	81.25	56.88	57.50	**83.12**	56.88	**91.25**	65.00	50.63	57.50	67.29±15.09
	EA-CSP	56.25	61.88	93.13	82.50	57.50	66.25	69.38	61.88	90.00	61.88	50.00	56.25	67.24±13.97
	CA-TSM	50.00	60.00	**95.63**	50.00	53.75	58.75	71.25	56.25	80.00	64.38	58.75	58.75	63.13±13.33
Type-II	DCN	97.50	95.00	61.50	56.25	60.25	62.25	60.75	50.33	51.33	55.33	49.83	51.75	62.67±16.32
	DAN	86.75	92.25	81.75	80.25	72.50	63.00	63.50	59.17	66.67	63.00	49.67	69.63	70.68±12.44
	DANN	84.25	88.00	80.50	77.75	77.75	59.50	67.75	59.83	71.33	58.33	48.67	54.00	68.97±12.84
Type-III	EEG-DG	87.50	90.50	85.00	88.25	85.50	65.25	64.25	70.75	58.50	54.13	52.50	54.75	71.41±15.02
	FedBS	94.10	94.60	88.00	82.50	85.80	66.67	69.50	64.50	67.00	64.25	60.60	57.90	74.62±13.41
	SHOT	98.25	95.75	91.75	**90.00**	78.00	50.00	63.50	50.00	64.83	51.13	50.17	50.50	69.49±19.99
	ASFA	98.50	**96.00**	90.75	77.75	50.25	61.50	49.16	**71.67**	51.13	50.17	49.00	**98.75**	70.39±21.09
	LSFT	**98.75**	**96.00**	91.00	84.75	**87.25**	55.75	55.25	52.75	63.50	55.50	54.75	55.50	70.90±18.74
	PDCC	95.25	88.75	87.25	85.75	81.75	**69.75**	78.00	64.75	68.75	**67.00**	**69.50**	61.00	**76.46**±11.10

TABLE VIII
STATISTICAL TESTS BETWEEN PDCC AND BASELINE MODELS

t-test	*t*-statistic	*p*-value	Significance level
PDCC vs CSP-LDA	6.7616	$\ll 0.001$	***
PDCC vs EA-CSP	4.9595	$\ll 0.001$	***
PDCC vs CA-TSM	6.3351	$\ll 0.001$	***
PDCC vs DCN	8.4064	$\ll 0.001$	***
PDCC vs DAN	9.2096	$\ll 0.001$	***
PDCC vs DANN	8.5142	$\ll 0.001$	***
PDCC vs EEG-DG	5.5718	$\ll 0.001$	***
PDCC vs FedBS	3.5446	0.001	***
PDCC vs SHOT	4.0248	0.0002	***
PDCC vs ASFA	2.1115	0.0406	*
PDCC vs LSFT	2.1277	0.0391	*

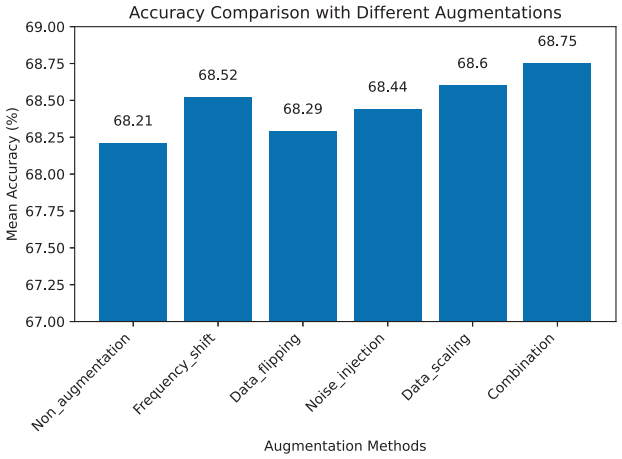

Fig. 3. EEG decoding accuracies corresponding to different data augmentation methods on subject 4 from the BNCI2014001-2 dataset.

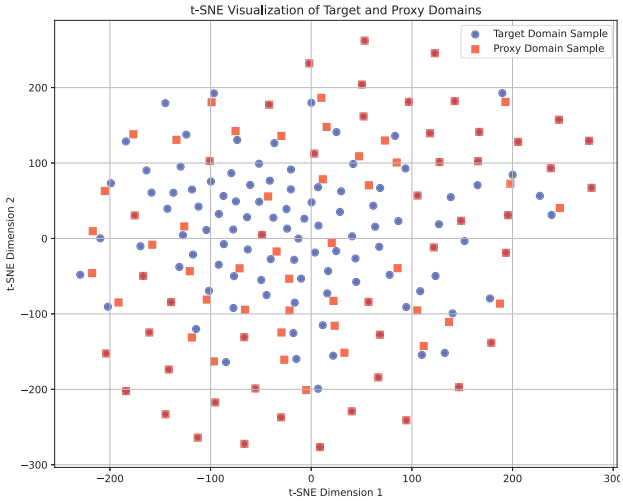

Fig. 4. Visualization of the proxy and target domain EEG samples from the subject 1 in the BNCI2014001-2 dataset.

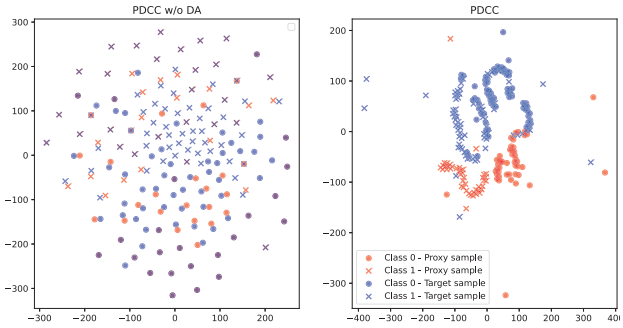

Fig. 5. Visualization of the proxy and target samples from the subject 1 in the BNCI2014001-2 dataset before and after domain adaptation.

struction, we visualize the sample selection results respectively obtained by these two metrics in Figure 6. Due to the additional incorporation of prediction confidence into the comprehensive score metric, the inter-class distance of selected samples are significantly larger than that obtained by considering the prediction consistency only, corresponding to better discriminative property. Further, such improved discriminative ability is beneficial for the conditional distribution modeling in the following domain adaptation stage.

In Figure 7, the bar chart shows that the prediction accuracy achieved by our proposed method, which utilizes the comprehensive score-based proxy domain, surpasses that of the prediction consistency-based one on the majority of the subjects in the BNCI2014001-2 dataset.

Figure 8 illustrates that the sample size determined by our method demonstrates greater robustness. In contrast, relying

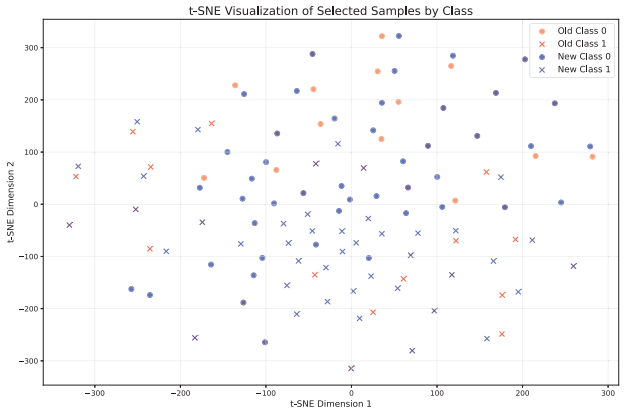

Fig. 6. Visualization of the selected EEG samples respectively obtained by the comprehensive score (in blue color) and the prediction consistency metric (in red color) for subject 1 in the BNCI2014001-2 dataset.

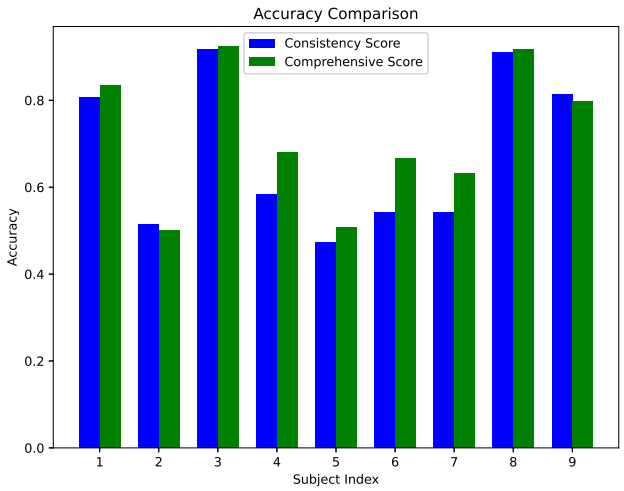

Fig. 7. Performance comparison between the comprehensive score-based proxy domain and the prediction consistency-based proxy domain on the BNCI2014001-2 dataset.

solely on prediction consistency often results in significant variations in the number of selected samples across different subjects. By comparison, our approach achieves a more stable and consistent selection of samples.

F. Ablation Analysis

We present the results of ablation studies to evaluate the individual and combined contributions of three key strategies, i.e., Ensemble Learning (EL), Proxy Domain (PD), and Transfer Learning (TL). The results are summarized in Table IX.

The baseline model does not employ any of the three strategies, instead using only an LDA classifier trained on source domain data to predict target domain data. Under this setting, the model achieves an average performance of 66.75% across the four datasets. Introducing TL alone increases the average performance to 68.91%, demonstrating its effectiveness in leveraging the auxiliary pre-trained knowledge. Similarly, using PD alone results in moderate improvement, with an average performance of 68.28%. Combining PD and TL further enhances performance to 69.27%, highlighting the

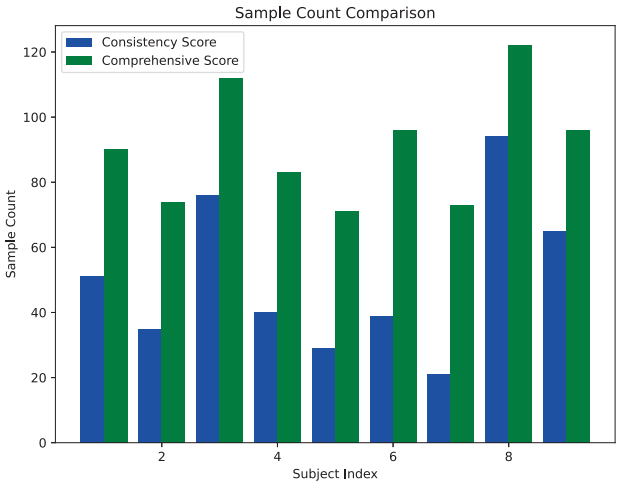

Fig. 8. Number of samples selected by traditional and proposed methods on the BNCI2014001-2 dataset.

TABLE IX
THE RESULTS OF ABLATION STUDY(%).

Strategy			Dataset				Average
EL	PD	TL	D1	D2	D3	D4	
✗	✗	✗	68.71	51.88	73.97	72.42	66.75
✗	✗	✓	71.27	55.03	75.05	74.28	68.91
✗	✓	✗	70.14	54.86	74.99	73.13	68.28
✗	✓	✓	72.78	56.21	72.84	75.23	69.27
✓	✗	✗	71.71	54.22	74.11	72.23	68.07
✓	✗	✓	73.16	56.34	75.27	76.31	70.27
✓	✓	✗	74.60	55.36	76.03	72.85	69.71
✓	✓	✓	76.91	58.59	77.55	76.46	72.38

D1, D2, D3, and D4 respectively denote the four datasets of BNCI2014001-2, BNCI2014001-4, BNCI2014002 and BNCI2015001.

complementary nature of these strategies. When all the three strategies (EL, PD, and TL) are applied simultaneously, the model achieves its best performance that the average accuracy is 72.38%, which is a substantial improvement over the baseline. This underscores the synergistic effect of integrating these strategies into our proposed method.

In summary, the ablation study demonstrates that each individual strategy positively impacts the model performance, and their combination yields the most significant enhancement. These results validate the effectiveness and robustness of the proposed PDCC model.

G. Parameters Sensitivity Analysis

In this subsection, we validated the performance of our proposed PDCC model in terms of the two important parameters, i.e., the threshold t and harmonic parameter α, which are involved in calculating the comprehensive score and further determining the proxy domain construction.

In Figure 9, we show the variation of the classification performance in terms of threshold t on the BNCI2014001-2 dataset. It is evident that parameter t has different impacts on different subjects. As t varies, the classification accuracy for most subjects demonstrates a general upward trend, with certain subjects exhibiting notable improvement at specific parameter values. For instance, subjects 1, 3, and

6 achieved relatively high classification accuracy, maintaining a stable performance as t increased. This indicates that PDCC demonstrates greater robustness for these subjects. In contrast, subjects 2 and 9 exhibited substantial fluctuations in accuracy across different t values, suggesting that the model is more sensitive to the choice of t when handling individual differences among these subjects. Therefore, it is reasonable by setting the parameter t within the range of [0.66, 0.70] to yield optimal results.

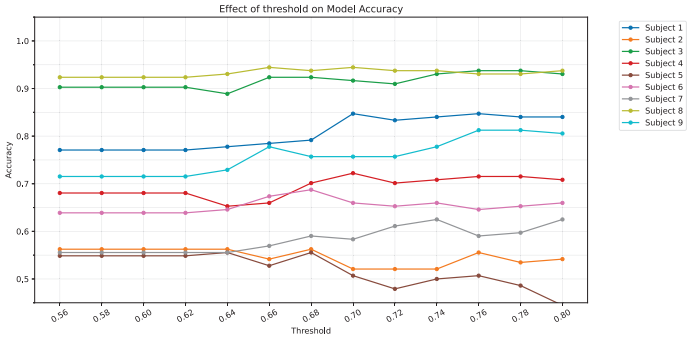

Fig. 9. Sensitivity for the threshold t on the BNCI2014001-2 dataset.

Generally, if the threshold is too low, excessive samples would be selected which are not representative for both source and target domains and may adversely impact the domain adaptation stage. Conversely, an excessively high threshold reduced the number of selected samples, omitting valuable proxy domain samples that effectively represent the source domain. This led to the insufficient capture of source domain knowledge, ultimately diminishing the subsequent domain adaptation performance. In Table X, we present the number of samples selected for constructing proxy domains under varying threshold values. As observed, the number decreases progressively as the threshold increases. For instance, when the threshold is 0.60, 32 samples were selected from BNCI2014001-4; however, no sample was selected when the threshold was enlarged to 0.75. In contrast, 346 samples were retained even at a higher threshold of 0.75 for the BNCI2015001 dataset. This disparity is attributed to the higher average prediction accuracy of this dataset, indicating a closer alignment between the source and target domains. Conversely, the significant distribution differences corresponding to the more involved classes in the BNCI2014001-4 dataset between the source and target domains resulted in fewer samples meeting the stricter criteria. These findings highlight the robustness and reliability of our proposed comprehensive score metric-based proxy domain construction method in reflecting domain alignment and ensuring sample quality.

TABLE X
THE NUMBERS OF SELECTED SAMPLES FOR THE PROXY DOMAIN CONSTRUCTION IN TERMS OF DIFFERENT THRESHOLD VALUES.

threshold (t)	**2014001-2**	**2014001-4**	**2014002**	**2015001**
0.60	144	32	160	400
0.65	127	25	137	394
0.70	102	16	78	374
0.75	70	0	34	346

In Figure 10, the sensitivity for the balance parameter α on the BNCI2014001-2 dataset is provided. It is observed that performance of PDCC is generally stable by setting α within [0.60, 0.70], indicating the importance of simultaneously taking prediction consistency and confidence into consideration in constructing the proxy domain.

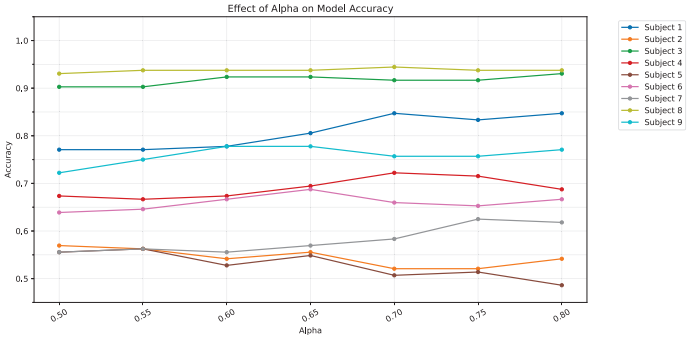

Fig. 10. Sensitivity for parameter α on the BNCI2014001-2 dataset.

V. CONCLUSION

In cross-subject EEG classification where labeled samples from multiple source subjects are available while the target subject's samples remain unlabeled, inter-subject data distribution divergences are often aligned by domain adaptation. In this paper, we have proposed a source-free framework to ensure the EEG data privacy of source subjects while leveraging its utility for cross-subject knowledge transfer. Our method began with augmenting the source subjects' data volume and aligning them to train multiple local source models. Next, a Proxy Domain was constructed by comprehensively considering the prediction Consistency and Confidence (PDCC) of source models on target EEG samples, which not only encapsulated the knowledge from the source domain but also characterized the data properties of target domain. Finally, domain adaptation was applied between the proxy domain and target domain. Experimental results on four benchmark EEG datasets demonstrated that our proposed PDCC method outperformed eleven existing approaches, including both some state-of-the-art non-source free and source-free knowledge transfer methods. Generally, PDCC is belonging to the feature representation-based transfer learning paradigm. As our future work, we will explicitly consider the differences among the source subjects to develop the multi-source extension of the PDCC model.

REFERENCES

[1] M. Y. Naser and S. Bhattacharya, "Towards practical BCI-driven wheelchairs: A systematic review study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1030–1044, 2023.
[2] Y. Chen, Y. Peng, J. Tang, T. A. Camilleri, K. P. Camilleri, W. Kong, and A. Cichocki, "EEG-based affective brain-computer interfaces: recent advancements and future challenges," *J. Neural Eng.*, vol. 22, no. 3, p. 031004, 2025.
[3] S. Gong, K. Xing, A. Cichocki, and J. Li, "Deep learning in EEG: Advance of the last ten-year critical period," *IEEE Trans. Cogn. Develop. Syst.*, vol. 14, no. 2, pp. 348–365, 2022.
[4] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[5] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016," *IEEE Trans. Cogn. Develop. Syst.*, vol. 14, no. 1, pp. 4–19, 2022.

[6] Y. Peng, W. Wang, W. Kong, F. Nie, B.-L. Lu, and A. Cichocki, "Joint feature adaptation and graph adaptive label propagation for cross-subject emotion recognition from EEG signals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 1941–1958, 2022.

[7] H. Wu, Z. Ma, Z. Guo, Y. Wu, J. Zhang, G. Zhou, and J. Long, "Online privacy-preserving EEG classification by source-free transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3059–3070, 2024.

[8] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," *J. Neural Eng.*, vol. 17, no. 1, p. 016041, 2020.

[9] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022.

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

[11] ——, "Different set domain adaptation for brain-computer interfaces: A label alignment approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1091–1108, 2020.

[12] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, 2020.

[13] Y. Peng, H. Liu, W. Kong, F. Nie, B.-L. Lu, and A. Cichocki, "Joint EEG feature transfer and semisupervised cross-subject emotion recognition," *IEEE Trans. Ind. Inform.*, vol. 19, no. 7, pp. 8104–8115, 2023.

[14] K. Xia, W. Duch, Y. Sun, K. Xu, W. Fang, H. Luo, Y. Zhang, D. Sang, X. Xu, F.-Y. Wang *et al.*, "Privacy-preserving brain-computer interfaces: A systematic review," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 5, pp. 2312–2324, 2023.

[15] J. N. Kundu, N. Venkat, R. M. V, and R. V. Babu, "Universal source-free domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn.*, 2020, pp. 4544–4553.

[16] J. Li, Z. Yu, Z. Du, L. Zhu, and H. T. Shen, "A comprehensive survey on source-free domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5743–5762, 2024.

[17] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2020, pp. 6028–6039.

[18] W. Zhang, Z. Wang, and D. Wu, "Multi-source decentralized transfer for privacy-preserving BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2710–2720, 2022.

[19] Y. Zhao, S. Feng, C. Li, R. Song, D. Liang, and X. Chen, "Source-free domain adaptation for privacy-preserving seizure prediction," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 2787–2798, 2024.

[20] D. Liang, A. Liu, L. Wu, C. Li, R. Qian, and X. Chen, "Privacy-preserving multi-source semi-supervised domain adaptation for seizure prediction," *Cogn. Neurodynamics*, vol. 18, pp. 3521–353, 2024.

[21] W. Zhang and D. Wu, "Lightweight source-free transfer for privacy-preserving motor imagery classification," *IEEE Trans. Cogn. Develop. Syst.*, vol. 15, no. 2, pp. 938–949, 2023.

[22] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz *et al.*, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, 2012.

[23] D. Steyrl, R. Scherer, J. Faller, and G. R. Müller-Putz, "Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier," *Biomed. Eng./Biomed. Techn.*, vol. 61, no. 1, pp. 77–86, 2016.

[24] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, "Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 313–319, 2012.

[25] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain–computer interface (BCI)," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 771–781, 2017.

[26] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[29] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.

[30] X.-C. Zhong, Q. Wang, D. Liu, Z. Chen, J.-X. Liao, J. Sun, Y. Zhang, and F.-L. Fan, "EEG-DG: A multi-source domain generalization framework for motor imagery EEG classification," *IEEE J. Biomed. Health Informatics*, vol. 29, no. 4, pp. 2484–2495, 2025.

[31] T. Jia, L. Meng, S. Li, J. Liu, and D. Wu, "Federated motor imagery classification for privacy-preserving brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, pp. 3442–3451, 2024.

[32] K. Xia, L. Deng, W. Duch, and D. Wu, "Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 11, pp. 3365–3376, 2022.

[33] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.