



OPEN ACCESS

EDITED BY

Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing
Medical University, China

REVIEWED BY

Zhihua Lu,
Ningbo University, China
Jun Niu,
Ningbo University, China
Detian Zhang,
Soochow University, China

*CORRESPONDENCE

Di Zhou

✉ sion2005@asasu.edu.cn

Yufang Dan

✉ 2018020@nbpt.edu.cn

RECEIVED 03 July 2024

ACCEPTED 03 October 2024

PUBLISHED 01 November 2024

CITATION

Dan Y, Zhou D and Wang Z (2024)
Discriminative possibilistic clustering
promoting cross-domain emotion
recognition.

Front. Neurosci. 18:1458815.

doi: 10.3389/fnins.2024.1458815

COPYRIGHT

© 2024 Dan, Zhou and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Discriminative possibilistic clustering promoting cross-domain emotion recognition

Yufang Dan^{1*}, Di Zhou^{2*} and Zhongheng Wang³

¹Ningbo Polytechnic, Institute of Artificial Intelligence Application, Zhejiang, China, ²Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Sichuan, China, ³Ningbo Vichnet Technology Co., Ltd., Zhejiang, China

The affective Brain-Computer Interface (aBCI) systems strive to enhance prediction accuracy for individual subjects by leveraging data from multiple subjects. However, significant differences in EEG (Electroencephalogram) feature patterns among subjects often hinder these systems from achieving the desired outcomes. Although studies have attempted to address this challenge using subject-specific classifier strategies, the scarcity of labeled data remains a major hurdle. In light of this, Domain Adaptation (DA) technology has gradually emerged as a prominent approach in the field of EEG-based emotion recognition, attracting widespread research interest. The crux of DA learning lies in resolving the issue of distribution mismatch between training and testing datasets, which has become a focal point of academic attention. Currently, mainstream DA methods primarily focus on mitigating domain distribution discrepancies by minimizing the Maximum Mean Discrepancy (MMD) or its variants. Nevertheless, the presence of noisy samples in datasets can lead to pronounced shifts in domain means, thereby impairing the adaptive performance of DA methods based on MMD and its variants in practical applications to some extent. Research has revealed that the traditional MMD metric can be transformed into a 1-center clustering problem, and the possibility clustering model is adept at mitigating noise interference during the data clustering process. Consequently, the conventional MMD metric can be further relaxed into a possibilistic clustering model. Therefore, we construct a distributed distance measure with Discriminative Possibilistic Clustering criterion (DPC), which aims to achieve two objectives: (1) ensuring the discriminative effectiveness of domain distribution alignment by finding a shared subspace that minimizes the overall distribution distance between domains while maximizing the semantic distribution distance according to the principle of "sames attract and opposites repel"; and (2) enhancing the robustness of distribution distance measure by introducing a fuzzy entropy regularization term. Theoretical analysis confirms that the proposed DPC is an upper bound of the existing MMD metric under certain conditions. Therefore, the MMD objective can be effectively optimized by minimizing the DPC. Finally, we propose a domain adaptation in Emotion recognition based on DPC (EDPC) that introduces a graph Laplacian matrix to preserve the geometric structural consistency between data within the source and target domains, thereby enhancing label propagation performance. Simultaneously, by maximizing the use of source domain discriminative information to minimize domain discrimination errors, the generalization performance of the DA model is further improved. Comparative experiments on several representative domain adaptation learning methods using multiple EEG datasets (i.e., SEED and SEED-IV) show that, in most cases, the proposed method exhibits better or comparable consistent generalization performance.

KEYWORDS

domain adaptation, probabilistic clustering, maximum mean discrepancy, fuzzy entropy, Electroencephalogram

1 Introduction

Within the research domain of affective computing (Mühl et al., 2014), Automatic Emotion Recognition (AER; Dolan, 2002) has garnered extensive interest and attention from researchers in the field of computer vision (Kim et al., 2013; Zhang et al., 2017). To date, numerous emotion recognition methods based on Electroencephalogram (EEG) have been successively proposed (Zheng, 2017; Li X. et al., 2018; Pandey and Seeja, 2019; Jenke et al., 2014; Musha et al., 1997). From the perspective of machine learning, EEG-based AER tasks can be formulated as classification or regression problems for processing (Kim et al., 2013; Zhang et al., 2017). In such tasks, state-of-the-art AER techniques often involve training classifiers on data from multiple subjects to achieve precise emotion recognition. However, classifiers that rely on specific subjects typically exhibit limited generalization capabilities due to significant variations in emotional expression patterns across different subjects (Pandey and Seeja, 2019). By optimizing feature representations and learning models (Li et al., 2018a, 2018b, 2019; Du et al., 2020; Song et al., 2018; Zhong et al., 2020; Zheng and Lu, 2015a, 2015b), the accuracy of emotion recognition has been significantly improved. Given the inherent individual differences in EEG-based AER, applying the learned classifiers to unseen subjects may yield unsatisfactory results based on qualitative and empirical observations (Ghifary et al., 2017; Lan et al., 2018; Jayaram et al., 2016; Zheng and Lu, 2016; Wang et al., 2022). To address this issue, one potential solution is to adopt subject-specific classifiers, but this approach is often impractical due to the scarcity of training data. Furthermore, even if this approach is feasible in certain scenarios, fine-tuning the classifier to maintain its good recognition performance is indispensable, partly because EEG signals from the same subject can change over time (Zhou et al., 2022). To tackle these challenges, the Domain Adaptation (DA) learning paradigm has emerged and has been widely and effectively applied (Patel et al., 2015; Dan et al., 2022; Tao et al., 2017, 2021, 2022; Zhang Y. et al., 2019). This paradigm aims to enhance the learning performance of the target domain (where labeled samples are scarce or absent) by transferring and leveraging prior knowledge from other related but differently distributed domains (i.e., source or auxiliary domains).

To achieve effective knowledge transfer across different professional domains, the crux lies in ensuring similarity or consistency in data distributions between the source domain and the target domain. Since the discrepancies in data distributions are particularly pronounced and more complex in sentiment analysis and other emotionally related fields, a significant challenge is posed for Domain Adaptation (DA) learning at present. Currently, a commonly adopted approach in the field of DA learning is to address distribution differences by identifying features (or samples) that remain constant across different domains (Pan and Yang, 2010; Patel et al., 2015). In order to more effectively leverage these domain-invariant features, traditional shallow DA models have gradually evolved into deep DA models. These deep DA models (Long et al., 2015, 2016; Chen et al., 2019; Lee et al., 2019; Ding et al., 2018; Tang and Jia, 2019) with their

profound feature transformation capabilities have made notable advancements in adaptive learning. Although deep DA models can mitigate the impact of distribution differences between domains when dealing with large datasets, they have not yet fully resolved the issue of domain shift. Deep DA methods exhibit robust performance. The specific mechanisms underlying these effects remain unclear, since these advantages may stem from various factors such as deep feature representations, model fine-tuning, or adaptive regularization. More importantly, the learning outcomes of these methods still lack adequate explanation and validation at both theoretical and practical levels.

To better characterize the generalization capability of shallow DA algorithms, existing theoretical research on DA has proposed [i.e., the generalization error bound (Ben-David et al., 2010)] for DA by the following inequality:

$$e_{\mathcal{T}}(h) \leq e_{\mathcal{S}}(h) + d_{\mathcal{H}}(\mathcal{D}_{\mathcal{S}}, \mathcal{D}_{\mathcal{T}}) + \min \left\{ \begin{array}{l} \varepsilon_{\mathcal{D}_{\mathcal{S}}} \left[\left| f_{\mathcal{S}}(x) - f_{\mathcal{T}}(x) \right| \right], \\ \varepsilon_{\mathcal{D}_{\mathcal{T}}} \left[\left| f_{\mathcal{S}}(x) - f_{\mathcal{T}}(x) \right| \right] \end{array} \right\} \quad (1)$$

Based on Equation 1, it can be seen that the upper bound of the expected error $e_{\mathcal{T}}(h)$ of the target hypothesis h is mainly determined by three aspects: the expected error $e_{\mathcal{S}}(h)$ of the source domain hypothesis, the distribution discrepancy $d_{\mathcal{H}}(\mathcal{D}_{\mathcal{S}}, \mathcal{D}_{\mathcal{T}})$ between the source and target domains, and the discrepancy in the labeling functions between the two domains. Therefore, to reduce the DA generalization error, in addition to retaining the discriminative information of the source domain (the first aspect), it is also necessary to consider reducing the distribution discrepancy between domains (the second aspect) and the discrepancy in the labeling functions between domains (the third aspect). Accordingly, current mainstream DA methods can be divided into those based on distribution alignment (including instance weighting and feature transformation) and those based on classifier model alignment (Gretton et al., 2007; Chu et al., 2013; Pan et al., 2011; Long et al., 2013; Luo et al., 2020; Baktashmotlagh et al., 2013; Ganin et al., 2016; Kang et al., 2022; Liang et al., 2019; Tao et al., 2012; Tao et al., 2015; Tao et al., 2016; Tao et al., 2019).

To address the challenges posed by domain distribution shift, early research endeavors adopted an instance weighting strategy. This strategy involves calculating the probability of each sample belonging to either the source or target domain, referred to as the instance's membership weight, and subsequently mitigating the domain shift issue by re-weighting the samples. Among these techniques, the Maximum Mean Discrepancy (MMD; Gretton et al., 2007) has been widely applied due to its simplicity and effectiveness. However, its optimization process is often isolated from the training of the classifier, making it difficult to achieve simultaneous optimization of both. In response to this limitation, Chu et al. (2013) proposed a domain-adaptive classifier that integrates instance weighting. In order to further transcend the

constraints imposed by the assumption of conditional distribution consistency in instance weighting methods, feature transformation methods have become a focal point of research in recent years (Pan et al., 2011; Long et al., 2013; Luo et al., 2020; Baktashmotlagh et al., 2013; Kang et al., 2022; Liang et al., 2019). For instance, Pan et al. (2011) introduced the Transfer Component Analysis (TCA) method. This approach aims to minimize the MMD distance between the source and target domain distributions by learning a transformation matrix while preserving the original variance of the data. Unfortunately, it does not take into account the semantic consistency alignment across different domains. To address this, Long et al. (2013) proposed the Joint Domain Adaptation (Joint DA, JDA) method, which not only considers feature distribution alignment but also accounts for conditional distribution alignment, and initializes the category labels of the target domain using pseudo-labels. More recently, Luo et al. (2020) presented a unified domain adaptation framework, DGA-DA, that synthesizes the ideas of TCA and JDA. This framework introduces a strategy of inter-domain dissimilar exclusion and retains the geometric structure information of domain data, thereby effectively facilitating the propagation of target labels. Most existing affective models are based on deep transfer learning methods utilizing domain-adversarial neural networks (DANN; Ganin et al., 2016), as seen in studies (Li et al., 2018; Li et al., 2018d; Du et al., 2020; Luo et al., 2018; Sun et al., 2022). DANN aims to find a shared feature representation for both source and target domains with indistinguishable distribution differences while maintaining the predictive capacity of the features on source samples for specific classification tasks. Additionally, the framework preserves the geometric structure information of domain data to ensure effective target label propagation. Baktashmotlagh et al. (2013) introduced the Domain Invariant Projection (DIP) algorithm. This algorithm applies a polynomial kernel to the MMD metric, aiming to construct a compact shared feature space and minimize the intra-class scatter through a clustering-based method.

A comprehensive review of the current state of Domain Adaptation (DA) research reveals that Maximum Mean Discrepancy (MMD) is a widely employed measure of distribution distance in the field of feature transformation. Traditional MMD-based DA methods primarily focus on reducing the distribution discrepancies between different domains. However, these methods often overlook the statistical (clustering) structure of the target domain data, which can adversely affect the inference of target domain labels. To address this issue, Kang et al. (2022) proposed an unsupervised DA method known as the Contrastive Adaptation Network. This approach hypothesizes the labels of the target domain through clustering and utilizes contrastive difference metrics in multiple fully connected layers to adjust the feature representations, aiming to minimize intra-class domain differences and maximize inter-class domain differences. During the training process, the hypotheses of target labels and feature representations are iteratively optimized in a cross-manner to enhance the model's generalization capability. Concurrently, inspired by clustering methods, Liang et al. (2019) developed an effective Domain Invariant Projection Ensemble method. This method leverages clustering principles to seek the optimal projections among various categories within the domain, thereby narrowing the semantic gap between domains and enhancing the cohesion of intra-domain

categories. Nevertheless, these methods essentially remain within the scope of MMD-based feature transformation DA methods.

It is worth noting that existing MMD-based methods did not fully consider the impact of intra-domain noise when measuring domain distribution distance. In real scenarios, noise inherently exists in domains. The intra-domain noise can lead to mean-shift problems on distance measure by traditional MMD methods and their variants. This phenomenon to some extent is affecting the generalization performance of MMD-based DA methods (Tao et al., 2023).

Fortunately, the proposed possibilistic clustering models (Krishnapuram and Keller, 1993; Dan et al., 2021) offer a comprehensive solution to the aforementioned issues. Unlike hard clustering models, possibilistic clustering effectively suppresses noise interference during the data clustering process (Dan et al., 2021). Inspired by this, the traditional MMD metric is relaxed into one-center center objective to tackle the mean shift problem of domain distributions in noisy environments. The possibilistic one-center center model with a fuzzy entropy regularization term is reconstructed. Then, we propose a distributed distance measure with Discriminative Possibilistic Clustering (DPC) criterion, and further develop a domain adaptation in Emotion recognition based on DPC model (EDPC). EDPC mainly comprises two jointly optimized components: DPC and classifier. First, DPC seeks a shared latent space to achieve discriminative alignment of overall domain distribution and inter-domain semantic distribution, the samples adhering to the principle of attract each other with similar class and repulse each other with different class. Clustering membership is adopted to indicate the likelihood of each sample being consistent with the overall domain distribution. It aims to suppress the impact of noisy data during domain matching and enhance the robustness and effectiveness of domain distribution clustering metrics. Second, in the classifier model learning phase, a graph Laplacian regularization term is introduced to preserve the local geometric structure of sample in the latent domain space to improve the performance of domain label propagation. Finally, we gain superior knowledge transfer performance by maximizing the utilization of source domain discriminative information minimizing the discriminative error in the target domain. The EDPC method will be evaluated in emotion recognition. Compared to other state-of-the-art methods, this approach demonstrates significant performance improvements in most cases. The major novelties of the proposed EDPC are summarized as follows:

- 1 To address the noisy problem in existed methods, we establish a robust distributed distance measure with Possibilistic Clustering by relaxing the traditional MMD metric criterion into a possibilistic one-center center model. From a theoretical point of view, we proposed measure method serves as an upper bound to the traditional MMD metric under certain conditions.
- 2 The samples follow the principle of "attract each other with similar class and repulse each other with different class," we extend and establish a robust distributed distance measure with Discriminative Possibilistic Clustering (i.e., DPC) metric criterion. Based on this, we seek to optimize a domain-invariant subspace that achieves joint alignment of overall domain distribution and inter-domain semantic distribution.
- 3 We propose a unified domain adaptation Emotion recognition model based on DPC (i.e., EDPC). An iterative optimization

algorithm is provided. We finally prove its consistent convergence and propose a generalization error bound for the model based on Rademacher complexity theory.

- 4 Extensive experiments on six real-world datasets validate the robust effectiveness of the proposed method.

2 Proposed method

In DA learning, the source domain is defined as $\mathcal{D}_S = \{x_i^s, y_i^s\}_{i=1}^n$,

where the sample set is defined as $X^s = [x_1^s, \dots, x_n^s] \in \mathbb{R}^{d \times n}$, and the corresponding class labels are defined as $Y^s = \{y_1, \dots, y_n\}^T \in \{0, 1\}^{n \times C}$. Here, $Y^s = \{y_1, \dots, y_n\}^T \in \{0, 1\}^{n \times C}$ is a one-hot encoded vector; if \hat{x}_i

belongs to the j -th class, then $y_j^i = 1$. The unlabeled target domain is defined as $\mathcal{D}_T = \{\hat{x}_j^t\}_{j=1}^m$, where the sample set and unknown sample

labels during training are $\hat{X}^t = [\hat{x}_1^t, \dots, \hat{x}_m^t] \in \mathbb{R}^{d \times m}$, $Y^t = \{y_1, \dots, y_m\}^T \in \mathbb{R}^{m \times C}$, respectively. We further

define $\hat{X} = [\hat{X}^s, \hat{X}^t] \in \mathbb{R}^{d \times N}$ and $Y = [Y^s, Y^t] \in \mathbb{R}^{N \times C}$, where $N = n + m$. Let μ_s and μ_t be the mean values of the samples in the source and target domains, respectively. Our work has the following assumptions:

- 1 The source domain distribution \mathbb{P} and the target domain distribution \mathbb{Q} are different, but they share the same feature space \mathcal{X} , i.e., $\mathbb{P}(\mathcal{X}_S) \neq \mathbb{Q}(\mathcal{X}_T)$ and $\mathcal{X}_S = \mathcal{X}_T$, where $\mathcal{X}_S, \mathcal{X}_T \in \mathcal{X}$ are the feature spaces of the source domain and the target domain, respectively.
- 2 The class-conditional probability distributions between domains are different but they share the same label space \mathcal{Y} , i.e., $\mathbb{P}(\mathcal{Y}_S | \mathcal{X}_S) \neq \mathbb{Q}(\mathcal{Y}_T | \mathcal{X}_T)$, where $\mathcal{Y}_S, \mathcal{Y}_T \in \mathcal{Y}$ are the label spaces of the source domain and the target domain, respectively.

The rest of this section is organized as follows. In Subsection 2.1, we give the general formulation of the proposed method. We describe the details of the general formulation in Subsection 2.2 about discriminative possibilistic clustering (i.e., DPC) formulation and Subsection 2.3 about the classifier in emotion recognition with DPC. The final detail formulation is showed in Subsection 2.4.

2.1 General formulation

To effectively align domains' distribution and achieve maximum knowledge transferring, this paper seeks to optimize a domain-invariant subspace $P \in \mathbb{R}^{d \times r}$, where r is the subspace dimension and making P be an orthogonal subspace (i.e., $P^T P = I, I \in \mathbb{R}^{r \times r}$ is the identity matrix). In other words, Our work is to achieve $\mathbb{P}(P(\mathcal{X}_S)) \approx \mathbb{Q}(P(\mathcal{X}_T))$ and $\mathbb{P}(\mathcal{Y}_S | P(\mathcal{X}_S)) \approx \mathbb{Q}(\mathcal{Y}_T | P(\mathcal{X}_T))$ in the optimized domain-invariant subspace P .

For the problem of DA in complex structures and noisy environments, we aim to improve the robustness of distribution distance metrics for DA and enhance generalization in the target domain. Based on the DA generalization error theory (Ben-David

et al., 2010), we explore to achieve the following two core objectives: First, we construct a robust distribution distance metric that can resist the impact of noise for addressing the issue of domain mean-shift. The differences of domains distribution can be selectively corrected. Second, we effectively perform semantic reasoning in the target domain by maintaining the geometric structure consistency of samples in domain and connecting the discriminative information of the source domain and minimizing the discriminative error in the target domain. The a highly generalizable target domain classifier be constructed. Therefore, our general framework can be described as following:

$$\begin{aligned} \Theta(\lambda, P, Y^t, W) &= \min R(f_W(P(\hat{X})), Y) \\ &+ \alpha \Omega_\lambda(P(\hat{X}^s), P(\hat{X}^t)) - \beta \Omega_{rep}(P(\hat{X}^s), P(\hat{X}^t)) \\ &s.t. P^T P = I \end{aligned} \tag{2}$$

where the first term is the empirical risk function of the decision function $f_W(P(\hat{X}))$ on the domain sample, W represents the model parameters, and $R(\cdot)$ is the empirical loss function. It can be chosen according to different application needs. The second term is a regularization term that measures the semantic distribution discrepancy between \mathcal{D}_S and \mathcal{D}_T in the domain-invariant subspace P . λ denotes the clustering membership degree. The last term aims to enhance the discriminative ability for the second term. The orthogonal constraint $P^T P = I$ serves as a regularization constraint on P . It does not need to be described in the objective function. Two hyper-parameters (i.e., $\alpha \geq 0$ and $\beta \geq 0$) are used to measure the semantic distribution discrepancy and the inter-classes distribution discrepancy of the two domains, respectively. Therefore, the subspace P , the membership degree λ , and the decision function $f_W(\cdot)$ can be learned simultaneously during optimizing the Problem 2.

2.2 Discriminative possibilistic clustering formulation

The rest of this subsection is organized as follows. In 2.2.1, we give the motivation of the proposed method DPC and proof the MMD metric can be relaxed and modeled as a special one-center clustering problem. We then explain all details of DPC metric in 2.2.2.

2.2.1 Motivation

In a certain Reproducing Kernel Hilbert Space (RKHS) \mathcal{H} , the original space data is transformed into feature representations in the RKHS through a nonlinear mapping $\phi: \mathbb{R}^d \rightarrow \mathcal{H}$. The corresponding

kernel function is defined as $K(\cdot, \cdot): \hat{X} \times \hat{X} \rightarrow \mathbb{R}$, where

$$K(\hat{x}_1, \hat{x}_2) = \phi(\hat{x}_1), \phi(\hat{x}_2)_{\mathcal{H}}, \hat{x}_1, \hat{x}_2 \in \hat{X}. \text{ Here, } \phi(\hat{x}) \text{ is a feature}$$

mapping function that maps samples from the original space to a high-dimensional or even infinite-dimensional space (i.e., the RKHS

\mathcal{H}). For uniform representation, the mapping result is defined as following:

$$x = \begin{cases} \hat{x}, & \text{linear kernel mapping} \\ k(\hat{x}, \cdot), & \text{non-linear kernel mapping} \end{cases}$$

where $k(\hat{x}, \cdot) = [k(\hat{x}, \hat{x}_1), k(\hat{x}, \hat{x}_2), \dots, k(\hat{x}, \hat{x}_n)]^T$ is called the empirical kernel mapping [32]. Then X^s and X^t respectively represent the unified representation of linear kernel mapping and nonlinear kernel mapping of \hat{X}^s and \hat{X}^t . Accordingly, $X = [X^s, X^t]$.

Therefore, we try to learn a DA learning machine $f(\phi(\hat{x})) = W^T \phi(\hat{x})$. Thus, the formal description of Problem 2 can

be rewritten in Equation (3):

$$\begin{aligned} \Theta(\lambda, P, Y^t, W) = & \min R(f_W(P(X)), Y) + \alpha \Omega_\lambda(P(X^s), P(X^t)) \\ & - \beta \Omega_{rep}(P(X^s), P(X^t)) \end{aligned} \quad (3)$$

s.t. $P^T P = I$

In classical DA research, a commonly adopted strategy to address the domain mean-shift problem is to reduce the distribution distance between the source domain and the target domain in a certain latent feature space. This involves identifying a feature descriptor space, where MMD is a frequently used method to measure the distribution difference between two domains. MMD utilizes the framework of RKHS to effectively quantify the gap between two distributions [as described in references (Bruzzone and Marconcini, 2010; Gretton et al., 2010)]. It is assumed that there exists a set \mathcal{G} containing all domain-invariant transformation matrices P (i.e., $G = \{P \in \mathbb{R}^r | P^T P = I\}$) in this framework. The maximum empirical mean discrepancy between the source domain distribution \mathbb{P} and the target domain distribution \mathbb{Q} can be defined in Equation (4):

$$MMD_G(X^s, X^t) = \left\| \frac{1}{n} \sum_{i=1}^n P(x_i^s) - \frac{1}{m} \sum_{j=1}^m P(x_j^t) \right\|_H^2 \quad (4)$$

According to Hoeffding's inequality theorem, when the domain sample size is sufficiently large (or approaches infinity), the expected difference and the empirical mean difference are approximately (or equal). To illustrate the generalized connection between the traditional MMD criterion and the mean clustering model, the following theorem is presented:

Theorem 1. The MMD metric can be relaxed and modeled as a special one-center clustering problem, where the clustering center of the one-center is μ and the sample clustering membership is the vector ζ .

Proof: From the definition of empirical MMD, we have (let $X \leftarrow P(X)$ for simplify):

$$\begin{aligned} MMD(X^s, X^t) &= \left\| \frac{1}{n} \sum_{i=1}^n x_i^s - \frac{1}{m} \sum_{j=1}^m x_j^t \right\|_H^2 \\ &= \left\| \frac{1}{n} \sum_{i=1}^n x_i^s - \mu + \mu - \frac{1}{m} \sum_{j=1}^m x_j^t \right\|_H^2 \\ &\leq \left\| \frac{1}{n} \sum_{i=1}^n x_i^s - \mu \right\|_H^2 + \left\| \frac{1}{m} \sum_{j=1}^m x_j^t - \mu \right\|_H^2 \\ &= \frac{1}{n^2} \left\| \sum_{i=1}^n x_i^s - n\mu \right\|_H^2 + \frac{1}{m^2} \left\| \sum_{j=1}^m x_j^t - m\mu \right\|_H^2 \\ &= \frac{1}{n^2} \left\| \sum_{i=1}^n (x_i^s - \mu) \right\|_H^2 + \frac{1}{m^2} \left\| \sum_{j=1}^m (x_j^t - \mu) \right\|_H^2 \\ &\leq \frac{1}{n^2} \sum_{i=1}^n \|x_i^s - \mu\|_H^2 + \frac{1}{m^2} \sum_{j=1}^m \|x_j^t - \mu\|_H^2 \\ &= \sum_{k=1}^N \zeta_k \|x_k - \mu\|_H^2 \end{aligned} \quad (5)$$

where the cluster center is defined as $\mu = \delta \mu_s + (1 - \delta) \mu_t$, $0 \leq \delta \leq 1$; when $n = m$,¹ let $\delta = 0.5$; the sample membership of the one-center center $\zeta = \{\zeta_k\}_{k=1}^N$ is defined in Equation (6):

$$\zeta_k = \begin{cases} \frac{1}{n^2}, & x_k \in X^s \\ \frac{1}{m^2}, & x_k \in X^t \end{cases} \quad (6)$$

By analyzing Equation 5, it can be inferred that the one-center is μ actually constitutes an upper bound of the traditional MMD metric. It means that the MMD metric can be simplified to a special form of the one-center objective function. In this case, it is possible to minimize the MMD between different domains by optimizing this special clustering objective.

According to Theorem 1 and the explanation in Baktashmotlagh et al. (2013), it can be recognized that the MMD metric standard for domain distribution is essentially related to the clustering model. It is possible to more effectively achieve distribution alignment between different domains promote domain adaptive learning by clustering domain data. It should be noted that traditional clustering models are usually sensitive to noise (Krishnapuram and Keller, 1993), which limits MMD-based DA methods and make them prone to domain mean-shift issues in noisy environment. To address this challenge, we further explore more robust clustering methods and propose a new discriminative domain distribution distance metric criterion in following subsection.

¹ When $n \neq m$, we can draw an equal number of samples from the source domain and the target domain, respectively.

2.2.2 Discriminative possibilistic clustering

2.2.2.1 Possibilistic clustering

Recently proposed probabilistic clustering methods have been proven to effectively mitigate the negative impact of noise on clustering results (Dan et al., 2021). In light of this, this section extends the original one-center method to the realm of probabilistic one-center. Then we propose a Possibilistic Clustering distribution distance metric in domain-invariant subspace \mathbf{p} , namely PC. We extend the hard clustering approach of MMD to a soft clustering form by incorporating the concept of possibilistic clustering entropy. In this framework, each sample determines its contribution based on its distance from the overall domain mean (that is, the greater the distance of the data sample, the lower its contribution, and conversely, the more likely it is considered as noise). In this way, PC allows for the attenuation of the impact of mean-shift caused by noise during domain alignment through adjustment. Therefore, the formula for the possibilistic clustering distribution distance metric can be defined as follows:

$$PC(X^s, X^t, \lambda) = \sum_{k=1}^N \lambda_k^b \|P(x_k) - P(\mu)\|_H^2 \quad (7)$$

s.t., $0 \leq \lambda_k \leq 1, k = 1, \dots, N$

where $\lambda = [\lambda_1, \dots, \lambda_N]$ is the possibilistic clustering membership vector, the parameter b is the weight exponent of λ_k is used to adjust the uncertainty or degree of samples belong to multiple categories. To avoid trivial solutions, we set $b = 2$ in the subsequent formulas. The detailed introduction to different values of b is referred to (Krishnapuram and Keller, 1993). According to Equation 7, $PC(X^s, X^t, \lambda)$ is a possibilistic one-center objective function (with the cluster center being μ). When $\lambda_k^2 = \zeta_k$, $PC(X^s, X^t, \lambda)$ represents the aforementioned special form of one-center. Next, we verify that the proposed $PC(X^s, X^t, \lambda)$ is an upper bound of the traditional MMD metric at a certain condition by the following theorem:

Theorem 2. When the possibilistic clustering membership satisfies $\lambda_k \in \left[\frac{1}{r}, 1\right] (k = 1, \dots, N)$, the possibilistic distribution distance metric $PC(X^s, X^t, \lambda)$ is an upper bound of the traditional MMD metric.

Proof: Combining Equations 5 and 7, we can obtain:

$$\begin{aligned} MMD(X^s, X^t) &\leq \sum_{k=1}^N \zeta_k \|P(x_k) - P(\mu)\|_H^2 \\ &\leq \sum_{k=1}^N \lambda_k^2 \|P(x_k) - P(\mu)\|_H^2 \\ &= PC(X^s, X^t, \lambda) \end{aligned} \quad (8)$$

According to the value range of ζ_k , when $\lambda_k \in \left[\left(\frac{1}{r}\right), 1\right]$ and $r = \min(n, m)$, the second inequality in Equation 8 holds true and the conclusion is proved.

Guided by Theorem 1 and Theorem 2, we are able to reconstruct the traditional MMD metric into a function that targets 1-center

probabilistic clustering. From this new perspective, we can gain a profound understanding that the objective of probability distribution distance measure is not only to facilitate effective adjustment of feature distributions across different domains but also to mitigate the negative transfer effects induced by intra-domain noisy data during the training process.

In the PC described in Equation 7, its primary objective is centered on reducing the statistical distribution discrepancy between the source domain and the target domain. However, this method fails to adequately emphasize the importance of preserving the semantic structural information of instances during the process of domain distribution matching. This approach may compromise the ability to distinguish between categories within the domains. To preserve the distinctiveness of the statistical distribution structures between domains, we further propose a Discriminative PC model (i.e., DPC). This model aims to implement discriminative statistical distribution alignment between the source and target domains while adjusting the semantic distribution through a clustering hypothesis based on likelihood.

The samples adhere to the principle of “attract each other with similar class and repulse each other with different class.” DPC criterion aims to achieve dual objectives: first, it reduces the distribution bias of similar samples within different domains, thereby minimizing semantic differences between domains; second, it increases the distribution gap between different classes of samples within different domains by enhancing the differentiability of domain samples.

2.2.2.2 DPC with Intra-class alignment

In the PC model shown in Equation 7, the process of domain distribution alignment does not consider the semantic structure information of samples. This oversight may compromise the local discriminative structure between different categories within the domain. To address this issue, Tao et al. (2016) proposes further considering the semantic distribution structure between domains during the alignment process and evaluating the contribution of each sample to semantic matching. Hence, we have the following DPC framework with semantic registration functionality by extending the PC model:

$$\begin{aligned} \Omega'_\lambda(P(X^s), P(X^t)) &= \min_{\lambda_{k,c}} \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c} - P^T \mu_c\|_H^2 \quad (9) \\ &\text{s.t.}, 0 \leq \lambda_{k,c} \leq 1 \end{aligned}$$

$$\text{where } \mu_c = \delta \mu_{s,c} + (1 - \delta) \mu_{t,c}, \quad \mu_{s,c} = \frac{1}{n} \sum_{i=1}^{n_c} x_{i,c}^s,$$

$\mu_{t,c} = \frac{1}{m} \sum_{j=1}^{m_c} x_{j,c}^t, c = 0, 1, 2, \dots, C$. C is the number of classes in the

target domain, n_c is the number of samples in the c -th class of the source domain, m_c is the number of samples in the c -th class of the

target domain, $n = \sum_{c=0}^C n_c$, and $m = \sum_{c=0}^C m_c$. When $c = 0$, $\mu_{s,c}$ and $\mu_{t,c}$

represent the mean of the entire source domain and the entire target

domain, respectively. Equation 9 is the form of feature distribution alignment. When $c \in [1, 2, \dots, C]$, $\mu_{s,c}$ and $\mu_{t,c}$ represent the mean of the corresponding classes in the source domain and target domain, respectively. $\lambda_{k,c}$ is the matching contribution value of x_k belonging to the c -th category in the domain.

To enhance the robustness and effectiveness of the possibilistic clustering distribution distance measure method in handling noisy data, a fuzzy entropy regularization term, which is related to the parameter $\lambda_{k,c}$, is introduced based on Equation 9. With this improvement, the DPC for semantic alignment can be redefined as following:

$$\begin{aligned} &\Omega_\lambda(P(X^s), P(X^t)) \\ &= \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c} - P^T \mu_c\|_H^2 \\ &+ \beta \sum_{c=0}^C \sum_{k=1}^N (\lambda_{k,c}^2 \ln \lambda_{k,c}^2 - \lambda_{k,c}^2) \\ &s.t., 0 \leq \lambda_{k,c} \leq 1 \end{aligned} \tag{10}$$

where the parameter β serves as an adjustable balancing factor, aimed at ensuring that the value of the relevant data $\lambda_{k,c}$ remains at a high level to avoid obtaining trivial solutions that lack discriminative power. The improved DPC model now becomes a function that monotonically decreases as the value of $\lambda_{k,c}$ decreases. This model uses the second term in Equation 10, namely fuzzy entropy, to mitigate the adverse impact of noisy data on the model's classification decisions. An increase in fuzzy entropy signifies an enhancement in the discriminatory information content of samples, which plays a positive role in strengthening the robust effectiveness of distribution distance measures. Furthermore, the introduction of a fuzzy entropy-regularized possibility distribution distance measure model can effectively limit the influence of noisy data in domain distribution alignment, reducing the interference of noise or outlier data in the domain adaptation learning process. For more details and empirical analysis on how fuzzy entropy improves robustness, refer to the discussion in reference (Gretton et al., 2010).

2.2.2.3 DPC with Inter-class discrimination

The intra-class alignment neglects inter-class discrimination. We therefore add an additional inter-class repulsion term into the DPC model to increase the inter-class distance across domain. It enhances the semantic discriminatory of samples in the domain-invariant subspace and improves the robustness and effectiveness of domain adaptation learning. Specifically, let $\Omega_{rep}(P(X^s), P(X^t))$ be the inter-domain different classes repulsion term, which is defined as the total difference between the mean of each class L^c and the mean of all other classes $L^{(cc):cc \in \{1, \dots, C\} - \{c\}}$ (excluding class c). That is:

$$\begin{aligned} &\Omega_{rep}(P(X^s), P(X^t)) \\ &= \sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_i \in L^{(c)}} P^T x_i - \frac{1}{\sum_{cc} n_{cc}} \sum_{x_j \in L^{(cc)}} P^T x_j \right\|^2 \\ &= \sum_{c=1}^C \text{tr}(P^T X M_{cc} X^T P) \end{aligned} \tag{11}$$

where,

$$M_{cc} = \begin{cases} \frac{1}{n_c n_c}, & x_i, x_j \in L^{(c)} \\ \frac{1}{n_{cc} n_{cc}}, & x_i, x_j \in L^{(cc)} \\ \frac{1}{n_c n_{cc}}, & \begin{cases} x_j \in L^{(c)}, x_i \in L^{(cc)} \\ x_i \in L^{(c)}, x_j \in L^{(cc)} \end{cases} \\ 0 \end{cases}$$

2.3 Induction Learning

The DPC criterion effectively addresses the challenges of domain distribution alignment and noise interference. Building upon this, we dedicate to achieving two core objectives in the process of target domain knowledge inference: (1) maintaining the consistency of geometric structures between the source domain and the target domain, ensuring that label information for neighboring samples remains consistent; and (2) striving simultaneously to minimize the structural risk loss between the source domain and the target domain. Through the description of the target task, the general form of the target risk function can be described in Equation (12):

$$R(f_W(P(X)), Y) = \Omega_Y + \Omega_W \tag{12}$$

where Ω_Y represents the joint knowledge transfer and label propagation loss, which preserves the geometric structure consistency of the sample in both the source and target domains, and Ω_W includes the structural risk loss terms of the source domain and target domain. Next, we will design these two terms separately.

2.3.1 Label Propagation

Firstly, we define the undirected weighted graph on the entire domain as $G = X, M$, and let $M \in \mathbb{R}^{N \times N}$ be the weight matrix and $M_{ij} = M_{ji} \geq 0$. The calculation method for M_{ij} is defined in Equation (13):

$$M_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), & x_i \in Ne(x_j) \text{ or } x_j \in Ne(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $x_k \in Ne(x_m)$ indicates that x_k is a neighbor of x_m , σ controls the local influence range of the Gaussian kernel function and is also a hyper-parameter. The larger σ , the greater local influence range. Conversely, the smaller local influence range. When σ is fixed, the value of M_{ij} decreases monotonically as the distance between x_i and x_j increases.

By combining source domain knowledge transfer and the graph Laplacian matrix (Long et al., 2013; Wang et al., 2018), label propagation modeling is performed as:

$$\Omega_Y = \min_Y \text{tr}(Y^T LY) \tag{14}$$

where $Y = [Y_s; Y_t] \in \mathbb{R}^{N \times C}$, Y_t is the label matrix of the target domain. If a sample in the target domain is unlabeled, the corresponding label value in Y_t is all zeros. Y_s is the label matrix of the source domain. $L = M - D \in \mathbb{R}^{N \times N}$ is the Laplacian graph matrix (Long et al., 2013) and D is a diagonal matrix with $D_{ii} = \sum_{j=1}^N M_{ij}$.

2.3.2 Design of Structural Risk

In our method, the source domain classifier f_s and target domain classifier f_t are defined as $f_s = (W_s')^T X^s + b_s$ and $f_t = W_t'^T X^t + b_t$, where $b_s(b_t)$ is the bias term for the source domain (target domain), and $W_s^o(W_t^o)$ is the parameter for the source domain (target domain). Let

$$\tilde{W}_s = [W_s^o, b_s], \tilde{X}^s = [X^s, 1], \tilde{W}_t = [W_t^o, b_t], \text{ and } \tilde{X}^t = [X^t, 1].$$

Then, the two classifiers can be rewritten as $\tilde{f}_s = \tilde{W}_s^T \tilde{X}^s$ and $\tilde{f}_t = \tilde{W}_t^T \tilde{X}^t$. Let $W = [\tilde{W}_s, \tilde{W}_t]$ and $\tilde{X} = [\tilde{X}^s, \tilde{X}^t]$. By combining

the two classifiers into a single classifier, we get: $F(W) = \tilde{X}^T W$.

According to the least squares loss function, the problem of minimizing structural risk for both domains can be described as:

$$\Omega_W = \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|x_{k,c}^T W - y_k\|_H^2 + \rho \|W\|_{2,1} \tag{15}$$

where the first term is the structural risk loss term with $y_k \in Y$. The second term is the constraint of the classification model. The features can be selected by applying $l_{2,1}$ regularization. It can effectively control the model complexity to a certain extent for preventing the target classification model from over-fitting.

Since the classification task ensures the reliability of predictions through the dual prediction of the label matrix Y and the decision function W , combining Equations 14 and 15 constitutes the target classification function. We can describe it as following:

$$\begin{aligned} R(f_W(P(X), Y)) &= \text{tr}(Y^T LY) + \\ &\sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|x_{k,c}^T W - y_k\|_H^2 + \rho \|W\|_{2,1} \tag{16} \\ \text{s.t., } 0 &\leq \lambda_{k,c} \leq 1, YY^T = I \end{aligned}$$

2.4 Final formulation

Combining the intra-class attraction term in Equation 10, the inter-class repulsion term in Equation 11, and the induction learning in Equation 16, the final optimization formulation of the EDPC method can be described as following:

$$\begin{aligned} &\Theta(\lambda, P, Y^t, W) \\ &= \min_{\lambda_{k,c}, Y, W} \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c}^T W - y_k\|_H^2 + \alpha \text{tr}(Y^T LY) \\ &\quad + \rho \|W\|_{2,1} + \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c} - P^T \mu_c\|_H^2 \\ &\quad + \theta \sum_{c=0}^C \sum_{k=1}^N (\lambda_{k,c}^2 \ln \lambda_{k,c}^2 - \lambda_{k,c}^2) \\ &\quad - \beta \sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_i \in L^{(c)}} P^T x_i - \frac{1}{\sum_{cc} n_{cc}} \sum_{x_j \in L^{(cc)}} P^T x_j \right\|^2 \\ \text{s.t., } 0 &\leq \lambda_{k,c} \leq 1, YY^T = I, P^T P = I \tag{17} \end{aligned}$$

where ρ, α, θ , and β are the balance parameters.

Once all model parameters are obtained, knowledge inference in the target domain can be achieved. By maximizing the use of discriminative information from the source domain, the two classifiers \tilde{f}_s and \tilde{f}_t are linearly combined, and this linear fusion model is used for target domain knowledge inference. The fusion form can be written as follows:

$$j = \arg \max_j (y_j^t = \nu \tilde{f}_s(x_j^t) + (1 - \nu) \tilde{f}_t(x_j^t))$$

where $\nu \in [0, 1]$ is an adjustable parameter that balances the two classifiers. To emphasize the importance of discriminative information from the source domain as prior knowledge, we let $\nu = 0.9$ based on experience.

3 Optimization

The optimization problem of EDPC is a non-convex problem with respect to $\lambda_{k,c}, P, W$, and Y . This paper adopts an alternating iterative optimization strategy to solve these parameters, ensuring that each optimization variable has a closed-form analytical solution.

3.1 Update $\lambda_{k,c}$ as given w, P , and Y

Since the third and fifth terms in Equation 17 do not involve the calculation of $\lambda_{k,c}$, the optimization solution of EDPC is described as follows:

$$\begin{aligned} \min_{\lambda_{k,c}} P_1 &= \min \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c} - P^T \mu_c\|_H^2 \\ &\quad - \theta \sum_{c=0}^C \sum_{k=1}^N (-\lambda_{k,c}^2 \ln \lambda_{k,c}^2 + \lambda_{k,c}^2) + \\ &\quad + \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|P^T x_{k,c}^T W - y_k\|_H^2 \\ \text{s.t. } 0 &\leq \lambda_{k,c} \leq 1 \tag{18} \end{aligned}$$

Theorem 3. The optimal solution to the original optimization problem of the objective function (18) is:

$$\lambda_{k,c} = \exp\left(-\frac{J}{\theta}\right) \tag{19}$$

where,

$$J = \sum_{c=0}^C \sum_{k=1}^N \left\| P^T x_{k,c}^T W - y_k \right\|_H^2 + \sum_{c=0}^C \sum_{k=1}^N \left\| P^T x_{k,c} - P^T \mu_c \right\|_H^2.$$

Proof: Taking the partial derivative of Equation 18 with respect to the variable $\lambda_{k,c}$ and setting it to zero, we get:

$$\begin{aligned} \frac{\partial P_1}{\partial \lambda_{k,c}} &= 2 \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c} \left\| P^T x_{k,c} - P^T \mu_c \right\|_H^2 \\ &\quad + 2\theta \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c} \ln \lambda_{k,c}^2 \\ &\quad + 2 \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c} \left\| P^T x_{k,c}^T W - Y \right\|_H^2 = 0 \end{aligned} \tag{20}$$

By combining like terms and rearranging Equation 20, the solution for $\lambda_{k,c}$ can be obtained as Equation 19, thus proving the theorem.

According to Theorem 3, the matching contribution of any sample can be derived from Equation 19.

3.2 Update W as given $Y, P,$ and $\lambda_{k,c}$

Since the first to third terms in Equation 17 do not involve the calculation of W , the optimization solution formula for EDPC is described as follows:

$$\begin{aligned} P_2 &= \min_W \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \left\| P^T x_{k,c}^T W - y_k \right\|_H^2 + \rho \|W\|_{2,1} \\ &= \min_W \lambda \left\| P^T X^T W - Y \right\|_H^2 + \rho \|W\|_{2,1} \end{aligned} \tag{21}$$

where, $\lambda \in \mathbb{R}^{N \times C}$ and each element is $\lambda_{k,c}^2$. $\lambda_{k,c}$ represents the membership value of x_k belonging to the c -th category.

Theorem 4. The optimal solution to the original optimization problem of the objective Function (21) is:

$$W = AY \tag{22}$$

where $A = (\lambda P^T X X^T P + \rho U)^{-1} P X$.

Proof: According to Equation 19, let $\frac{\partial P_2}{\partial W} = 0$ and solve for W , that is

$$\frac{\partial P_2}{\partial W} = 2\lambda \left[X \left(X^T W - Y \right) \right] + 2\rho U W = 0 \tag{23}$$

where, $\frac{\partial \rho \|W\|_{2,1}}{\partial W} = U W$, where U is a diagonal matrix with diagonal elements $U_{ii} = \frac{1}{\|w_i\|}$, ($i=1, \dots, d$), and w_i is the i -th row vector of matrix W . By rearranging Equation 23, the analytical solution for W can be derived as given in Equation 22.

3.3 Update Y by fixing $w, P,$ and $\lambda_{k,c}$

Since the first, second, and fifth terms in Equation 17 do not involve the calculation of Y , and substituting the result $W = AY$ into Equation 17, and the constraint $Y Y^T = I$ can reduce the interference information in the obtained label matrix Y , the objective form for optimizing and solving Y is represented as:

$$\begin{aligned} P_3 &= \min_{Y^T Y = I} \text{atr} \left(Y^T L Y \right) \\ &\quad + \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \left\| P^T x_k^T W - y_k \right\|_H^2 \\ &= \min_{Y^T Y = I} \text{atr} \left(Y^T L Y \right) \\ &\quad + \lambda \left\| P^T X^T A Y - Y \right\|_H^2 \\ &= \min_{Y^T Y = I} \text{tr} \left(Y^T H Y \right) \end{aligned} \tag{24}$$

where, $H = \alpha L + \lambda B^T B$, $B = P^T X^T A - I$. The optimization Problem (24) is a standard singular value decomposition problem. Y consists of the eigenvectors of matrix H , so the optimal solution for Y can be obtained by solving the singular value decomposition of matrix H .

3.4 Update P by fixing $w, Y,$ and $\lambda_{k,c}$

From Equation 17, it can be seen that the optimization solution for EDPC can be described as:

$$\begin{aligned} P_4 &= \min_{P^T P = I} \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \left\| P^T x_{k,c}^T W - y_k \right\|_H^2 \\ &\quad + \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \left\| P^T x_{k,c} - P^T \mu_c \right\|_H^2 \\ &\quad - \beta \sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_i \in I^{(c)}} P^T x_i - \frac{1}{n_{cc}} \sum_{x_j \in I^{(c)}} P^T x_j \right\|_H^2 \end{aligned} \tag{25}$$

Theorem 5. The optimal solution to the original optimization problem of the objective Function (25) is:

$$P = \frac{1}{2} \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 x_k W y_k O^{-1} \tag{26}$$

where ϖ is the balancing parameter of the constraint term $P^T P = I$.

$$O = \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|x_{k,c} - \mu_c\|_H^2 - \beta X M_{cc} X^T + \varpi + \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 x_k x_k^T W W^T$$

Proof: According to Equation 25, let $\partial P_4 / \partial P = 0$, solve for P , it is shown in Equation (27):

$$\begin{aligned} \frac{\partial P_4}{\partial P} = & P \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 \|x_{k,c} - \mu_c\|_H^2 - \beta P X M_{cc} X^T + \varpi P \\ & + P W W^T \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 x_k x_k^T - \frac{1}{2} W \sum_{c=0}^C \sum_{k=1}^N \lambda_{k,c}^2 x_k x_k^T y_k = 0 \end{aligned} \tag{27}$$

The solution for P can be obtained as Equation 26, thus Theorem 5 is proved.

4 Algorithm

4.1 Algorithm description

In the context of unsupervised domain adaptation where the target domain lacks labeled data, achieving semantic alignment between domains relies on initial label information from the target domain. The initial label information of the target domain samples can be obtained through the following three strategies (Liang et al., 2019): (1) by a random method; (2) by setting all labels to zero; (3) by performing clustering on the target domain data using a model trained from the source domain data. Strategies (1) and (2) are cold start methods. Strategy (3) is a warm start method, which is usually more beneficial for the subsequent learning process. Therefore, we choose the third strategy to initialize the prior information of the target domain, thereby initializing $\lambda_{k,c}$, W , P , and Y . EDPC utilizes an iterative optimization strategy, which is a common approach in multi-objective optimization. The iterative process of the algorithm will stop when the following conditions are met: $\left| \Theta(\lambda_{k,c}^z, P^z, W^z, Y^z) - \Theta(\lambda_{k,c}^{z-1}, P^{z-1}, W^{z-1}, Y^{z-1}) \right| < \varepsilon$, where $\Theta(\lambda_{k,c}^z, P^z, W^z, Y^z)$ represents the value of the objective function at the z -th iteration, and ε is a predefined threshold. The complete learning process of the proposed method is given in Algorithm 1.

4.2 Generalization Analysis

Rademacher complexity is an effective measure of a function set's capacity to fit noise (Ghifary et al., 2017; Tao and Dan, 2021).

ALGORITHM 1 : Domain adaptation learning based on EDPC

Input: The source domain data $\{\hat{X}_s, Y_s\}$, the target domain data \hat{X}_t , unknown target domain labels Y_t (initialized via clustering), and parameters $\rho, \alpha, \vartheta, \beta, \varpi$, the threshold for iteration termination ε , and the maximum number of iterations Z .

Output: The contribution matrix λ , which represents the matching contributions of each instance at the mean points of various categories in the overall domain; the shared subspace P ; the decision function W on the source and target domain datasets; and the label matrix Y .

Procedure:

- 1 Initialize the label values for the unlabeled data in the target domain.
- 2 Calculate the mean values $\mu_{t,c}$ and $\mu_{s,c}$ for different categories in the target domain and source domain, respectively, where $c = 0, 1, 2, \dots, C$.
- 3 Then calculate the mean values μ_c for different categories in the overall domain as $\mu_c = \frac{1}{2}(\mu_{s,c} + \mu_{t,c})$.
- 4 Obtain the initial values $\lambda_{k,c}^0$ for $\lambda_{k,c}$ using Equation 19.
- 5 Obtain the initial values W^0 for W using Equation 22.
- 6 Obtain the initial values Y^0 for Y using Equation 24.
- 7 Obtain the initial values P^0 for P using Equation 26.
- 8 Compute the value of the objective function $\Theta(\lambda_{k,c}^0, P^0, W^0, Y^0)$.
- 9 Repeat the following steps sequentially from $z = 0$ to Z :
 - {
 - 9.1 Update the values of $\lambda_{k,c}$ to $\lambda_{k,c}^z$ using Equation 19.
 - 9.2 Obtain the updated values W^z for W using Equation 22.
 - 9.3 Obtain the updated values Y^z for Y using Equation 24.
 - 9.4 Obtain the updated values P^z for P using Equation 26.
 - 9.5 Update the value of the objective function to $\Theta(\lambda_{k,c}^z, P^z, W^z, Y^z)$.
 - 9.6 If $\left| \Theta(\lambda_{k,c}^z, P^z, W^z, Y^z) - \Theta(\lambda_{k,c}^{z-1}, P^{z-1}, W^{z-1}, Y^{z-1}) \right| < \varepsilon$, terminate the repetition and return the matrices λ, P, W , and Y ; otherwise, go back to step 9.1 and continue the calculation until the condition in step 9.5 is satisfied.
 - }

Therefore, we will derive the generalization error bound for the proposed method using Rademacher complexity. Let $H := \{\mathcal{X} \rightarrow \mathcal{Y}\}$ be a set of hypothesis functions in the RKHS H space, where \mathcal{X} is a compact set and \mathcal{Y} is a label space. Given a loss function $loss(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ and a neighborhood distribution \mathcal{D} on \mathcal{X} , the expected loss between two hypothesis functions $h, \tilde{h} \in H$ is defined as:

$$\mathcal{L}_{\mathcal{D}}(h, \tilde{h}) = \mathbb{E}_{x \sim \mathcal{D}} [\text{loss}(h(x), \tilde{h}(x))]$$

The difference in domain distributions between the source domain distribution \mathbb{P} and the target domain distribution \mathbb{Q} can be defined in Equation (29):

$$\text{disc}(\mathbb{P}, \mathbb{Q}) = \sup_{h, \tilde{h} \in \mathcal{H}} \left\{ \mathcal{L}_{\mathbb{P}}(h, \tilde{h}) - \mathcal{L}_{\mathbb{Q}}(h, \tilde{h}) \right\} \quad (29)$$

Let $f_{\mathbb{P}}$ and $f_{\mathbb{Q}}$ be the true label functions for \mathbb{P} and \mathbb{Q} , respectively, and let the corresponding optimized hypothesis functions be:

$$h_{\mathbb{P}}^* := \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{L}_{\mathbb{P}}(h, f_{\mathbb{P}})$$

$$h_{\mathbb{Q}}^* := \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{L}_{\mathbb{Q}}(h, f_{\mathbb{Q}})$$

Their corresponding expected loss is denoted as $\mathcal{L}_{\mathbb{P}}(h_{\mathbb{P}}^*, h_{\mathbb{P}}^*)$. Our EDPC method aims to achieve the empirical loss target $\mathcal{L}_{\mathbb{P}}(h_{\mathbb{P}}^*, h_{\mathbb{P}}^*)$ through the objective function $R(Y, W)$.

The following theorem provides the generalization error bound for the proposed method:

Theorem 6 (Generalization Error Bound; Nie et al., 2010). Let $H := \{f \in \mathcal{H} : \mathcal{X} \rightarrow \mathbb{R}, \|f\|_{\mathcal{H}} \leq 1 \text{ and } \|f\|_{\infty} \leq r\}$ be a function set in the reproducing kernel Hilbert space (RKHS) \mathcal{H} . Consider $X_{\mathcal{X}}^{\mathbb{P}} = (x_1^s, \dots, x_{n_s}^s) \sim \mathbb{P}$ and $X_{\mathcal{X}}^{\mathbb{Q}} = (x_1^t, \dots, x_{n_t}^t) \sim \mathbb{Q}$ as datasets from the source domain and the target domain, respectively. Assume the loss function $\text{loss}(\cdot)$ is q -Lipschitz, mapping $\text{loss}(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \rightarrow [0, q]$. For $a, b \in \mathcal{Y} \times \mathcal{Y}$, the condition $|\text{loss}(a) - \text{loss}(b)| = q|a - b|$ holds. The generalization error bound for any hypothesis function $h \in \mathcal{H}$, with a probability of at least $1 - \delta$, having Rademacher complexity $\mathfrak{R}_{X_{\mathcal{X}}^{\mathbb{P}}}(H)$ on $X_{\mathcal{X}}^{\mathbb{P}}$, is shown in Equation (30):

$$\begin{aligned} \mathcal{L}_{\mathbb{Q}}(h, f_{\mathbb{Q}}) - \mathcal{L}_{\mathbb{Q}}(h_{\mathbb{Q}}^*, f_{\mathbb{Q}}) &\leq \mathcal{L}_{\mathbb{P}}(h, h_{\mathbb{P}}^*) + 2q\mathfrak{R}_{X_{\mathcal{X}}^{\mathbb{P}}}(H) \\ &+ 3q\sqrt{\frac{\log \frac{2}{\delta}}{2N}} + 8q\sqrt{\Omega(\mathbb{P}(X_s), \mathbb{P}(X_t))} + R(W, Y) \end{aligned} \quad (30)$$

where $\mathfrak{R}_{X_{\mathcal{X}}^{\mathbb{P}}}(H)$ is the Rademacher complexity, $\Omega(\mathbb{P}(X^s), \mathbb{P}(X^t))$ is the discriminative possibilistic clustering distribution distance measure and is composed by possibilistic clustering distribution distance measure with intra-class alignment item $\Omega_{\lambda}(\mathbb{P}(X^s), \mathbb{P}(X^t))$

and discriminative item $\Omega_{rep}(\mathbb{P}(X^s), \mathbb{P}(X^t))$ (i.e., $\Omega(\mathbb{P}(X^s), \mathbb{P}(X^t)) = \alpha\Omega_{\lambda}(\mathbb{P}(X^s), \mathbb{P}(X^t)) - \beta\Omega_{rep}(\mathbb{P}(X^s), \mathbb{P}(X^t))$).

Theorem 6 demonstrates that the discriminative possibilistic clustering distribution distance measure $\Omega(\mathbb{P}(X^s), \mathbb{P}(X^t))$ and the model alignment function $R(W, Y)$ can jointly control the generalization error bound of the proposed method. Consequently, by minimizing both the discriminative possibilistic distribution distance

between domains and the model bias, the proposed method can effectively enhance its generalization performance in domain adaptation. Experimental results on real-world datasets also support this conclusion.

5 Experiments

To validate the effectiveness of the proposed EDPC in the cross-domain emotion recognition, this section systematically compares and analyzes the performance of the EDPC method with current state-of-the-art unsupervised domain adaptation techniques on several key EEG datasets (i.e., SEED and SEED-IV).

5.1 Databases description

To ensure a fair comparison with state-of-the-art (SOTA) methods, extensive experiments were conducted for effective validation using two well-known open datasets: SEED (Zheng and Lu, 2015a) and SEED-IV (Zheng et al., 2018). The SEED datasets comprises data collected from 15 subjects, each participating in three sessions held at different times. Each session includes 15 trials, featuring three types of emotional stimuli: negative, neutral, and positive. Similarly, the SEED-IV datasets also involves 15 subjects, each undergoing three sessions at different times. Each session in SEED-IV consists of 24 trials, with four emotional stimuli: happy, sad, fearful, and peaceful.

EEG signals for both datasets (SEED and SEED-IV) were simultaneously recorded using a 62-channel ESI Neuroscan system. During EEG signal preprocessing, the data were down-sampled to a rate of 200 Hz. Environmental noise was manually removed, and the data were filtered using a 0.3 Hz-50 Hz band-pass filter. To make a fair comparison with the existing studies on the two benchmark databases, we also use the pre-computed differential entropy (DE) features as the model input. Specifically, for each trial, the EEG data was divided into a number of 1-s segments, and the DE features were extracted from each 1-s segment at the given five frequency bands [Delta (1–3 Hz), Theta (4–7 Hz), Alpha (8–13 Hz), Beta (14–30 Hz), and Gamma (31–50 Hz)] from the 62 channels. Then, for each 1-s segment, it was represented by a 310-dimensional feature vector (5 frequency bands \times 62 channels), which was further filtered by a linear dynamic system method for smooth purpose (Shi and Lu, 2010).

5.2 Experiment Settings

Before delving into the detailed analysis of the experimental results, it is essential to fine-tune the hyperparameters in the EDPC strategy. Empirical evidence suggests that the hyperparameters θ and α serve to balance the trade-off between fuzziness and local structure preservation in the objective Function (17) for both the source and target domains. Meanwhile, the other two hyperparameters β and ρ can still be adjusted to, respectively, balance the influence of class exclusion and feature selection. Therefore, these two parameters play a crucial role in the final performance of the algorithm.

Given that parameter setting remains a challenging issue in the field of machine learning, this study adopts an experience-based

parameter space exploration strategy to determine the optimal parameter configuration. This, in turn, allows for the evaluation of various methods on the datasets and the recording of the best performance for each method. Except in special cases, all related methods will undergo fine-tuning to achieve optimal results. Regarding the potential geometric structure of the data, it is associated with the neighborhood size chosen when constructing the Laplacian matrix. Experimental observations show that the model's performance is slightly sensitive to changes in neighborhood size when the neighborhood is small. Therefore, when constructing the nearest neighbor graph in EDPC, this study uses grid search within the range $\{3,5,10,15,17\}$ to determine the optimal number of neighbors and reports the highest recognition accuracy results obtained under this optimal parameter configuration.

Additionally, for all methods implemented in this paper, a Gaussian kernel $K(x, x_i) = \exp(-x - x_i^2 / 2\sigma^2)$ is used in both the source and target domains, where σ is determined by minimizing the Maximum Mean Discrepancy (MMD) to establish a benchmark. Based on prior experience, σ is initially set to the square root of the average norm of the binary training data, and for multiclass classification, σ is adjusted to $\sigma\sqrt{C}$ (where C represents the number of classes). The underlying geometric structure relies on k neighbors for computing the Laplacian matrix. In our experiments, we observed that performance slightly varies when k is not large. Consequently, to construct the nearest neighbor graph in EDPC, we conduct a grid search to determine the optimal number of nearest k neighbors from the set $\{3,5,10,15,17\}$, and report the best recognition accuracy from the optimal parameter configuration.

Before presenting the detailed evaluation, it is essential to explain the tuning process for the hyper-parameters of EDPC. Based on prior experience, the parameter β is used to balance the fuzzy entropy and the alignment of domain probability distributions in the objective Function (16). The parameters α and ρ are adjustable and used to balance the importance of structure description and feature selection. Given that parameter uncertainty remains an open issue in machine learning, we rely on previous work experience to determine these parameters. Consequently, we evaluate all methods on the datasets by empirically searching the parameter space to identify the optimal settings and report the best results for each method. Except for special cases, all parameters of relevant methods are fine-tuned to achieve optimal results.

Since unsupervised domain adaptation lacks target labels for standard cross-validation, we employ a leave-one-subject-out strategy on the SEED and SEED-IV datasets (detailed in Section 6.2). We identify the optimal parameter values from the set $\{10^{-6}, 10^{-5}, \dots, 10^5, 10^6\}$ by achieving the highest average accuracy on these datasets using the aforementioned method. This strategy typically constructs an effective EDPC model for unsupervised domain adaptation, and a similar approach is used to find optimal parameter values for other domain adaptation methods. In the subsequent sub-sections, a series of experiments is designed to test the sensitivity of the proposed EDPC method to parameter selection (see Section 6.4.1), verifying that EDPC can maintain stable performance across a wide range of parameter values. Additionally, the hyper-parameters for other methods are selected according to their original literature.

5.3 Experimental protocols

To fully verify the robustness and stability of the proposed method, this paper employs three different validation protocols (leave-one-subject-out; Zhang et al., 2020) to compare the proposed method with the latest methods.

- 1 Single-subject cross-session leave-one-session-out cross-validation. In line with existing methods, a time series cross-validation approach is utilized here, where past data is leveraged to predict current or future data. For each subject, the first two sessions are designated as the source domain, while the latter session serves as the target domain. The final results are determined by calculating the average accuracy and standard deviation across all subjects.
- 2 Cross-subject single-session leave-one-subject-out cross-validation. This validation scheme is the most widely used in emotion recognition tasks based on EEG data (Li J. et al., 2020; Luo et al., 2018). In this approach, one session's data from a subject is treated as the target domain, while the data from the remaining subjects serve as the source domain. The training and validation process is repeated until each subject has been used as the target once. Consistent with other studies, we only consider the first session for this type of cross-validation.
- 3 Cross-subject cross-session leave-one-subject-out cross-validation. To comprehensively assess the robustness of the model on unseen subjects and trials, this paper employs a rigorous leave-one-out cross-subject cross-session method for evaluation. In this approach, all session data from a single subject are designated as the target domain, while data from all sessions of the remaining subjects serve as the source domain. The training and validation process is repeated until each subject's sessions have been used as the target domain once. Given the variations between subjects and sessions, this evaluation protocol presents a substantial challenge to the effectiveness of models in EEG-based emotion recognition tasks.

5.4 Experimental results

Specifically, in the following tables of experimental results, the bold values in each table are the best accuracy performance results achieved by the compared methods. Pacc denotes the average accuracy performance of each method. In following tables, EDPC denotes the method proposed by us. During the implementation of the experiments, the features of the data were initially extracted using shallow technical means. When labeled as EDPC+ResNet101, it indicates that we employed the deep neural network ResNet101 for data feature extraction in the experimental process.

5.4.1 Single-subject cross-session

We calculate the average and standard deviation of each subject's experimental results, the cross-session validation results for each subject on different datasets (i.e., SEED and SEED-IV) show in Tables 1, 2, respectively. When the proposed EDPC method is compared with the traditional machine learning methods on both

SEED and SEED-IV, the EDPC obtained the best accuracy performance, even DICE. It indicates that the EDPC method with discrimination has a more significant noise suppression effect.

Additionally, in the experiments on the SEED datasets, the results from deep learning methods show that the EEGMatch method achieved the best performance among the deep learning methods. This is possibly because the mixup technique provided richer and more effective data information, which aided the model training, although the increased data volume naturally led to higher computational costs. Nevertheless, the EDPC method still obtained comparable performance and ranked closely its behind. This demonstrates the advantage of the EDPC method in distinguishing the single subject across sessions.

We can observe from the Table 3 that the EDPC method achieved the best performance on SEED-IV datasets with four different emotions (SEED has three different emotions), no matter what in the traditional machine learning or the deep learning methods. It signs that the EDPC method with “sames attract and opposites repel” characteristics can get more accuracy performance on finer-grained emotion recognition. The EDPC has stronger scalability in more nuanced emotion recognition tasks.

5.4.2 Cross-subject single-session

Table 3 presents the model results for the recognition task using the leave-one-subject-out method within a single session, comparing them with the performance of the latest methods in the literature. All results are presented as mean ± standard deviation. The MDDD method achieved the best performance, the possible reason is: a more balanced impact of noise may arise when the different subjects in the same session faced consistent consistent environment, the MDDD adopted ensemble learning approach can effectively handle this kind of data. The MDDD method needs higher computational costs since it requires training and combining multiple models.

However, the EDPC achieved the best accuracy (82.34) with a standard deviation of 7.52 except MDDD method. It still maintains a

comparable performance advantage overall among traditional machine learning methods. The recognition performance of EDPC surpasses that of the DICE method, indicating that the EDPC method handles noise better than the DICE method. It shows that the clustering hypothesis with fuzzy entropy can overcome the influence of noise and outliers in unsupervised classification.

When we compared to the latest deep learning methods, particularly deep transfer learning networks based on DANN (e.g., ATDD-DANN, R2GSTNN, BiHDM, BiDANN, WGAN-GP, PR-PL, EEGMatch), the proposed EDPC method demonstrated the best performance from Table 3. It effectively addresses issues of individual differences and noisy labels in aBCI applications, indicating that the EDPC method has better generalization performance and discrimination across subjects within the same session.

5.4.3 Cross-subject cross-session

To validate the efficiency and stability of the EDPC method under cross-subject and cross-session conditions, this study uses cross-subject cross-session leave-one-out cross-validation on the SEED and SEED-IV databases to verify the proposed EDPC method. As shown in Tables 4, 5, the proposed EDPC method achieves the best performance when compared to traditional machine learning methods and deep learning methods. Moreover, Table 3 obtained better performance than Table 5 since the cross-subject cross-session is more complicated. We easy to see that the performance of EDPC method is better than the MDDD method. All these results demonstrate that the proposed EDPC method has higher recognition accuracy and better generalization ability in the face of more complex individual and environmental differences, indicating better emotional validity.

6 Discussion

To comprehensively study the model’s performance, this section evaluates the effects of different settings in EDPC. Please note that all

TABLE 1 The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using single-subject cross-session leave-one-subject-out cross-validation.

Methods	Pacc	Methods	Pacc
Traditional machine learning methods			
RF (Breiman, 2001)	76.42 ± 11.15	KNN* (Coomans and Massart, 1982)	72.96 ± 12.10
TCA* (Pan et al., 2011)	77.63 ± 11.49	CORAL (Sun et al., 2016)	84.18 ± 9.81
SA* (Li Y. et al., 2020)	67.79 ± 7.43	GFK* (Gong et al., 2012)	79.28 ± 7.44
DICE (Liang et al., 2019)	81.58 ± 7.55	EDPC	82.31 ± 6.44
MDDD (Luo et al., 2024)	81.27 ± 5.47		
Deep learning methods			
DAN (Long et al., 2015)	89.16 ± 7.90	SimNet (Pinheiro, 2018)	86.88 ± 7.83
DDC (Tzeng et al., 2014)	91.14 ± 5.61	ADA (Li J. et al., 2020)	89.13 ± 7.13
DANN (Ganin et al., 2016)	89.45 ± 6.74	MMD (Li J. et al., 2020)	84.38 ± 12.05
JDA-Net (Li J. et al., 2020)	91.17 ± 8.11	DCORAL (Sun et al., 2016)	88.67 ± 6.25
PR-PL (Zhou et al., 2022)	93.18 ± 6.55	PARSE (Zhang and Etemad, 2022)	89.85 ± 5.06
EEGMatch (Zhou et al., 2023)	94.70 ± 4.10	EDPC+ResNet101	93.76 ± 5.82

Here, the model results reproduced by us are indicated by ‘*’. The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

TABLE 2 The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED-IV database using single-subject cross-session leave-one-subject-out cross-validation.

Methods	Pacc	Methods	Pacc
Traditional machine learning methods			
RF (Breiman, 2001)	60.27 ± 16.36	KNN (Coomans and Massart, 1982)	54.18 ± 16.28
TCA* (Pan et al., 2011)	59.49 ± 12.07	CORAL* (Sun et al., 2016)	66.88 ± 14.67
SA* (Li Y. et al., 2020)	56.94 ± 11.45	GFK* (Gong et al., 2012)	60.66 ± 10.00
DICE (Liang et al., 2019)	69.68 ± 12.52	EDPC	71.39 ± 7.22
MDDD (Luo et al., 2024)	68.81 ± 9.25		
Deep learning methods			
DCORAL (Sun et al., 2016)	65.10 ± 13.20	DAN (Long et al., 2015)	60.20 ± 10.20
DDC (Tzeng et al., 2014)	68.80 ± 16.60	MEERNet (Chen et al., 2021)	72.10 ± 14.10
PR-PL (Zhou et al., 2022)	74.62 ± 14.15	PARSE (Zhang and Etemad, 2022)	70.24 ± 8.47
EEGMatch (Zhou et al., 2023)	72.91 ± 8.34	EDPC+ResNet101	76.58 ± 10.29

Here, the model results reproduced by us are indicated by '*'. The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

TABLE 3 The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using cross-subject single-session leave-one-subject-out cross-validation.

Methods	Pacc	Methods	Pacc
Traditional machine learning methods			
TKL (Li et al., 2018c)	63.54 ± 15.47	T-SVM* (Li et al., 2018c)	68.57 ± 9.54
TCA (Pan et al., 2011)	63.64 ± 14.88	TPT* (Suykens and Vandewalle, 1999)	73.86 ± 11.05
KPCA (Suykens and Vandewalle, 1999)	61.28 ± 14.62	GFK (Gong et al., 2012)	71.31 ± 14.09
SA* (Li Y. et al., 2020)	66.00 ± 10.89	DICA (Ma et al., 2019)	69.40 ± 07.80
DNN (Suykens and Vandewalle, 1999)	61.01 ± 12.38	SVM (Suykens and Vandewalle, 1999)	58.18 ± 13.85
DICE (Liang et al., 2019)	74.22 ± 7.33	EDPC	82.34 ± 7.52
MDDD (Luo et al., 2024)	84.57 ± 9.49		
Deep learning methods			
DGCNN (Song et al., 2018)	79.95 ± 9.02	DAN (Long et al., 2015)	83.81 ± 8.56
RGNN (Zhong et al., 2020)	85.30 ± 6.72	BiHDM (Li Y. et al., 2020)	85.40 ± 7.53
WGAN-GP (Luo et al., 2018)	87.10 ± 7.10	MMD (Li J. et al., 2020)	80.88 ± 10.10
ATDD-DANN (Du et al., 2020)	90.92 ± 1.05	JDA-Net (Li J. et al., 2020)	88.28 ± 11.44
R2G-STNN (Li et al., 2019)	84.16 ± 7.63	SimNet* (Pinheiro, 2018)	81.58 ± 5.11
BiDANN (Li et al., 2018c)	83.28 ± 9.60	DResNet (Ma et al., 2019)	85.30 ± 8.00
ADA (Li J. et al., 2020)	84.47 ± 10.65	DANN (Ganin et al., 2016)	81.65 ± 9.92
PR-PL (Zhou et al., 2022)	93.06 ± 5.12	PARSE (Zhang and Etemad, 2022)	82.11 ± 5.83
EEGMatch (Zhou et al., 2023)	92.45 ± 06.85	EDPC+ResNet101	94.79 ± 4.28

Here, the model results reproduced by us are indicated by '*'. The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

results presented in this section are based on the SEED datasets, using the cross-subject single-session cross-validation evaluation protocol. The bold values in Table 6 is the best accuracy performance result achieved of this method. It's best to review in color mode.

6.1 Ablation study

This ablation study systematically explores the effectiveness of different components in the proposed model and presents the

corresponding contributions of each component to the overall performance of the model. As shown in Table 6, adding 5 labeled data points per category in the target domain achieves a recognition accuracy (95.98 ± 5.22) very close to the recognition accuracy of EDPC (unsupervised learning; 92.19 ± 4.70). This decline indicates that prior label information in the target domain significantly enhances model performance and highlights the great potential of transfer learning in aBCI applications. Moreover, simultaneously preserving the local structure of data in both the source and target domains helps to improve model performance; otherwise, the

TABLE 4 The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using cross-subject cross-session leave-one-subject-out cross-validation.

Methods	Pacc	Methods	Pacc
Traditional machine learning methods			
RF (Breiman, 2001)	69.60 ± 7.64	KNN (Coomans and Massart, 1982)	60.66 ± 7.93
SVM* (Suykens and Vandewalle, 1999)	62.24 ± 5.48	Adaboost (Zhu et al., 2006)	71.87 ± 5.70
TCA* (Pan et al., 2011)	65.31 ± 6.04	CORAL (Sun et al., 2016)	69.22 ± 4.11
SA (Li Y. et al., 2020)	61.41 ± 9.75	GFK* (Gong et al., 2012)	67.36 ± 6.52
DICE* (Liang et al., 2019)	73.56 ± 4.23	EDPC	76.82 ± 7.14
MDDD (Luo et al., 2024)	76.60 ± 6.79		
Deep learning methods			
DCORAL* (Sun et al., 2016)	80.87 ± 6.04	DAN* (Long et al., 2015)	82.51 ± 3.71
DDC (Tzeng et al., 2014)	82.17 ± 4.96	DANN* (Ganin et al., 2016)	84.79 ± 6.44
PR-PL (Zhou et al., 2022)	85.56 ± 4.78	PARSE (Zhang and Etemad, 2022)	82.44 ± 5.00
EEGMatch (Zhou et al., 2023)	86.30 ± 5.04	EDPC+ResNet101	87.42 ± 6.15

The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

TABLE 5 The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED-IV database using cross-subject cross-session leave-one-subject-out cross-validation.

Methods	Pacc	Methods	Pacc
Traditional machine learning methods			
RF (Breiman, 2001)	50.98 ± 9.20	KNN (Coomans and Massart, 1982)	40.83 ± 7.28
SVM (Suykens and Vandewalle, 1999)	51.78 ± 12.85	Adaboost (Zhu et al., 2006)	53.44 ± 9.12
TCA (Pan et al., 2011)	56.56 ± 13.77	CORAL (Sun et al., 2016)	49.44 ± 9.09
SA (Li Y. et al., 2020)	64.44 ± 9.46	GFK (Gong et al., 2012)	45.89 ± 8.27
KPCA (Suykens and Vandewalle, 1999)	51.76 ± 12.89	DNN (Suykens and Vandewalle, 1999)	49.35 ± 9.74
DICE (Liang et al., 2019)	66.75 ± 7.25	EDPC	67.88 ± 5.21
MDDD (Luo et al., 2024)	64.90 ± 10.25		
Deep learning methods			
DGCNN (Song et al., 2018)	52.82 ± 9.23	DAN (Long et al., 2015)	58.87 ± 8.13
RGNN (Zhong et al., 2020)	73.84 ± 8.02	BiHDM (Li Y. et al., 2020)	69.03 ± 8.66
BiDANN (Li et al., 2018c)	65.59 ± 10.39	DANN (Ganin et al., 2016)	54.63 ± 8.03
PR-PL (Zhou et al., 2022)	74.92 ± 7.92	PARSE (Zhang and Etemad, 2022)	69.78 ± 8.22
EEGMatch (Zhou et al., 2023)	73.60 ± 7.53	EDPC+ResNet101	76.11 ± 6.69

The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

recognition accuracy significantly decreases (92.29 ± 5.67 and 92.83 ± 4.81 , respectively). When the l21 norm in W is replaced with the l2 norm, the model's recognition accuracy drops to 92.13 ± 5.83 . This result demonstrates that using the l21 constraint achieves better sample selection and denoising effects. For the pseudo-labeling method, when switching from a fixed mode to a linear dynamic update, the corresponding accuracy increases from 91.48 to 94.71. When using an adaptive pseudo-labeling method based on nonlinear dynamics, the accuracy further improves to 94.79. When employing multi-kernel learning, the accuracy further increases to 95.66. These results indicate that multi-kernel learning helps rationalize the importance of different kernels in various scenarios and enhances the model's generalization ability.

Next, we analyze the impact of different hyperparameters on the overall performance of the model. According to the experimental

results, it can be observed that when the dynamic learning rate ρ varies from 0 to 100, the model accuracy continuously improves from 92.79 to 93.54. This indicates that a dynamic learning rate ρ is superior to a fixed value in terms of recognition accuracy. Additionally, the results suggest that the value of β directly affects the importance of the discriminative term in the model. When the discriminative term is removed, the model accuracy drops to 90.56, whereas when $\beta = 100$, the model accuracy reaches 94.74, which is close to the performance of EDPC. This demonstrates that the discriminative term plays an indispensable role in the model.

Additionally, θ and θ are two balancing parameters used to optimize the performance of EDPC by weighting fuzzy entropy and local retention (referring to the source and target domains). As shown in Figure 1a, when θ is 0, the performance is poor. Performance jumps significantly when θ is 1, and it continues to improve steadily as θ

TABLE 6 The ablation study of our proposed model.

	Pacc
Ablation setting	
target prior information (5 labeled samples per category)	95.98 ± 5.22
only preserving the local structures on the source	92.29 ± 5.67
only preserving the local structures on the target	92.83 ± 4.81
imposing l_2 -norm on W	92.13 ± 5.83
fixed pseudo-labeling	91.48 ± 5.33
dynamic pseudo-labeling	94.71 ± 4.15
multiple kernel leaning	95.66 ± 3.20
Hyper-parameter controlling strategy	
$\rho = 0$ (Ignore the regularization term of W)	92.79 ± 4.40
fixed $\rho = 100$ for W regularization	93.54 ± 5.16
$\beta = 0$ (Ignore the regularization term of discriminative)	90.56 ± 3.25
fixed $\beta = 100$ for the regularization term of discriminative	94.74 ± 4.53
The proposed model	
EDPC + ResNet101	94.79 ± 4.28

The bold value in table corresponds to the method with the best accuracy performance result among all the compared methods.

increases. When θ reaches 400 or higher, the performance stabilizes. Therefore, based on the experimental results, setting θ around 400 can achieve the best EDPC performance.

From Figure 1b, it can be seen that when α is 0, the performance is poor. When α is 1, there is a performance jump. As α increases further, the performance improves slowly and becomes jittery. According to the trend observed in the experimental curve, α can be adjusted within the range of [100, 225] to achieve optimal EDPC performance.

Figure 1c shows that when δ is within the range of [0.5, 0.85], the performance is high, indicating that the source domain contains a significant amount of discriminative information, and the means of different categories play a more important role. When δ takes larger values up to 1 (completely removing the target domain's mean information), the performance declines, further demonstrating that the mean information of the target domain also plays an auxiliary role.

6.2 Effect of noisy labels

To further validate the robustness of the model in the presence of noisy labels, we conducted an experiment where noise was randomly added to the source labels at different proportions, and the corresponding

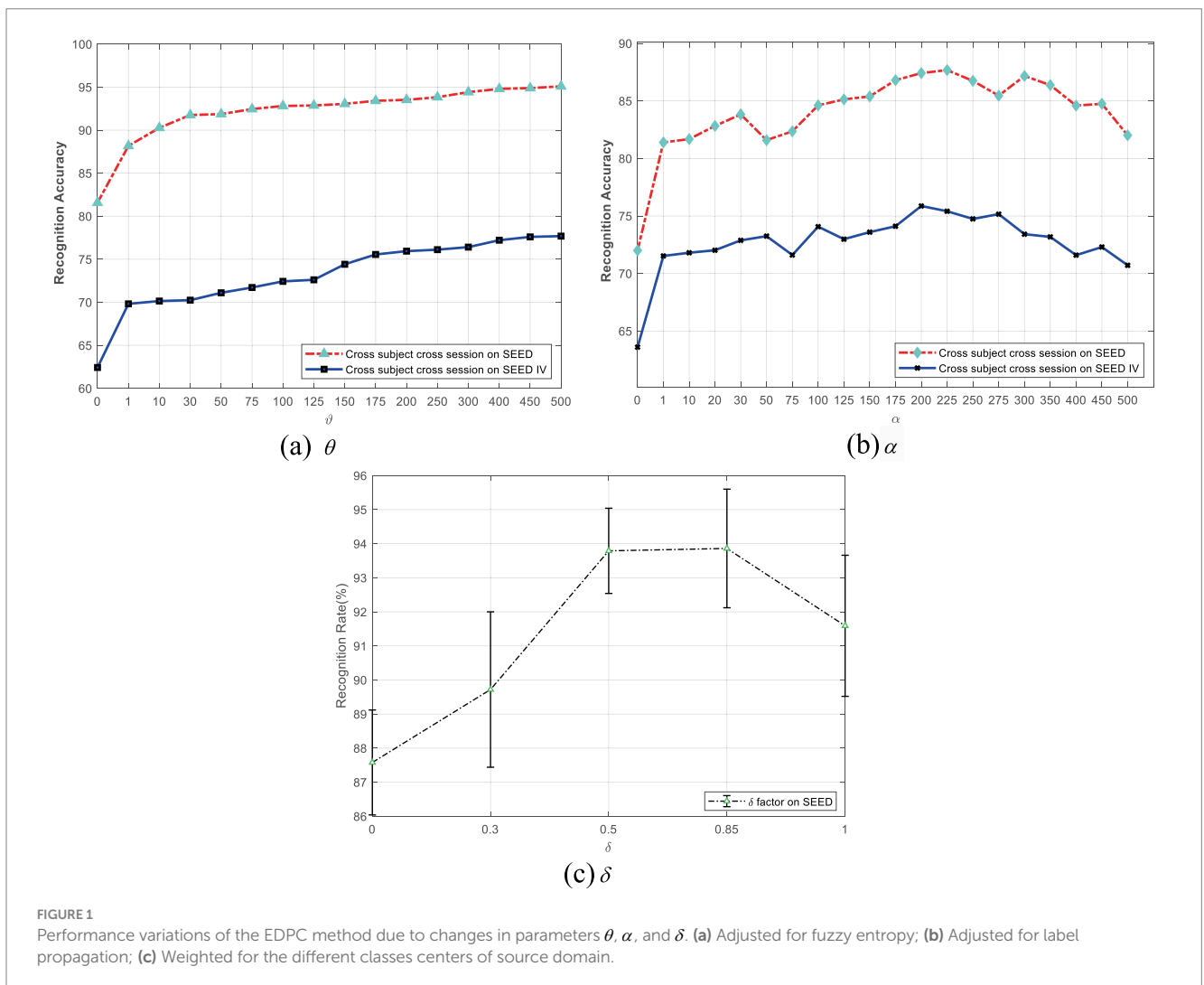


FIGURE 1 Performance variations of the EDPC method due to changes in parameters θ , α , and δ . (a) Adjusted for fuzzy entropy; (b) Adjusted for label propagation; (c) Weighted for the different classes centers of source domain.

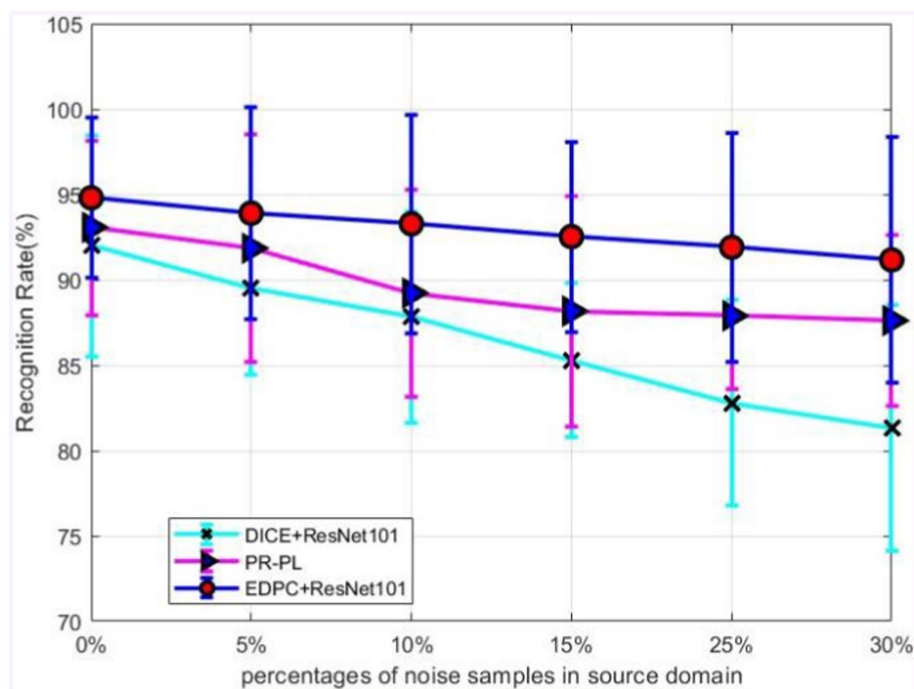


FIGURE 2
Robustness on source domain with different noise levels.

model performance on unseen target data was tested. Specifically, we replaced a certain proportion of the true labels with randomly generated labels and trained the model using semi-supervised learning. Then, the trained model's performance was evaluated on the target domain. It is important to note that noise was only added to the source domain data, while the target domain was used for model evaluation. In the implementation, the noise proportions were adjusted to 5, 15, 25, and 30% of the source domain data.

The experimental results shown in Figure 2 indicate that as the amount of noise increases, the accuracy of the proposed EDPC decreases at the slowest rate, demonstrating that EDPC is a reliable model with a high tolerance for noisy data. In future work, recent methods such as [52, 53] could be integrated to further reduce the more prevalent noise in EEG signals and enhance the model's stability in cross-corpus applications.

6.3 Confusion matrices

To qualitatively study the performance of the model in each emotion category, this section visualizes the confusion matrix and compares the results with state-of-the-art models (i.e., BiDANN, BiHDM, RGNN, PR-PL, DICE ResNet101). As shown in the Figure 3, all models excel at distinguishing positive emotions from other emotions (with recognition rates above 90%), but they perform relatively poorly in distinguishing negative emotions from neutral emotions. For example, the emotion recognition rate in RGNN [25] is even below 80% (specifically, 79.14%).

Additionally, the PR-PL method slightly outperforms the EDPC method in recognizing positive emotions, likely due to the use of adversarial networks in PR-PL, which also increases its

computational cost. Compared to existing methods (Figures 3a–c,e), the EDPC method is optimal, particularly in distinguishing neutral and negative emotions (even surpassing PR-PL). Furthermore, the overall performance of this method is superior to the DICE method (as shown in the comparison between Figures 3e,f).

6.4 Convergence

The proposed algorithm in this paper employs an iterative optimization strategy. To demonstrate the convergence of the algorithm, experiments were conducted on the MATLAB platform. The hardware configuration used for implementation includes 64GB of memory, a 2.5GHz CPU, and an 8-core Intel i7-11850H processor. The Figure 4 shows the convergence process of EDPC at different iteration counts. From the displayed results, it is evident that the proposed algorithm approaches convergence at around 30 iterations. In the algorithm, the objective function of each sub-problem optimization is a decreasing function, thereby proving that the EDPC method has good convergence properties.

7 Conclusion

The article analyzes the issue of mean shift that may arise during distribution distance measure in domain adaptation due to potential noise in the field. It proposes a novel distribution distance measure method with DPC criterion. This method determines the membership degree of each instance (i.e., the closer the distance, the more likely it is non-noise data) by

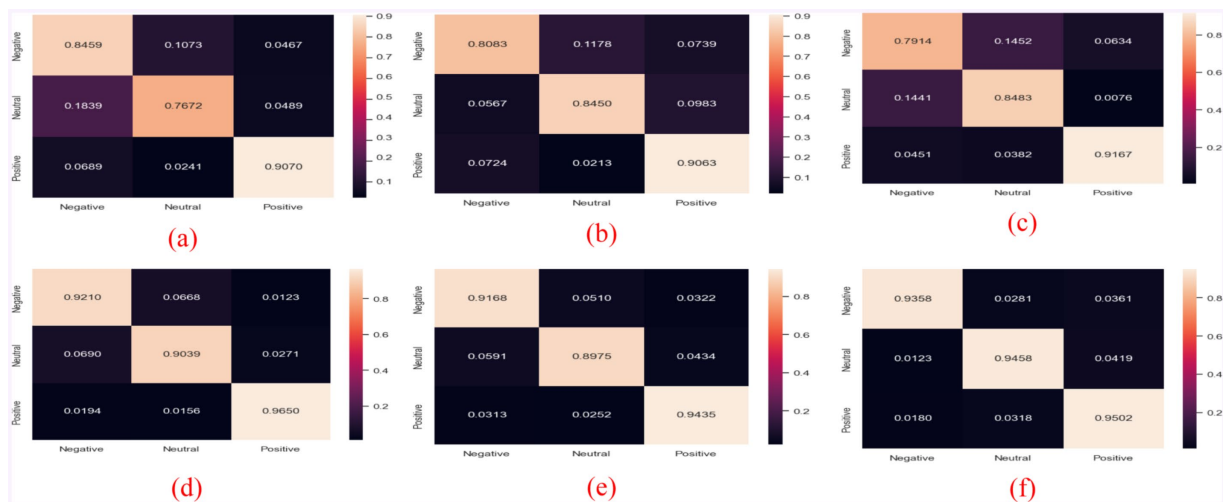


FIGURE 3 Confusion matrices of different models: (a) BiDANN; (b) BiHDM; (c) RGNN; (d) PR-PL; (e) DICE+ResNet101; and (f) EDCP+ResNet101.

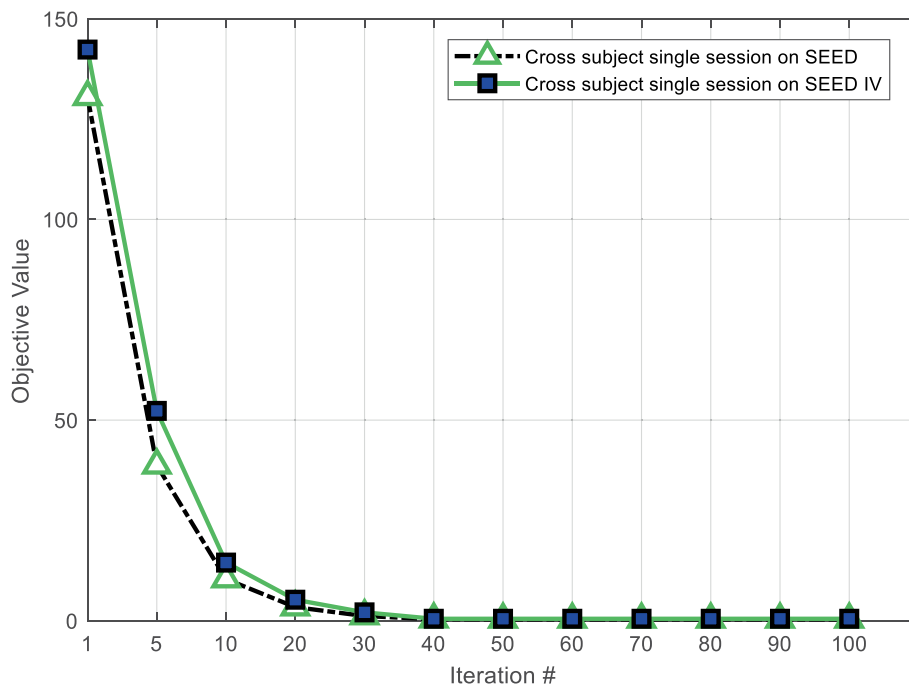


FIGURE 4 Convergence.

measuring the distance between each instance and the overall domain mean. It uses membership degree and fuzzy trimming to mitigate the impact of noise data on final performance. Additionally, an outlier repulsion term is added during the domain adaptation process to further improve classifier discrimination accuracy. Based on this, a Emotion recognition DA method based on DPC is proposed (namely EDCP). It minimizes domain distribution differences while introducing a graph Laplacian

matrix to learn a target domain label matrix and minimizes domain discrimination differences to learn a classifier for both the source and target domains. Theoretically, it is proven that the proposed algorithm has consistent convergence and an effective generalization error bound. Finally, the proposed EDCP method is extensively compared with the latest shallow and deep DAL methods on real datasets, validating its robustness and classification accuracy. Existing research results indicate that

multi-source domain adaptation can effectively avoid the “negative transfer” situation caused by a single source domain. However, multi-source domain adaptation increases computational complexity. Therefore, constructing the EDPC model based on multi-source domain adaptation is a direction worthy of further research in this article.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YD: Formal analysis, Methodology, Validation, Writing – original draft. DZ: Data curation, Validation, Writing – review & editing. ZW: Data curation, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was

References

- Baktashmotlagh, M., Harandi, M. T., Lovell, B. C., and Salzmann, M. (2013). “Unsupervised Domain Adaptation by Domain Invariant Projection” in *IEEE International Conference on Computer Vision*. IEEE Computer Society.
- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. W. (2010). A theory of learning from different domains. *Mach. Learn.* 79, 151–175. doi: 10.1007/s10994-009-5152-4
- Breiman, (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Bruzzone, L., and Marconcini, M. (2010). Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 770–787. doi: 10.1109/TPAMI.2009.57
- Chen, Z. L., Zhang, J. Y., Liang, X. D., and Lin, L. (2019). “Blending-target domain adaptation by adversarial meta-adaptation networks,” in *Proceeding of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA-United States*. June 15–20, 2019.
- Chen, H., Li, Z., Jin, M., and Li, J., (2021). “Meernet: Multi-source eeg-based emotion recognition network for generalization across subjects and sessions.” in *Proceeding of 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021. pp. 6094–6097.
- Chu, W. S., Torre, F. D. L., and Cohn, J. F., (2013). “Selective transfer machine for personalized facial action unit detection,” in *Proceeding of 2013 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA*. June 23–28, 2013. pp: 3515–3522
- Coomans, D., and Massart, L. D. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules. *Anal. Chim. Acta* 136, 15–27. doi: 10.1016/S0003-2670(01)95359-0
- Dan, Y. F., Tao, J. W., Fu, J. J., and Zhou, D. (2021). Possibilistic clustering-promoting semi-supervised learning for eeg-based emotion recognition. *Front. Neurosci.* 15:690044. doi: 10.3389/fnins.2021.690044
- Dan, Y. F., Tao, J. W., and Zhou, D. (2022). Multi-model adaptation learning with possibilistic clustering assumption for EEG-based emotion recognition. *Front. Neurosci.* 16. doi: 10.3389/fnins.2022(16):855421
- Ding, Z. M., Li, S., Shao, M., and Fu, Y., (2018). “Graph adaptive knowledge transfer for unsupervised domain adaptation.” in *Proceeding of European Conference on Computer Vision, Munich, Germany*. September 8–14, 2018, pp. 36–52.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194. doi: 10.1126/science.1076358
- Du, X., Ma, C., Zhang, G., Li, J., Lai, Y.-K., Zhao, G., et al. (2020). An efficient lst m network for emotion recognition from multichannel eeg signals. *IEEE Trans. Affect. Comput.* 13:1. doi: 10.1109/TAFFC.2020.3013711
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2016). Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17, 2096–2030. doi: 10.48550/arXiv.1505.07818
- Ghifary, M., Balduzzi, D., Kleijn, W. B., and Zhang, M. (2017). Scatter component analysis: a unified framework for domain adaptation and domain generalization. *IEEE Trans. Pattern Anal. Mach. Intel.* 99:10.1109/TPAMI.2016.2599532, 1–1.
- Gong, B., Shi, Y., Sha, F., and Grauman, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. *IEEE Conf. Comp. Vision Pattern Recogn.* 2012, 2066–2073. doi: 10.1109/CVPR.2012.6247911
- Greton, A., Borgwardt, K. M., Rasch, M., Scholkopf, B., and Smola, A. J. (2007). “A kernel method for the two-sample-problem.” in *Proceeding of the 21st Annual Conference on Neural Information Processing Systems, Vancouver, B.C., Canada*. Dec. 3–6.
- Greton, A., Harchaoui, Z., and Fukumizu, K. J., Sriperumbudur, B. K. (2010). A fast, consistent kernel two-sample test.” in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*. pp:673–681
- Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545
- Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834
- Kang, G. L., Jiang, L., Wei, Y., Yang, Y., and Hauptmann, A. (2022). Contrastive adaptation network for single- and multi-source domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intel.* 44, 1793–1804. doi: 10.1109/TPAMI.2020.3029948
- Kim, M. K., Kim, M., Oh, E., and Kim, S. P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Comput. Math. Methods Med.* 2013, 1–13. doi: 10.1155/2013/573734
- Krishnapuram, R., and Keller, J.-M. (1993). A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* 1, 98–110. doi: 10.1109/91.227387
- Lan, Z., Sourina, O., Wang, L., Scherer, R., and Muller-Putz, G. R. (2018). Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/TCDS.2018.2826840
- Lee, S. M., Kim, D. W., and Kim, N., Jeong, S.-G., (2019). “Drop to adapt: Learning discriminative features for unsupervised domain adaptation.” in *Proceeding of 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South)*. October 27–November 2. PP:90–100.

supported by the Ningbo Natural Science Foundation project (No. 2022J180), and by Basic Public Welfare Research Program of Zhejiang Province (No. ZCLTY24F0301), and by Special Polymer Materials for Automobile Key Laboratory of Sichuan Province (No. TZGC2023ZB-06), and by Key Laboratories of Sensing and Application of Intelligent Optoelectronic System in Sichuan Provincial Universities (No. ZNGD2309).

Conflict of interest

ZW was employed by Ningbo Vichnet Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018). Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* 12:162. doi: 10.3389/fnins.2018.00162
- Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2018a). EEG emotion recognition based on graph regularized sparse linear regression. *Neural. Process. Lett.* 47, 1–19. doi: 10.1007/s11063-017-9609-3
- Li, Y., Zheng, W., Cui, Z., Zhang, T., and Zong, Y. (2018b). “A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition.” in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Li, Y., Zheng, W., Zong, Y., Cui, Z., Zhang, T., and Zhou, X. (2018c). A bi-hemisphere domain adversarial neural network model for eeg emotion recognition. *IEEE Trans. Affect. Comput.* 12, 494–504. doi: 10.1109/TAFFC.2018.2885474
- Li, Y., Jin, M., Zheng, W. L., and Lu, B. L. (2018d). Cross-subject emotion recognition using deep adaptation networks. *Neural Inform. Proces.* 11305, 403–413. doi: 10.1007/978-3-030-04221-9_36
- Li, Y., Zheng, W., Wang, L., Zong, Y., and Cui, Z. (2019). From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Trans. Affect. Comput.* 13, 568–578. doi: 10.1109/TAFFC.2019.2922912
- Li, J., Qiu, S., du, C., Wang, Y., and He, H. (2020). Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Trans., Cogn. Dev. Syst.* 12, 344–353. doi: 10.1109/TCDS.2019.2949306
- Li, Y., Wang, L., Zheng, W., Zong, Y., Qi, L., Cui, Z., et al. (2020). A novel bi-hemispheric discrepancy model for EEG emotion recognition. *IEEE Trans., Cogn. Dev. Syst.* 13, 354–367. doi: 10.1109/TCDS.2020.2999337
- Liang, J., He, R., Sun, Z. N., and Tan, T. (2019). Aggregating Randomized Clustering-Promoting Invariant Projections for Domain Adaptation. *IEEE Trans. Pattern Anal. Mach. Intel.* 41, 1027–1042. doi: 10.1109/TPAMI.2018.2832198
- Long, M. S., Wang, J. M., Ding, G. G., Sun, J., and Yu, P. S. (2013). “Transfer Feature Learning with Joint Distribution Adaptation.” in *Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE*.
- Long, M. L., Cao, Y., Wang, J., and Jordan, M. (2015). “Learning transferable features with deep adaptation networks.” in *Proceedings of the 32nd International Conference on Machine Learning, Lille*. pp. 97–105.
- Long, M. S., Wang, J. M., and Jordan, M. I. (2016). “Unsupervised domain adaptation with residual transfer networks.” in *Proceeding of the 30th Annual Conference on Neural Information Processing Systems, Barcelona, Spain, December 5–10, 2016*. pp. 136–144.
- Luo, Y., Zhang, S. Y., Zheng, W. L., and Lu, B. L. (2018). Wgan domain adaptation for EEG-based emotion recognition, International Conference on Neural Information Processing.
- Luo, L. K., Chen, L. M., Hu, S. Q., Lu, Y., and Wang, X. (2020). Discriminative and geometry aware unsupervised domain adaptation. *IEEE Trans. Cybernetics* 50, 3914–3927. doi: 10.1109/TCYB.2019.2962000
- Luo, T., Zhang, J., Qiu, Y. W., Zhang, L., Hu, Y., Yu, Z., et al. (2024). MDDD- manifold-based domain adaptation with dynamic distribution for non-deep transfer learning in cross-subject and cross-session EEG. arXiv:2404.15615v1
- Ma, B. Q., Li, H., Zheng, W. L., and Lu, B. L. (2019). Reducing the subject variability of eeg signals with adversarial domain generalization, in Neural Information Processing, 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I. pp. 30–42.
- Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Comput. Interfaces* 1, 66–84. doi: 10.1080/2326263X.2014.912881
- Musha, T., Terasaki, Y., Haque, H. A., and Ivamitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artif. Life Robot.* 1, 15–19. doi: 10.1007/BF02471106
- Nie, F. P., Huang, H., Cai, X., and Ding, C. (2010). “Efficient and robust feature selection via joint L2,1-norms minimization.” in *Proceedings of the 23rd International Conference on Neural Information Processing Systems. Curran Associates Inc.* pp. 1813–1821.
- Pan, S. J., Tsang, I. W., and Kwok, J. (2011). et al, Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* 22, 199–210. doi: 10.1109/TNN.2010.2091281
- Pandey, P., and Seeja, K. (2019). “Emotional state recognition with EEG signals using subject independent approach” in *Data Science and Big Data Analytics (Springer), Women In Research*. 117–124. Available at: <https://api.semanticscholar.org/CorpusID:70190352>
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Patel, V. M., Gopalan, R., Li, R., and Chellappa, R. (2015). Visual domain adaptation: a survey of recent advances. *IEEE Signal Process. Mag.* 32, 53–69. doi: 10.1109/MSP.2014.2347059
- Pinheiro, P. O. (2018). “Unsupervised domain adaptation with similarity learning.” in *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 8004–8013.
- Shi, L. C., and Lu, B. L. (2010). “Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning.” in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology, 2010*. pp. 6587–6590.
- Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11:1. doi: 10.1109/TAFFC.2018.2817622
- Sun, B., Feng, J., and Saenko, K. (2016). “Return of frustratingly easy domain adaptation.” in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, ser. AAAI’16*. AAAI Press. p. 2058–2065.
- Sun, Y., Gao, Y., Zhao, Y., Liu, Z., Wang, J., Kuang, J., et al. (2022). Neural network-based tracking control of uncertain robotic systems: predefined-time nonsingular terminal sliding-mode approach. *IEEE Trans. Ind. Electron.* 69, 10510–10520. doi: 10.1109/TIE.2022.3161810
- Suykens, J., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural. Process. Lett.* 9, 293–300. doi: 10.1023/A:1018628609742
- Tang, H., and Jia, K. (2019). “Discriminative adversarial domain adaptation.” in *Proceeding of the 34th National Conference on Artificial Intelligence, New York, USA-United States*. Feb. 7–12.
- Tao, J., Chung, F. L., and Wang, S. (2012). On minimum distribution discrepancy support vector machine for domain adaptation. *Pattern Recogn.* 45, 3962–3984. doi: 10.1016/j.patcog.2012.04.014
- Tao, J. W., and Dan, Y. F. (2021). Multi-source co-adaptation for EEG-based emotion recognition by mining correlation information. *Front. Neurosci.* 15:677106. doi: 10.3389/fnins.2021.677106
- Tao, J. W., Dan, Y. F., and Zhou, D. (2021). Robust multi-source co-adaptation with adaptive loss minimization. *Signal Process. Image Commun.* 99:6455. doi: 10.1016/j.image.2021.116455
- Tao, J. W., Dan, Y. F., Zhou, D., and He, S. S. (2022). Robust latent multi-source adaptation for cephalogram-based emotion recognition. *Front. Neurosci.* 16:850906. doi: 10.3389/fnins.2022.850906
- Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Pattern Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006
- Tao, J. W., Wen, S. T., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neural Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009
- Tao, J. W., Wen, S. T., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowledge-Based Syst.* 98, 76–94. doi: 10.1016/j.knsys.2016.01.021
- Tao, J. W., Zhou, D., Liu, F. Y., and Zhu, B. (2019). Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Pattern Recogn.* 87, 296–316. doi: 10.1016/j.patcog.2018.10.023
- Tao, J. W., Dan, Y. F., and Zhou, D. (2023). Possibilistic distribution distance metric: a robust domain adaptation learning method. *Front. Neurosci.* 17:1247082.
- Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. (2014). Deep domain confusion: Maximizing for domain invariance, CoRR, vol. abs/1412.3474. Available at: <http://arxiv.org/abs/1412.3474>
- Wang, L., Ding, Z., and Fu, Y. (2018). Adaptive graph guided embedding for multi-label annotation. In: *IJCAI*. 2798–2804.
- Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., and Lu, W. (2022). “Generalizing to Unseen Domains: A Survey on Domain Generalization.” in *IEEE Transactions on Knowledge and Data Engineering*.
- Zhang, G. Y., Etamad, A., (2022). PARSE: Pairwise Alignment of Representations in Semi-Supervised EEG Learning for Emotion Recognition, arXiv preprint arXiv: 2202.05400.
- Zhang, Y., Tian, F., Wu, H., Geng, X., Qian, D., Dong, J., et al. (2017). Brain MRI tissue classification based fuzzy clustering with competitive learning. *J. Med. Imaging Health Inform.* 7, 1654–1659. doi: 10.1166/jmhi.2017.2181
- Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2020). Alzheimer’s disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fusion* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002
- Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019). Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition. *IEEE Access* 7, 127600–127614. doi: 10.1109/ACCESS.2019.2937657
- Zheng, W. (2017). Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Trans., Cogn. Dev. Syst.* 9, 281–290. doi: 10.1109/TCDS.2016.2587290
- Zheng, W. L., and Lu, B. L. (2016). “Personalizing EEG-based affective models with transfer learning.” in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press. pp. 2732–2738.
- Zheng, W. L., and Lu, B. L. (2015a). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497
- Zheng, W. L., Zhang, Y. Q., Zhu, J. Y., and Lu, B.-L. (2015b). “Transfer components between subjects for EEG-based emotion recognition.” in *Proceedings*

of *International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, 2015. pp. 917–922.

Zheng, W.-L., Liu, W., Lu, Y., Lu, B. L., and Cichocki, A. (2018). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybernetics* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 13, 1290–1301. doi: 10.1109/TAFFC.2020.2994159

Zhou, R., Zhang, Z., Yang, X., Zhang, L., Li, L., Huang, G., et al. (2022). A novel transfer learning framework with prototypical representation based pairwise learning for cross-subject cross-session EEG-based emotion recognition. *ArXiv abs/2202.06509*.

Zhou, R. S., Ye, W. S., Zhang, Z. G., Luo, Y., Zhang, L., Li, L., et al., (2023). EEGMatch: learning with incomplete labels for semi-supervised EEG-based cross-subject emotion recognition. *arXiv:2304.06496v1*

Zhu, J., Arbor, A., and Hastie, T. (2006). Multi-class adaboost. *Stat. Interface* 2, 349–360. doi: 10.4310/SII.2009.v2.n3.a8