



Multi-Model Adaptation Learning With Possibilistic Clustering Assumption for EEG-Based Emotion Recognition

Yufang Dan^{1,2*†}, Jianwen Tao^{1†} and Di Zhou^{3*}

¹ Institute of Artificial Intelligence Application, Ningbo Polytechnic, Ningbo, China, ² Key Laboratory of 3D Printing Equipment and Manufacturing in Colleges and Universities of Fujian Province, Fujian, China, ³ Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Liang Yu,
Shanghai Jiao Tong University, China
Tingyang Chen,
Wuhan University of Technology,
China

*Correspondence:

Yufang Dan
2018020@nbpt.edu.cn
Di Zhou
sion2005@sasu.edu.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 15 January 2022

Accepted: 25 February 2022

Published: 04 May 2022

Citation:

Dan Y, Tao J and Zhou D (2022)
Multi-Model Adaptation Learning With
Possibilistic Clustering Assumption
for EEG-Based Emotion Recognition.
Front. Neurosci. 16:855421.
doi: 10.3389/fnins.2022.855421

In machine learning community, graph-based semi-supervised learning (GSSL) approaches have attracted more extensive research due to their elegant mathematical formulation and good performance. However, one of the reasons affecting the performance of the GSSL method is that the training data and test data need to be independently identically distributed (IID); any individual user may show a completely different encephalogram (EEG) data in the same situation. The EEG data may be non-IID. In addition, noise/outlier sensitiveness still exist in GSSL approaches. To these ends, we propose in this paper a novel clustering method based on structure risk minimization model, called multi-model adaptation learning with possibilistic clustering assumption for EEG-based emotion recognition (MA-PCA). It can effectively minimize the influence from the noise/outlier samples based on different EEG-based data distribution in some reproduced kernel Hilbert space. Our main ideas are as follows: (1) reducing the negative impact of noise/outlier patterns through fuzzy entropy regularization, (2) considering the training data and test data are IID and non-IID to obtain a better performance by multi-model adaptation learning, and (3) the algorithm implementation and convergence theorem are also given. A large number of experiments and deep analysis on real DEAP datasets and SEED datasets was carried out. The results show that the MA-PCA method has superior or comparable robustness and generalization performance to EEG-based emotion recognition.

Keywords: semi-supervised learning, multi-model adaptation, clustering assumption, encephalogram, fuzzy entropy, emotion recognition

INTRODUCTION

Emotion is a psychological experience from human beings of the world, which is complex and changeable (Dolan, 2002; Zhang et al., 2016, 2019b). Different human beings have different emotional understanding on the same thing and may make misjudgment about the emotion occasionally, let alone machines. Therefore, emotion recognition has attracted great attention from researchers (Kim et al., 2013; Mühl et al., 2014; Zhao et al., 2015, 2016; Chu et al., 2017). In this paper, we mainly recognize the corresponding emotion by the internal changes of the human body which include the heart rate, blood pressure, respiratory rate,

magneto encephalogram, electroencephalogram (Mühl et al., 2014), and so on. Generally, most existing EEG-based emotion recognition systems are divided into two steps: data preparation and classifier training (Lan et al., 2018; Zhang et al., 2020). EEG feature extraction methods are comprehensively sorted out in Jenke et al. (2014). In order to improve the recognition accuracy, there existed many EEG-based emotion recognition approaches (Musha et al., 1997; Kim et al., 2013). A satisfactory state of emotion detection based on brain computer interface (BCI) is to detect the emotional state by a real-time EEG signal without inputting signal from the subjects (Zhang et al., 2019b). Different feedback is given to different emotional states in the meantime. These proposed methods (Zhang et al., 2016, 2017) are used to recognize multiple emotion classes from EEG. The latest affective BCIs adopted machine learning algorithms and relied on a few support vectors (Jenke et al., 2014; Mühl et al., 2014). For recording of the EEG signal of the expected target emotion, it is necessary to provide emotional stimulation of the expected concrete emotion to subjects. In the training/calibration stage, the EEG datasets with labels are used to train the emotion recognizer. Many researchers have reported sound classification performance on emotion recognition from real-time EEG data (Mühl et al., 2014).

Due to the high cost of obtaining labeled data, semi-supervised learning (SSL) technology has appeared. It only needs a small, labeled data and a large, unlabeled data to learn a model, which solves the problem on supervised learning needing a large number of labeled samples. Tu and Sun (2013) presented an EEG classifier *via* SSL feature extraction strategy. Tao et al. (2015, 2016, 2017) and Wu and Deng (2018) showed a SSL method for reducing the possible negative impact from random initialization parameters on neural networks. Zu et al. (2019) proposed that remote sensing image classification method based on SSL can effectively improve the accuracy of land cover classification and has a higher efficiency in remote sensing image classification since graph-based semi-supervised learning (GSSL) (Li and Zhou, 2011; Liu et al., 2012; Wang et al., 2012), with its good performance, has been extensively studied. The manifold regularization (MR) (Belkin et al., 2006; Gao et al., 2010; Nie et al., 2010) is a popular GSSL method. A general MR framework was presented by Nie et al. (2010).

In general, the clustering assumption is a basic assumption in GSSL: similar samples should belong to the same class (Chapelle, 2006; Zhu and Goldberg, 2009; Xue et al., 2011; Zhou et al., 2014; Wang et al., 2019). In other words, each sample only have one label, which we called hard classification. However, in the real applications of emotion recognition, its performance will be discounted by this assumption—for example, for different subjects in different scenes, crying may be understood as sad and happy.

To handle the limitation from this assumption, Wang et al. (2012) and Zhang et al. (2019b) proposed a novel clustering assumption that can significantly boost the classifier performance. Assuming that similar samples have the same label membership, each sample may have multiple membership values (Zhang et al., 2019a), not only one. However, this method has a constraint in that the sum of the membership values of each sample is 1. This constraint may lead to the membership values

of some noise being close to or even greater than those of normal samples; it may lead to misrecognition.

According to this problem in the SSCCM method, Dan et al. (2021) proposed SSPCA that relaxed the constraint in SSCCM and added a fuzzy entropy regularization term (Kosko, 1986; Krishnapuram and Keller, 1993; Zhang et al., 2019c) for increasing the samples' discriminative information to get a membership function with better generalization and that can also reduce the negative impact of noise and outlier on recognition performance to improve the robustness of the method. Wang and Chen (2013) designed SA-SSCCM. Specifically, the SSCCM method is upper boundary and the LS-SVM method is lower boundary, respectively—that is, if unlabeled data is good for model training, the classification result of SA-SSCCM is close to SSCCM; if unlabeled data penalizes model training, the classification result of SA-SSCCM is close to that of the LS-SVM method, and the interference of noise data to the SA-SSCCM model training is avoided. However, both SSPCA method and SA-SSCCM method require training data, and the test data should be independently identically distributed (IID). Due to the difference among different subjects in real emotion recognition applications, it may cut recognition accuracy. It is hard to guarantee that two datasets are IID.

Toward the problem of reduced recognition accuracy caused by non-IID of training data and test data, this paper adopts domain adaptation learning (DAL) (Bruzzone and Marconcini, 2010; Tao et al., 2021) related to computer vision and machine learning (Bishop, 2006; Zhu, 2008). Generally, DAL includes instance-based DAL, feature-based DAL, and model-based DAL (Pan and Yang, 2010). The instance-based DAL and feature-based DAL need to access instances from the source domain during the model learning. When the source dataset is relatively large, the training efficiency will be reduced. The model-based DAL uses the pre-trained source classifier on some source datasets to learn an effective target classifier, which has good classification effectiveness and high efficiency. Much more DAL categories in Pan and Yang (2010) can be found. Therefore, this paper proposes a multi-model adaptation learning with possibilistic clustering assumption for EEG-based emotion recognition (MA-PCA).

The main ideas are as follows: firstly, according to manifold learning (Tenenbaum et al., 2000; Belkin and Niyogi, 2001; Gao et al., 2010; Nie et al., 2010), there is similarity among samples within its local. According to formula (1) in Dan et al. (2021), the local weighted mean (LWM) point is determined by the convex hull of the k -nearest neighbors. It represents the mean value of the local. Therefore, the neighbors in the local should have consistency with the mean value—that is, the labels of each neighbor in the local and its corresponding LWM should be similar (or consistent). Then, it is assumed that each neighbor has a similar label membership to its corresponding LWM (Bottou and Vapnik, 1992; Atkeson et al., 1997; Xue and Chen, 2007); secondly, the classification prediction results are mutually verified by the decision function and the membership function to improve the classification reliability; thirdly, a fuzzy entropy regularization term is proposed to increase the sample discrimination information; then, we can get a membership function with better generalization, and the negative impact of noise and outlier will be relaxed on

recognition performance. Finally, a classification model with better generalization performance is obtained by adding a multi-model adaptation regularization term for IID and non-IID on training data and test data, respectively. The major contributions of this work are the following:

- (1) A multi-model adaptation learning with possibilistic clustering assumption for EEG-based (MA-PCA) is proposed.
- (2) Since a multiple auxiliary discriminant model is good for SSL with a small labeled instance, the regular term has Laplacian local consistency and the regularize term has different weights of multiple-source models. It aims to expand the discriminant space of the target domain and guarantee the local structure consistency among inner samples of source domain and target domain. At the end, it solves the non-IID problem between the training data and test data, too.
- (3) Finally, our comprehensive experiments on real datasets (i.e., DEAP, SEED) show that the method has better robustness and generalization.

The remainder of this work is organized as follows: In section “Proposed Framework,” our framework MA-PCA will be designed which includes MA-PCA formulation, optimization, and convergence analysis, and section “Algorithm of MA-PCA” arranges the corresponding optimal algorithm of MA-PCA. The experimental results and analysis on two real EEG datasets (i.e., DEAP and SEED) are presented in section “Experimental Evaluation.” Finally, we conclude in section “Conclusion.”

MA-PCA FRAMEWORK

This section will introduce the concept of our multi-model adaptation learning with possibilistic clustering assumption for EEG-based (MA-PCA) framework in detail. It mainly uses multiple-source models which are obtained from existing relevant source datasets to learn the robust semi-supervised classification model. Therefore, the two core components are organically unified into MA-PCA: (1) any instance should have a similar label membership with its corresponding LWM. The fuzzy entropy regularization term is added to reach the amount of discrimination information to improve the classification accuracy and robustness and (2) assuming that multiple-source models can help SLL, the existing multi-source models are used for multi-source domain adaptation learning to establish a robust target domain classification model. At the same time, considering IID and non-IID, the best source model is found by multiple-source models with different weights to train the target model.

Notations

We denote $X = \{x_1, x_2, \dots, x_i, x_{i+1}, \dots, x_n\}$ as a feature dataset, where n is sample number ($l \ll n$) and $Y_l = \{y_1, y_2, \dots, y_l\}^T \in \mathbb{R}^{l \times M}$ is a label set about dataset $X_l = \{x_i\}_{i=1}^l$. $X_u = \{x_j\}_{j=l+1}^n$ is an unlabeled feature dataset, where x_i is d dimensions ($x_i \in R^d$) vector of the i -th sample. We compute LWM \hat{x}_i about x_i :

$$\hat{x}_i = \frac{\sum_{x_j \in Ks(x_i)} D_{ij} x_j}{\sum_{x_j \in Ks(x_i)} D_{ij}}, \tag{1}$$

where the k nearest neighbors of x_i are arranged in $Ks(x_i)$, and the Euclidean distance algorithm is used to find these neighbors. We design an undirected weight graph $G = (X, D)$, where $D \in \mathbb{R}^{n \times n}$ is weight matrix, $D_{ji} = D_{ij} \geq 0$, and the element is measured as follows:

$$D_{ij} = \begin{cases} \exp(-\tau \|x_i - x_j\|^2), & x_i \text{ is one of the neighbors of } x_j \\ 0 & \text{otherwise} \end{cases},$$

where τ is a changeable parameter in Gaussian kernel function. If the distance between x_i and x_j is smaller, D_{ij} is higher, and *vice versa*. Therefore, the clustering problem is changed into a graph problem in this paper.

Basic Formulation of MA-PCA

Since both SSPCA and SA-SSCCM methods require that the training data and test data meet the IID assumption, this paper reasonably combines the SSPCA method with a multi-model adaptation learning method (i.e., MA-PCA). This proposed method not only improves the robustness on noises/outliers but also solves the problems of insufficient label data and noisy data affecting the performance of the model and the different distribution of training data and test data. We therefore propose the following basic formula of MA-PCA:

$$Q(W, v_m(x_j), \gamma) = \min \Omega_B(W, v_m(x_j)) + \beta \Omega_M(W, \gamma), \tag{2}$$

where $\Omega_M(W, \gamma)$ is the multi-model adaptation term, and $\Omega_B(W, v_m(x_j))$ is used for reducing the negative impact of noises/outliers. We have the following function:

$$\begin{aligned} & \Omega_B(W, v_m(x_j)) \\ &= \min_{w, v_m(x_j)} \sum_{i=1}^l \|W^T x_i - y_i\|^2 + \lambda_s \sum_{i=1}^l \|W^T \hat{x}_i - y_i\|^2 \\ &+ \sum_{m=1}^M \sum_{j=l+1}^n v_m^2(x_j) \|W^T x_j - c_m\|^2 \\ &+ \lambda_s \sum_{m=1}^M \sum_{j=l+1}^n v_m^2(x_j) \|W^T \hat{x}_j - c_m\|^2 + \lambda \|W^T\|_{\mathcal{H}}^2 \\ &+ C \sum_{m=1}^M \sum_{j=l+1}^n (v_m^2(x_j) \ln v_m^2(x_j) - v_m^2(x_j)) \\ & \text{s.t. } 0 \leq v_m(x_j) \leq 1, m = 1, \dots, M, j = l+1, \dots, n, \end{aligned} \tag{3}$$

where λ_s, λ, C are balance parameters that can be adjusted to avoid overfitting during model training. The details about the other parameters are provided in Dan et al. (2021).

The Multi-Model Adaptation Term in MA-PCA

In our domain adaptation learning, given is $\{W_i^s\}_{i=1}^q$ as a multiple-source model set, where q is source model number, and W_i^s is the i -th source model. Each source model is obtained by learning the specified public dataset. This paper expects the

classification results of the target model to be consistent with those of the source domain models. In other words, this paper will learn the target classifier $f(x) = \mathbf{W}^T \mathbf{x}$ in the whole sample space, regularize the weight parameters of different source domain models to control its complexity, and make each target instance be close to the source domain models. According to this criterion, it can be realized by introducing the multi-source domain model adaptation regularization function on the target domain. The formula is described as follows:

$$\begin{aligned} \Omega_M(\mathbf{W}, \gamma) &= \sum_{i=1}^q \gamma_i \int_X \|\mathbf{W}^T \mathbf{x} - \mathbf{W}_i^T \mathbf{x}\|^2 dx + \eta \|\gamma\|_2^2 \\ &\triangleq \sum_{i=1}^q \gamma_i \frac{1}{n} \sum_{j=1}^n \|\mathbf{W}^T \mathbf{x}_j - \mathbf{W}_i^T \mathbf{x}_j\|_F^2 + \eta \|\gamma\|_2^2 \\ &= \sum_{i=1}^q \gamma_i \text{tr} \left[(\mathbf{W} - \mathbf{W}_i^s)^T \mathbf{S} (\mathbf{W} - \mathbf{W}_i^s) \right] + \eta \|\gamma\|_2^2, \end{aligned} \tag{4}$$

where $\gamma = [\gamma_1, \dots, \gamma_q]^T$, and γ_i is a weight of the i -th source model. We constraint that the sum of γ_i is 1 (i.e., $\sum_{i=1}^q \gamma_i = 1$). It is better to explore the contribution among the source models. The divergence matrix of the target domain is $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. $\eta \in \mathbb{R}^+$ is a balance parameter to control the contribution of $\|\gamma\|_2^2$. This parameter can be changed to avoid overfitting on multiple-source models. In addition, in the second equation in (4), we employ the sampling frequency as a weight to access the real distribution for the target domain.

Remark 1: The divergence matrix \mathbf{S} is important for connecting the source classifiers and the target classifier. It will promote the learning of the target classification model to the real distribution direction of the target domain, thereby improving the generalization performance of adaptive learning, which is essentially different from other domain adaptation regularization terms based on DAL models (Bottou and Vapnik, 1992; Duan et al., 2012a). In order to better fit the model idea, this paper refers to formula (4) as the construction of the regularization term for divergence-constrained multi-model adaptation.

Final Formulation

We expect better model adaptation performance for EEG-based emotion recognition by combining SSPCA with scatter-constrained multi-source classifier model. Therefore, a unified framework MA-PCA is obtained to learn $\mathbf{W}, \mathbf{v}_m(\mathbf{x}_j), \gamma$ by combining formulas (3) and (4). The optimization problem of MA-PCA can be described as follows:

$$\begin{aligned} Q(\mathbf{W}, \mathbf{v}_m(\mathbf{x}_j), \gamma) &= \min_{\mathbf{w}, \mathbf{v}_m(\mathbf{x}_j)} \sum_{i=1}^l \|\mathbf{W}^T \mathbf{x}_i - \mathbf{y}_i\|^2 + \lambda_s \sum_{i=1}^l \|\mathbf{W}^T \hat{\mathbf{x}}_i - \mathbf{y}_i\|^2 \\ &\quad + \sum_{m=1}^M \sum_{j=l+1}^n \mathbf{v}_m^2(\mathbf{x}_j) \|\mathbf{W}^T \mathbf{x}_j - \mathbf{c}_m\|^2 \end{aligned}$$

$$\begin{aligned} &+ \lambda_s \sum_{m=1}^M \sum_{j=l+1}^n \mathbf{v}_m^2(\mathbf{x}_j) \|\mathbf{W}^T \hat{\mathbf{x}}_j - \mathbf{c}_m\|^2 + \lambda \|\mathbf{W}^T\|_{\mathcal{H}}^2 \\ &+ C \sum_{m=1}^M \sum_{j=l+1}^n (\mathbf{v}_m^2(\mathbf{x}_j) \ln \mathbf{v}_m^2(\mathbf{x}_j) - \mathbf{v}_m^2(\mathbf{x}_j)) \\ &+ \beta \left\{ \sum_{i=1}^q \gamma_i \text{tr} \left[(\mathbf{W} - \mathbf{W}_i^s)^T \mathbf{S} (\mathbf{W} - \mathbf{W}_i^s) \right] \right\}, \end{aligned} \tag{5}$$

where $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_q]^T$, $\gamma \in \mathbb{R}^{q \times 1}$, $\gamma^T \mathbf{1}_q = 1$, and q are the number of source domain models. When $\beta = 0$, MA-PCA degenerated to SSPCA. When $\beta > 0$, β is used as a balance parameter. When γ_i is constant 1, it indicates that there is only one single-source domain, and its distribution is the same as the target domain. At this time, MA-PCA approximates the SA-SSCCM method. When $0 \leq \gamma_i < 1$, MA-PCA is a multi-model adaptation learning method based on the possibility clustering assumption, and the distribution of source domain and target domain can be identical or non-identical.

OPTIMIZATION

The objective function (5) is a non-convex function on $(\mathbf{W}, \mathbf{v}_m(\mathbf{x}_j), \gamma)$. In this paper, the strategy of alternating iterative optimization is adopted to realize the optimal solution of decision model \mathbf{W} , membership model $\mathbf{v}_m(\mathbf{x}_j)$, and contribution coefficient γ of the source models, respectively, and each iteration has an optimal solution.

Update \mathbf{W} as Given $\mathbf{v}_m(\mathbf{x}_j)$ and γ

Firstly, fixing $\mathbf{v}_m(\mathbf{x}_j)$ and γ to solve \mathbf{W} : for ease of calculation, the following formula (5) is transformed into matrix form. \mathbf{W} can be written as $\mathbf{W} = \sum_{i=1}^n x_i \alpha_i$. Under the Representation Theorem, the formula (5) exists in reproducing kernel Hilbert space, and the kernel of \mathbf{W} can be rewritten as follows: $\mathbf{W} = \sum_{i=1}^n K(x_i, x) \alpha_i$ (Belkin et al., 2006). Therefore, formula (5) is mapped into a finite dimensional space of the optimization α_i and can be reformulated as follows:

$$\begin{aligned} Q(\alpha) &= \min_{\alpha} \text{tr} \left((\alpha^T \mathbf{K}_l \mathbf{K}_l - \mathbf{Y}) (\alpha^T \mathbf{K}_l \mathbf{K}_l - \mathbf{Y})^T \right) \\ &\quad + \lambda_s \text{tr} \left((\alpha^T \overline{\mathbf{K}}_l \overline{\mathbf{K}}_l - \mathbf{Y}) (\alpha^T \overline{\mathbf{K}}_l \overline{\mathbf{K}}_l - \mathbf{Y})^T \right) \\ &\quad + \text{tr} \left((\alpha^T \mathbf{K}_u \mathbf{K}_u \mathbf{J} - \mathbf{L}) \hat{\mathbf{V}} (\alpha^T \mathbf{K}_u \mathbf{K}_u \mathbf{J} - \mathbf{L})^T \right) \\ &\quad + \lambda_s \text{tr} \left((\alpha^T \overline{\mathbf{K}}_u \overline{\mathbf{K}}_u \mathbf{J} - \mathbf{L}) \hat{\mathbf{V}} (\alpha^T \overline{\mathbf{K}}_u \overline{\mathbf{K}}_u \mathbf{J} - \mathbf{L})^T \right) \\ &\quad + \lambda \text{tr} (\alpha^T \mathbf{K} \mathbf{K} \alpha) + \beta \sum_{i=1}^a \gamma_i (\mathbf{W}_i^T - \alpha^T \mathbf{K}) \mathbf{K} \\ &\quad \cdot \mathbf{K} (\mathbf{W}_i - \mathbf{K} \alpha), \end{aligned} \tag{6}$$

where the details about $\mathbf{J}, \mathbf{L}, \hat{\mathbf{V}}$ can be found in Dan et al. (2021).

By solving the derivation of (6) w.r.t. α and letting it be equal to 0 (i.e., $\partial Q(\alpha)/\partial \alpha = 0$), we obtain the following:

$$\alpha = H^{-1}B, \tag{7}$$

where

$$H = K_l K_l K_l K_l + \lambda \overline{K_l K_l K_l K_l} + K_u K_u J \hat{V} J^T K_u K_u + \lambda_s \overline{K_u K_u J \hat{V} J^T K_u K_u} + \lambda K K + \beta \sum_{i=1}^q \gamma_i K K K K,$$

and

$$B = K_l K_l Y^T + \lambda_s \overline{K_l K_l Y^T} + K_u K_u J \hat{V} L + \lambda_s \overline{K_u K_u J \hat{V} L^T} + \beta \sum_{i=1}^q \gamma_i K K K W_i^s,$$

Finally, the solution of W is $K\alpha$.

Update $v_m(x_j)$ as Given W and γ

Then, we fix W and γ to solve $v_m(x_j)$. The optimal problem of the objective function in (5) can be rewritten as follows:

$$\begin{aligned} Q(v_m(x_j)) &= \min_{v_m(x_j)} \sum_{m=1}^M \sum_{j=l+1}^n v_m(x_j)^2 \|W^T x_j - c_m\|^2 \\ &+ \lambda_s \sum_{m=1}^M \sum_{j=l+1}^n v_k(x_j)^2 \|W^T \hat{x}_j - c_m\|^2 + \lambda \|W^T\|_{\mathcal{H}}^2 \\ &+ C \sum_{m=1}^M \sum_{j=l+1}^n (v_m(x_j)^2 \ln v_m(x_j)^2 - v_m(x_j)^2), \end{aligned} \tag{8}$$

By solving the derivation of Eq. (5) w.r.t. $v_m(x_j)$ and letting it be equal to zero (i.e., $\partial Q(v_m(x_j))/\partial v_m(x_j) = 0$), then

$$\begin{aligned} \partial Q(v_m(x_j))/\partial v_m(x_j) &= 2v_m(x_j) \|W^T x_j - c_m\|^2 \\ &+ 2\lambda_s v_m(x_j) \|W^T \hat{x}_j - c_m\|^2 \\ &+ C [2v_m(x_j) \log v_m(x_j)^2] = 0, \end{aligned}$$

we can get:

$$v_m(x_j) = \exp\left(\frac{-(\|W^T x_j - c_m\|^2 + \|W^T \hat{x}_j - c_m\|^2)}{2C}\right), \tag{9}$$

Since x_j means any one instance, the general presentation of $v_m(x_j)$ is as follows:

$$v_m(x) = \exp\left(\frac{-(\|W^T x - c_m\|^2 + \|W^T \hat{x} - c_m\|^2)}{2C}\right), \tag{10}$$

Update γ by Fixing W and $v_m(x_j)$

We define $A_i = \text{tr}[(W - W_i^s)^T S (W - W_i^s)]$ ($i = 1, \dots, q$) which is corresponding to the adaptation of the i -th source model to the target domain. Given W and $v_m(x_j)$, the objective function (5) can be rewritten in (11) as follows:

$$\begin{aligned} \min_{\gamma} \gamma^T A + \eta \|\gamma\|_2^2 \\ \text{s.t. } \gamma^T \mathbf{1}_q = 1, 0 \leq \gamma \leq 1, \end{aligned} \tag{11}$$

where $A = (A_1, \dots, A_q)^T$. The optimal estimation of γ becomes the optimal weight division problem of multiple model adaptation learning with scatter constraints. Theoretically, if $\eta = 0$, the optimal γ_i will be 1; otherwise, if $A_i = \min_{j=1, \dots, q} A_j$, γ_i is 0; if $\eta \rightarrow +\infty$, the optimal γ_i will tend to be the same weight $1/q$. Thus, the following theorem is obtained:

Theorem 1 (Karasuyama and Mamitsuka, 2013). The following equation is the optimal solution of (11):

$$\gamma_i = \begin{cases} \frac{\rho - A_i}{\zeta}, & i = 1, 2, \dots, \zeta \\ 0, & i = \zeta + 1, \dots, q \end{cases}$$

where $\rho = (\eta + \sum_{i=1}^q A_i)/\zeta$, $\zeta = |\{i | \rho - A_i > 0, i = 1, 2, \dots, q\}|$.

Theorem 1 presents that there are ζ non-zero entries in the optimal γ . According to the target domain, this optimization can select source domains automatically with a different γ . If γ is bigger, there is a higher similarity between the source domain and the target domain. Since the optimal γ could be calculated *via* the optimal ζ , Karasuyama and Mamitsuka, 2013) presented a special method to obtain the optimal ζ . This algorithm effectiveness can be proved under the given right amount of source domains.

Next, the algorithm adopts a coordinate descent strategy to solve (11), which is close to the method in Geng et al. (2012). In each iteration cycle, when other entries are fixed, we select two items to update, and $\gamma^T \mathbf{1}_q = 1$ must be satisfied at the end of each iteration. Suppose that in an iteration cycle the i -th and j -th entries are selected, the following iterative formula can therefore be obtained:

$$\begin{cases} \text{(I)} \gamma_i^* = 0, \gamma_j^* = \gamma_i + \gamma_j, \text{ if } 2\eta(\gamma_i + \gamma_j) + (A_j - A_i) \leq 0 \\ \text{(II)} \gamma_j^* = 0, \gamma_i^* = \gamma_i + \gamma_j, \text{ if } 2\eta(\gamma_i + \gamma_j) + (A_i - A_j) \leq 0 \\ \text{(III) or } \gamma_i^* = 1/2(\gamma_i + \gamma_j) + 1/4\eta(A_j - A_i), \\ \gamma_j^* = \gamma_i + \gamma_j - \gamma_i^*. \end{cases} \tag{12}$$

We iteratively traverse all paired entries in γ and optimize any two entries in γ by (12) until the optimization function (5) converges. Intuitively, the updating criteria in (12) tends to be that the larger the value to γ_i , the smaller the A_i . Since A_i measures the distribution distance between the i -th source model and the target domain, the smaller the A_i , the higher the correlation between the i -th source domain and the target domain.

Remark 1: After obtaining the optimal solution of W and V , the label matrix of the samples in the target domain can also be obtained, and the influence of noise has been effectively suppressed. The label and the membership of out of sample can also be calculated by W and V . Finally, the performance of W relies on the learned graph $G = XD$ and the multiple-source domain adaptation model.

Convergence Analysis

Since the objective function (5) is a multi-objective optimization function, it is difficult to guarantee the global optimal solution. It is worthy to note that the algorithm in this paper adopts an alternate iterative strategy for optimizing. Since the objective function of a single optimization variable [i.e., equations (6), (8), and (11)] is convex and in closed form, the iterative optimal solution can be obtained. Therefore, this work only deduces the asymptotic convergence of the algorithm based on the iterative target value of the objective function. The derivation process is shown in (13) as follows:

$$\begin{aligned}
 Q(W^{itr}, v^{itr}, \gamma^{itr}) &\geq \\
 Q(W^{itr+1}, v^{itr}, \gamma^{itr}) &\geq \\
 Q(W^{itr+1}, v^{itr+1}, \gamma^{itr}) &\geq \\
 Q(W^{itr+1}, v^{itr+1}, \gamma^{itr+1}) &> \varepsilon > 0,
 \end{aligned}
 \tag{13}$$

where $W^{itr}, v^{itr}, \gamma^{itr}$ are the optimal solutions at the itr -th iteration. ε is a very small constant. The objective function will converge to a local optimum. The derivation process proves that the iterative target value of the algorithm shows a downward trend. When the value drops to a certain threshold (at least greater than ε), we stop the iterative. Finally, the objective function will converge to the local target value of each single optimization variable.

ALGORITHM DESCRIPTION

The optimization of MA-PCA adopts an alternating iteration strategy. The most semi-supervised learning methods are often optimized by iterative learning. In addition, the membership value of initialized unlabeled instances can be obtained by any of the following methods: some fuzzy clustering method, randomization strategy, or all initialized to zero. Therefore, the learning of MA-PCA method starts with labeled instances to initialize the decision model W . When the objective function converges, the iteration terminates. The algorithm in this paper specifically adopts a window-based stopping criterion to better control the algorithm convergence: given a window size h , computing $\zeta = |Max_{\Theta_{itr}} - Min_{\Theta_{itr}}| / Max_{\Theta_{itr}}$ at the itr -th iteration ($\Theta_{itr} = \{Obj_{itr-h+1}, \dots, Obj_{itr}\}$) means Θ_{itr} is composited by the historical target value in this window. When $\zeta < \varepsilon$, the iteration terminates. The details of this algorithm are shown in **Table 1**.

TABLE 1 | Algorithm description of MA-PCA.

Input: the target domain with data X_l and its labels Y_l , unlabeled data X_u , regular term parameter $\lambda, \lambda_s, C, \eta, \beta$. There are q source classifier models $\{W_i^s\}_{i=1}^q$, iteration termination threshold ε , and maximum number of iterations N .
Output: the target classifier model W , the label membership function v , and the contribution coefficient γ of multi-source model.
Procedure:
1. Initialize the label memberships of unlabeled data, $\gamma_i^0 = 1/q (i = 1, \dots, q)$;
2. Obtain the initial W^0 by Eq. (7);
3. Obtain the initial v^0 by Eq. (10);
4. Calculate the $Q(W^0, v^0, \gamma^0)$ of objective function
for $itr = 1$ to N do
{
5.1 Fix the current v^{itr} and γ^{itr} for updating W^{itr} to W^{itr+1} by Eq. (7)
5.2 Fix the current W^{itr} and γ^{itr} for updating v^{itr} to v^{itr+1} by Eq. (10)
5.3 Fix the current W^{itr} and v^{itr} for updating γ^{itr} to γ^{itr+1} by Eq. (12)
Until $itr > N$ or $\zeta < \varepsilon$, return the optimal W, v , and γ
}

EXPERIMENT

In this part, we comprehensively compare the proposed method with several state-of-the-arts on two widely used benchmark databases, including SEED (Zheng and Lu, 2015) and DEAP (Koelstra et al., 2012), for EEG-based emotion recognition (Mansour et al., 2009).

Datasets

According to Lan et al. (2018) and Zhong et al. (2020), there exist certain significant differences between SEED and DEAP since they can be generated by different subjects, sessions, EEG devices, experimental schemes, emotional stimuli, etc., Detailed information about these two datasets can be viewed in Lan et al. (2018). In the following experiments, we adopt differential entropy (Lan et al., 2018; Zhong et al., 2020) as the data feature in emotion recognition, which has also been widely used in the preceding literatures (Shi et al., 2015; Zheng et al., 2015, 2016; Chai et al., 2016, 2017; Lan et al., 2018; Zhong et al., 2020) for domain adaptation emotion recognition.

Baseline Setting

We will systematically compare our method with such state-of-the-arts as SSPCA (Dan et al., 2021), a baseline without domain adaptation, FastDAM (Duan et al., 2012b), Multi-KT (Tommasi et al., 2014) with l_2 -norm constraint on p , A-SVM (Yang et al., 2007), and DSM (Duan et al., 2012a). Since existing deep domain adaptation frameworks have achieved many inspiring results on emotion recognition as well as visual recognition, we also additionally present comparisons with several deep (CNN-based) domain adaptation methods with deep features: DAN (Long et al., 2015) and Reverse Grad (Ganin and Lempitsky, 2015).

It was noted that, in the DA schema, automatic parameter tuning is not possible for source classifiers using cross-validation due to the reason that training and test data are from different

data distributions. Therefore, all methods evaluated in this paper on the dataset are empirically searched for the parameter space to optimize parameter settings to obtain the best results for each method. The parameters of all methods are adjusted to obtain the best results, except for the specially specified parameters.

For the method SSPCA without domain adaptation, this experiment will fuse the decision values of all classifiers obtained from the independent training of labeled samples in each source domain and target domain. DSM, Multi-KT, A-SVM, and FastDAM are domain adaptation methods. For A-SVM, this experiment also fuses the decision values of all basic classifiers, and each classifier is learned from a labeled sample in a source domain.¹

There are some hyper-parameters in the objective function (5) that need to be determined. First of all, this experiment sets the most important parameters (e.g., γ_i) as optimization variables in the iterative optimization process, and only a few crucial parameters that are proposed in MA-PCA need to be pre-defined (e.g., λ , λ_s , C , η , β), and considering that parameter determination is an open problem in the field of machine learning, we have determined parameters empirically in a past work. Since the exponent of γ_i plays the role of avoiding trivial solutions in the process of optimizing γ_i , as proved in Hou et al. (2017), the larger the exponent of γ_i is, the closer all weight values are to be same. In order to reflect the differences among different source domains, this experiment will set the index of γ_i to be 2 by experience. The validity of this decision will be verified in the experimental results in the following section. The hyperparameter λ , λ_s , C , η , β is adjusted within the range of $\{10^{-4}, 10^{-3}, \dots, 10^3, 10^4\}$. Finally, we search the nearest neighbor number k from the set $\{35, 10, 15, 17\}$ to construct the nearest neighbor graph in MA-PCA (also SSPCA) and obtain the first-ranked recognition accuracy from the optimal parameter configuration.

For non-linear learning methods MA-PCA, FastDAM, and multi-KT, Gaussian kernel $K_{ij} = \exp(-\sigma\|x_i - x_j\|^2)$ is default kernel function, $\sigma = 1/d$, and d is the feature dimension. In FastDAM, $\gamma_i = \frac{\exp(-\delta \text{Dist}(X_i^s, X))}{\sum_j \exp(-\delta \text{Dist}(X_j^s, X))}$ ($i = 1, \dots, S$) is the weight value of each source domain, $\delta = 100$. For the benchmark method SSPCA, the target domain samples are directly mapped to the source domain without any domain adaptation, and the decision values of all classifiers obtained from the independent training of labeled samples from each source domain and target domain are equally fused.

Emotion Recognition Within Dataset

Note that different subjects even from the same dataset still have different EEG feature distributions due to individual characteristics. We therefore practice the so-called leave-one-out cross-validation strategy conducted also in Lan et al. (2018) to

evaluate the emotion recognition performance of MA-PCA—that is, one subject remained to be the target domain, and the others from the dataset are constructed as multiple sources. In this multi-source scenario, we follow the same setting as Tao et al. (2021) to evaluate our method compared with other state-of-the-arts on SEED and DEAP, respectively.

Performance Comparison

The emotion recognition performance of MA-PCA and the rest of the comparison methods within the DEAP and SEED datasets are visualized in **Figure 1**. It can be seen from the bar plot that the recognition performance of all DA methods is better than that of SSPCA, and MA-PCA achieves the best performance (about 21% performance improvement over SSPCA), followed by DSM on DEAP dataset. Besides this, those multi-source adaptation methods, including our method, unsurprisingly achieved more accuracy gains than the no-adaptation method SSPCA on SEED. We can observe that our method MA-PCA demonstrates the best performance on SEED by upgrading the average accuracy. An interesting observation is that all methods work better on SEED than on DEAP, which has also been reported in Lan et al. (2018) and Tao et al. (2021). The reason for this phenomenon might be that the larger distribution discrepancy between different subjects from DEAP prevented boosting performance in these methods (Mansour et al., 2009; Lan et al., 2018).

Finally, MA-PCA achieved almost the best performance on both datasets. A possible explanation is that the distribution discrepancy may exist in the same dataset (i.e., DEAP or SEED), and MA-PCA can learn a more robust target classifier for domain adaptation by discriminatively selecting a set of prelearned base classifiers in the non-IID scenario of multi-subject adaptation.

Multi-Kernel Learning

As well known, the choice of kernel is a challenging issue in the kernel learning method. Recently, multiple kernel learning (MKL) has been effectively proposed for conquering this choice issue that existed in single kernel learning methods. Consequently, we also evaluate the performance boost in our method by using MKL (called MKMA-PCA for short) for each source domain. To this end, the first step is to construct a new space spanned by multiple kernel mapping features. We firstly denote by $\{\phi_a\}_{a=1}^{\mathcal{U}}$ an empirical kernel function set, which, respectively, projected X_a into \mathcal{U} different spaces. Then, an orthogonally integrated space can be constructed by concatenating these \mathcal{U} spaces. We denote the mapping features in this final space by $\tilde{\phi}(x_i) = [\phi_1(x_i)^T, \phi_2(x_i)^T, \dots, \phi_{\mathcal{U}}(x_i)^T]^T \in \mathbb{R}^{\mathcal{U}n_a}$, where $x_i \in X_a$. Correspondingly, the kernel matrix in this final space can be easily deduced as $K_{new} = [\tilde{K}_1; \tilde{K}_2; \dots; \tilde{K}_{\mathcal{U}}]$, where \tilde{K}_i is the i -th kernel matrix from the \mathcal{U} feature spaces. Aiming to exploit the multiple kernel spaces, we therefore employ four kernel mapping functions, including the Gaussian kernel used above. The other additionally employed kernels are inverse square distance kernel function, Laplacian kernel function, and inverse distance kernel function, respectively, denoted as $\mathbf{K}_{ij} = 1/(1 + \sigma\|x_i - x_j\|^2)$, $\mathbf{K}_{ij} = \exp(-\sqrt{\sigma}\|x_i - x_j\|)$, and $\mathbf{K}_{ij} = 1/(1 + \sqrt{\sigma}\|x_i - x_j\|)$.

¹For each source domain, we train one SVM by using the corresponding labeled samples. Then, for each test instance x , the decision values from p SVM classifiers are converted into probability values by using the sigmoid function [i.e., $g(t) = 1/(1 + \exp(-t))$]. Finally, we average the p probability values as the final prediction of the test instance x .

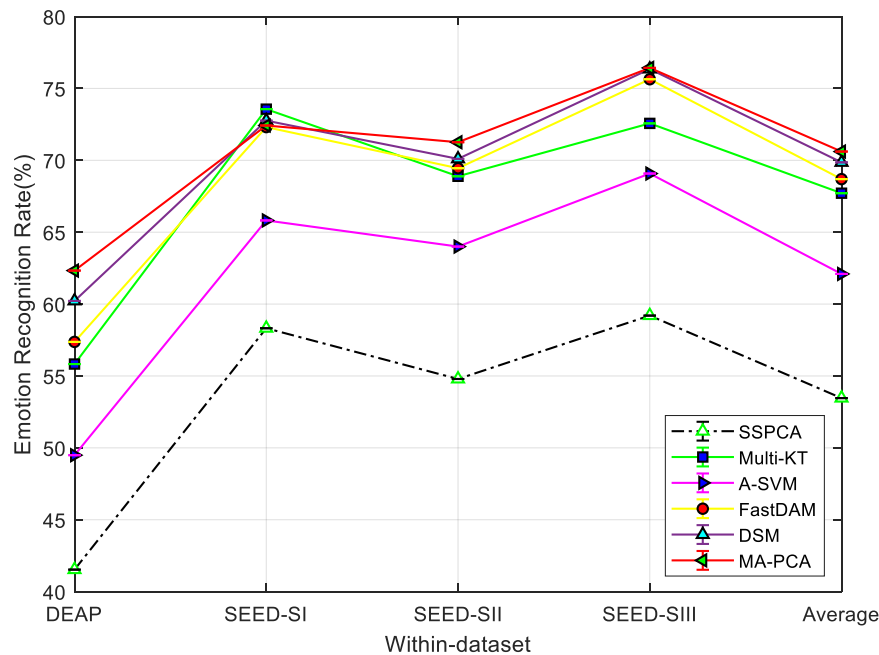


FIGURE 1 | Domain adaptation emotion recognition on within dataset. SI, session I; SII, session II; SIII, session III.

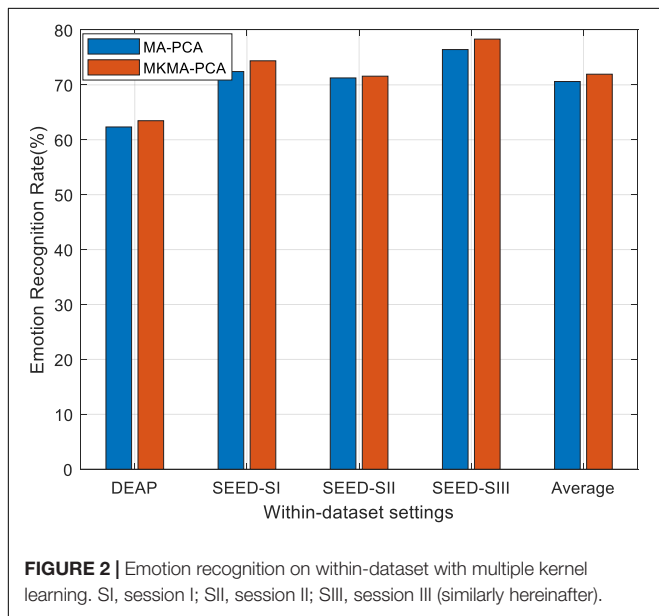


FIGURE 2 | Emotion recognition on within-dataset with multiple kernel learning. SI, session I; SII, session II; SIII, session III (similarly hereinafter).

The observation from **Figure 2**, in which MKMA-PCA significantly outperforms MA-PCA, justifies that our MA-PCA with MKL can further boost the recognition performance on DEAP and SEED. This also proves the importance of kernel choice in those kernel-based learning models.

Emotion Recognition Cross-Dataset

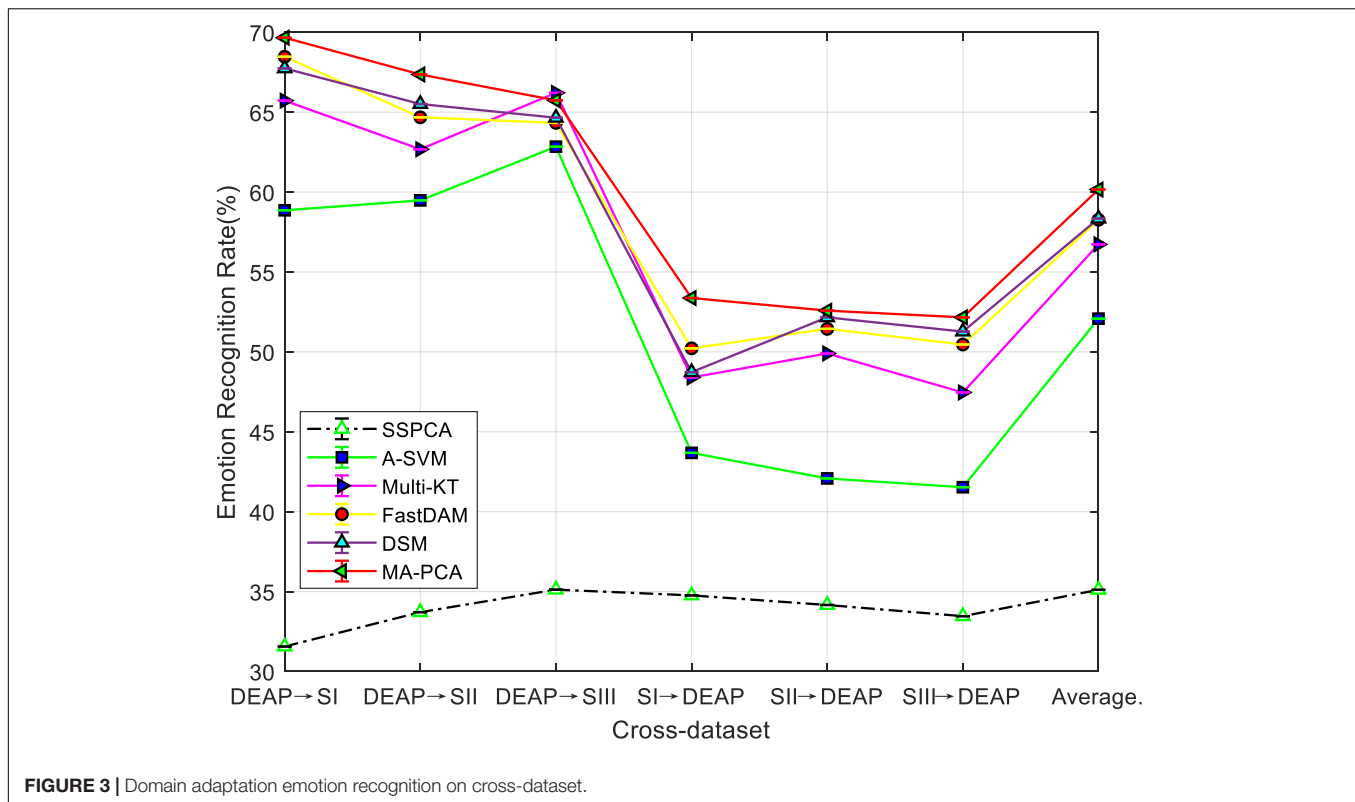
It is more challenging on emotion recognition when across datasets, with the differences in acquisition pathways, subjects’

characteristics, and behaviors. The previous experiments show the performance comparison of MA-PCA with other DA methods within the dataset (i.e., across subjects). This subsection further evaluates the robust effectiveness of MA-PCA when adapting across datasets. In this experimental scenario, multiple different protocols were constructed using different EEG devices and emotional stimuli by sampling the training and testing datasets separately. Therefore, six experimental settings, namely, DEAP → session I, DEAP → session II, DEAP → session III, session I → DEAP, session II → DEAP, and session III → DEAP, were set up to demonstrate that MA-PCA has robust effectiveness on emotion recognition with cross-dataset. For simplicity of expression, session I, session II, and session III in SEED are coded as SI, SII, and SIII, respectively, [for detailed experimental setup information, see the literature Tao et al. (2021)].

Performance Comparison

We aim to evaluate the performance of our method MA-PCA using the emotion recognition results on cross-dataset from DEAP and SEED. The experimental results are, respectively, plotted in **Figure 3**, which shows the average results for six possible combinations.

It can be seen from the bars in **Figure 3** that the no-adaptation method SSPCA has the worst performance than others in all cases, which witnesses the existence of distribution discrepancy between DEAP and one of the sessions from SEED. In this context, the importance of domain adaptation (DA) will be indispensable. This is justified by the observation in **Figure 3** that DA may reduce technical differences in cross-dataset applications, and our MA-PCA consistently outperformed other DA methods in most cases of the cross-dataset settings. A possible



reason for this may be that the clustering hypothesis with fuzzy entropy in our MA-PCA could weaken the impact of noise from these datasets. A clear phenomenon can be observed in **Figure 3**—due to the large distribution difference between the different datasets, the average recognition accuracy of all methods is correspondingly lower than the results obtained in **Figure 1** within the dataset.

Emotion Recognition With Multi-Source Prior

As reported in preceding works about domain adaptation learning, multiple-source domains can improve the adaptation performance to some extent by integrating multiple prior knowledge. Nevertheless, in concrete applications, multi-source adaptation also incurs another challenge, i.e., source scalability issue, since multi-source learning could lead to the so-called negative transfer problem. In this scenario, how to discriminately exploit multiple sources becomes a challenge worthy to be addressed in multi-source adaptation learning frameworks. To this end, we will explore in this part the different reliabilities of the prior sources in the emotion recognition task (Tao et al., 2019). We evaluate the performance of all baseline domain adaptation methods with multiple prior sources on the designed cross-dataset settings. The average accuracies of all methods are plotted in **Figure 4**, where A-SVM employs the average prior model.

When there exists a very large distribution discrepancy between different domain datasets, it is hard for A-SVM to eliminate the inter-domain distribution bias. Therefore, the results in **Figure 4** shows that A-SVM is inferior compared with the other multi-source adaptation methods in most settings.

A-SVM even has a downgraded performance tendency with the increase of source domains in some scenarios, which indicates the existence of “negative transfer” phenomenon in A-SVM. Another interesting observation from **Figure 4** is that all DA methods, except A-SVM, achieve more improvement by leveraging multiple-source knowledge than that by bridging only one source (i.e., cross-subject settings) when the number of source domains increases. This proves that it is beneficial to leverage multiple sources for boosting the recognition performance. Moreover, MA-PCA and DSM conquer others by touching on the top performance due to their designed weights for discriminately screening the optimal sources. Our method MA-PCA obtains more gains over DSM in some scenarios. This may be attributed to the adopted strategy in MA-PCA, which can efficiently select the most relevant source domains through optimal weighted multi-source adaptive regularization.

Adaptive Emotion Recognition With Deeply Extracted Features

In this subsection, we will particularly evaluate our method MA-PCA with deeply extracted features by comparing it with several recently proposed deep adaptation models on cross-dataset emotion recognition using multi-source settings.

In practical tasks, our method MA-PCA can be trained on the deeply transformed features of all domains, which follows the same setup with that in Zhu et al. (2017) and Zhou et al. (2018). Concretely, some pre-trained deep models (e.g., VGG16 and DAN, etc.) are first fine-tuned using the source domain, then the deep features can be extracted from EEG signals in both

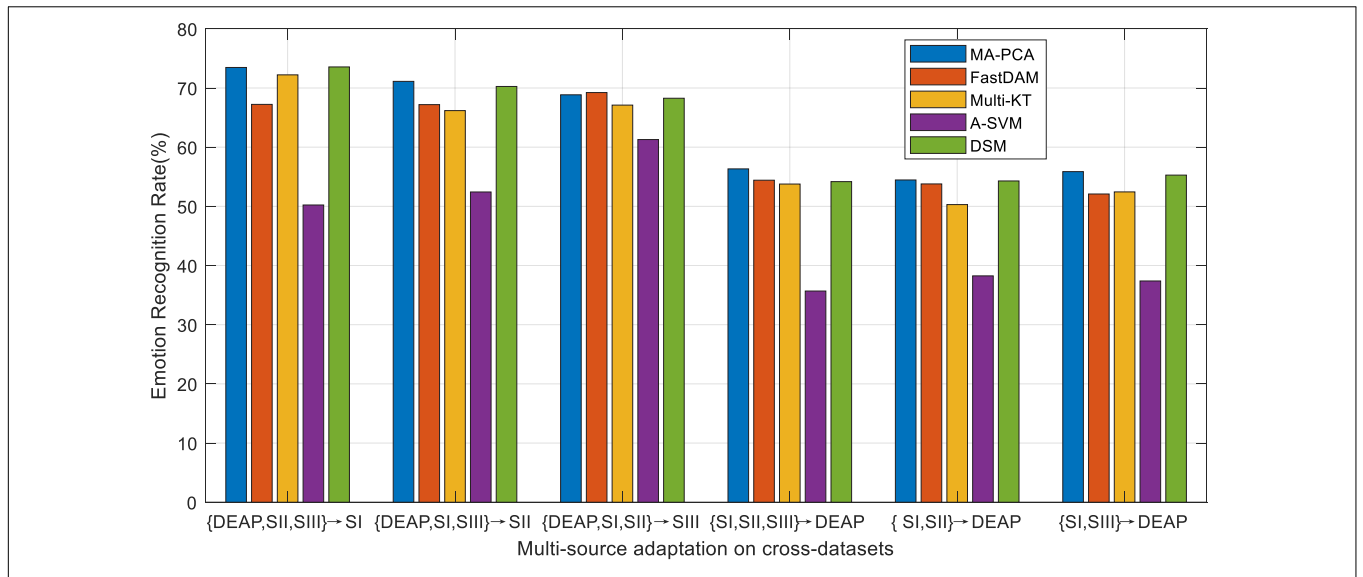


FIGURE 4 | Emotion recognition with multi-source adaptation settings.

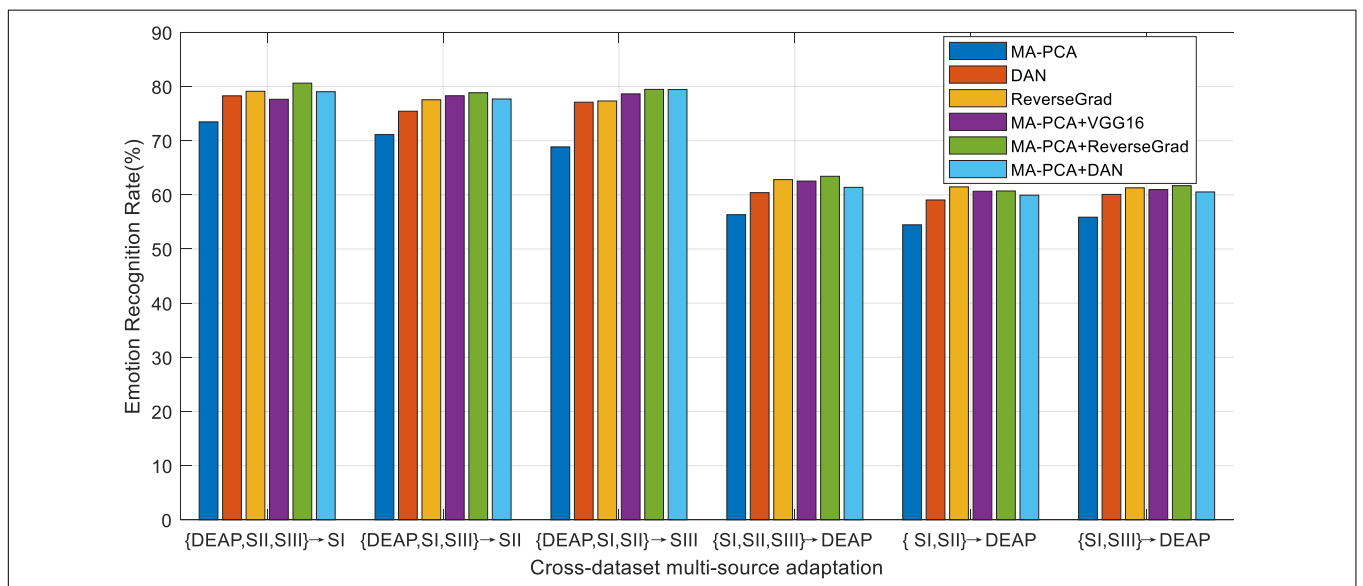


FIGURE 5 | Adaptive emotion recognition using deeply extracted features.

TABLE 2 | Multi-source adaptation emotion recognition accuracies of derived methods as well as MA-PCA.

Method	{DEAP,SII,SIII} → SI	{DEAP,SI,SIII} → SII	{DEAP,SI,SII} → SIII	{SI,SII,SIII} → DEAP	{SI,SII} → DEAP	{SI,SIII} → DEAP
MA-PCA_NTS	72.81	70.52	68.57	55.90	54.20	55.81
MA-PCA_NSS	71.30	70.05	65.87	53.17	53.77	55.43
MA-PCA_NOS	71.61	69.86	66.28	53.49	54.23	55.66
MA-PCA	73.47	71.12	68.85	56.33	54.46	55.87

Values in bold denote the best recognition rates.
 SI, session I; SII, session II; SIII, session III.

source and target domains with this CNN model, and finally the recognition model would be trained on these extracted features. In the context of our experiments, we denote our methods with VGG16 (respectively DAN) model as MA-PCA + VGG16 (MA-PCA + DAN respectively). As for DAN and ReverseGrad, we use their released source codes to fine-tune the pre-trained models by, respectively, using the pre-tuned parameters in their works (Ganin and Lempitsky, 2015; Long et al., 2015). Note that these deep adaptation methods typically aim to learn domain-invariant representations. Differently from the deep adaptation frameworks, our proposed method explores to learn a domain-invariant recognition model with strong generalization ability from the source domain to the target one.

We plot the mean results of all methods in **Figure 5**, from which we can observe that our deep adaptation method MA-PCA + VGG16 and MA-PCA + DAN significantly outperform MA-PCA. This indicates the advantage of deep features due to its robust feature representation. Furthermore, MA-PCA + DAN also obtains comparable recognition performance with respect to other deep adaptation methods. This may be attributed to the classification-level constraint in MA-PCA, where most of the source discriminative structures are expected to be preserved by the guidance of target classification. In some cases, shown in **Figure 5**, MA-PCA + VGG16 or MA-PCA + DAN even achieves the top-one performance compared with other deep adaptation frameworks. This phenomenon shows that the proposed MA-PCA can become an effective surrogate to the deep adaptation model by just exploiting the deep features extracted from any one of the state-of-the-art deep models.

Ablation Study

In our method MA-PCA, there exist several hyper-parameters needed to be tuned. These hyper-parameters are mainly used to trade off different components of the proposed framework. We therefore, respectively, set these parameters into their extreme values to explore the importance of each component in MA-PCA. To this end, we set $S = I$ to denote MA-PCA without target domain divergence information by MA-PCA_NTS and set $\gamma_i = 1/q$ and $\gamma_i = \frac{\exp(-Dist(X_i^s, X))}{\sum_i \exp(-Dist(X_i^s, X))}$ ($i = 1, 2, \dots, q$) to, respectively, denote by MA-PCA_NSS and MA-PCA_NOS the case where MA-PCA weights each source model by mean components and measures its distance from the target domain, respectively.

The performance of these derived methods is evaluated on cross-dataset recognition tasks, and the performance results are shown in **Table 2**. It is easy to see from **Table 2** that the performance of all derived methods is more or less degraded, and the performance of the MA-PCA_NTS method without target divergence constraints is slightly degraded. The performance of the MA-PCA_NSS method with an average weight on the source domain model decreases significantly; The performance of MA-PCA_NOS with the distance-weighted method for the source domain model is better than that of the MA-PCA_NSS method. However, the performance of these two derived methods is weaker than that of the MA-PCA optimization weighting method for the source domain model, which indicates that the proposed optimization mechanism

for source domain model selection is effective. In addition, an interesting observation is that the overall recognition accuracy is below 60% when the multi-source domains are all from the SEED dataset. However, when the multi-source domain has the DEAP dataset, the overall recognition accuracy of all methods is close to 70%, even higher than 70%. It indicates that the diversity of source domains can improve the robustness and generalization of MA-PCA and its derived methods.

CONCLUSION

To deal with cross-subject/dataset EEG-based emotion recognition task, we proposed a multi-model adaptation method with possibilistic clustering assumption, i.e., MA-PCA, by exploiting the knowledge of the correlation between the source and target domains in the objective function. It suppresses the influence of noise/abnormal data and weakens the impact of model performance caused by the different distributions of training data and test data (i.e., source domain and target domain). In MA-PCA, the fuzzy entropy regularization term is used to weaken the influence of noisy data, and multi-domain adaptation learning method is used to establish a robust classification model to weaken the influence of different data distributions. The comprehensive experiments performed on two public datasets verify the effectiveness of MA-PCA in dealing with cross-subject/dataset emotion recognition. In most scenarios, our MA-PCA (or MA-PCA-VGG16/DAN) obtains the best results or comparable performance with respect to several representative baselines. Since the implementation of MA-PCA algorithm needs an iterative optimization procedure, how to improve the efficiency of MA-PCA and seek a more efficient algorithm would be an issue worthy of further study in our future research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the research institution special program of Ningbo Polytechnic under Grant No. NZ21JG006 and Open Project of Key Laboratory of 3D Printing Equipment and Manufacturing in Colleges and Universities of Fujian Province under Grant No. 3DPM1901.

REFERENCES

- Atkeson, C.-G., Moore, A.-W., and Schaal, S. (1997). Locally weighted learning. *Artif. Intell. Rev.* 11, 11–73. doi: 10.1007/978-94-017-2053-3_2
- Belkin, M., and Niyogi, P. (2001). “Laplacian eigenmaps and spectral techniques for embedding and clustering,” in *Proceedings of the 14th International Conference on Neural Information Processing Systems*, Cambridge, MA) doi: 10.1162/0899766041732396
- Belkin, M., Niyogi, P., and Sindhvani, V. (2006). Manifold regularization: a geometric framework for learning from examples. *J. Mach. Learn. Res.* 7, 2399–2434.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.
- Bottou, L., and Vapnik, V. (1992). Local learning algorithms. *Neur. Comput.* 4, 888–900. doi: 10.1162/neco.1992.4.6.888
- Bruzzone, L., and Marconini, 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
- Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., and Li, Y. (2016). Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* 79, 205–214. doi: 10.1016/j.combiomed.2016.10.019
- Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014.
- Chapelle, O. (2006). *Scholkopf Bernhard, and Zien Alexander, Semi-Supervised Learning*. Cambridge, MA: MIT Press.
- Chu, W. S., De la Torre, F., and Cohn, J. F. (2017). Selective transfer machine for personalized facial action unit detection. *IEEE Transact. Patt. Anal. Mach. Intellig.* 39, 529–545. doi: 10.1109/CVPR.2013.451
- Dan, Y., Tao, J., Fu, 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
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194.
- Duan, L., Tsang, I. W., and Xu, D. (2012a). Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 465–479. doi: 10.1109/TPAMI.2011.114
- Duan, L., Dong, X., and Shih-Fu, C. (2012b). “Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach,” in *Proceedings of the CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, (New York, NY: IEEE), 1338–1345.
- Ganin, Y., and Lempitsky, V. (2015). “Unsupervised domain adaptation by back propagation,” in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, 1180–1189. doi: 10.1007/s10916-011-9759-1
- Gao, J., Wang, S.-T., and Deng, Z.-H. (2010). Global and local preserving based semi-supervised support vector machine. *Acta Electron. Sin.* 38, 1626–1633.
- Geng, B., Tao, D., Xu, C., Yang, L., and Hua, X. S. (2012). Ensemble manifold regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 1227–1233. doi: 10.1109/tpami.2012.57
- Hou, C., Nie, F., Tao, H., and Yi, D. (2017). Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Trans. Knowl. Data Eng.* 29, 1998–2011. doi: 10.1109/TKDE.2017.2681670
- Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transact. Affect. Comput.* 5, 327–339. doi: 10.1109/taffc.2014.2339834
- Karasuyama, M., and Mamitsuka, H. (2013). Multiple graph label propagation by sparse integration. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 1999–2012. doi: 10.1109/TNNLS.2013.2271327
- 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. *Computat. Math. Methods Med.* 2013:573734. doi: 10.1155/2013/573734
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signals. *IEEE Transact. Affect. Comput.* 3, 18–31. doi: 10.1109/t-affc.2011.15
- Kosko, B. (1986). Fuzzy entropy and conditioning. *Inform. Sci.* 40, 165–174. doi: 10.1016/0020-0255(86)90006-x
- Krishnapuram, R., and Keller, J.-M. (1993). A possibilistic approach to clustering [J]. *IEEE Trans. Fuzzy Syst.* 1, 98–110. doi: 10.1109/91.227387
- Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2018). Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Transact. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/tcds.2018.2826840
- Li, Y.-F., and Zhou, Z.-H. (2011). “Improving semi-supervised support vector machines through unlabeled instances selection,” in *Proceedings of the 25th AAAI Conference of Artificial Intelligence*, San Francisco, CA, 386–391. doi: 10.3233/BME-130935
- Liu, W., Wang, J., and Chang, S.-F. (2012). Robust and scalable graph-based semi-supervised learning. *Proc. IEEE* 100, 2624–2638.
- Long, M., 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, 97–105.
- Mansour, Y., Mohri, M., and Rostamizadeh, A. (2009). “Domain adaptation with multiple sources,” in *Proceedings of the Conference on Neural Information Processing Systems*, (Vancouver, BC: MIT Press), 1041–1048.
- 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. Interfac.* 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., Xu, D., Tsang, I., and Zhang, C. (2010). Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction. *IEEE Transact. Image Proces.* 19, 1921–1932. doi: 10.1109/TIP.2010.2044958
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359.
- Shi, X., Guo, Z., Lai, Z., Yang, Y., Bao, Z., Zhang, D., et al. (2015). A framework of joint graph embedding and sparse regression for dimensionality reduction. *IEEE Trans. Image Process.* 24, 1341–1355. doi: 10.1109/TIP.2015.2405474
- Tao, J., Dan, Y., and Zhou, D. (2021). Robust multi-source co-adaptation with adaptive loss minimization. *Sign. Proces. Image Commun.* 99:116455. doi: 10.1016/j.image.2021.116455
- Tao, J., Wen, S., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neur. Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009
- Tao, J., Wen, S., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowl. Based Syst.* 98, 76–94. doi: 10.1016/j.knsys.2016.01.021
- Tao, J., Zhou, D., Liu, F., and Zhu, B. (2019). Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Patt. Recogn.* 87, 296–316. doi: 10.1016/j.neunet.2019.02.007
- Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Patt. Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006
- Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi: 10.1126/science.290.5500.2319
- Tommasi, T., Orabona, F., and Caputo, B. (2014). Learning categories from few examples with multi model knowledge transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 928–941. doi: 10.1109/tpami.2013.197
- Tu, W., and Sun, S. (2013). Semi-supervised feature extraction for EEG classification[J]. *Patt. Anal. Appl. Paa* 16, 213–222. doi: 10.1007/s10044-012-0298-2
- Wang, Q.-W., Li, Y.-F., and Zhou, Z.-H. (2019). “Partial label learning with unlabeled data,” in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI’19)*, Macao, 3755–3761.
- Wang, Y., and Chen, S. (2013). Safety-aware semi-supervised classification[J]. *IEEE Trans. Neur. Netw. Learn. Syst.* 24, 1763–1772. doi: 10.1109/tnnls.2013.2263512
- Wang, Y.-Y., Chen, S.-C., and Zhou, Z.-H. (2012). New semi-supervised classification method based on modified cluster assumption. *IEEE Transact. Neur. Netw. Learn. Syst.* 23, 689–702. doi: 10.1109/TNNLS.2012.2186825
- Wu, M.-S., and Deng, X.-G. (2018). Semi-supervised pattern classification method based on Tri-DE-ELM. *Comp. Eng. Appl.* 54, 109–114.

- Xue, H., and Chen, S. (2007). "Alternative robust local embedding," in *Proceedings of International Conference on Wavelet Analysis Pattern Recognition*, Beijing, 591–596.
- Xue, H., Chen, S.-C., and Yang, Q. (2011). Structural regularized support vector machine: a framework for structural large margin classifier. *IEEE Transact. Neur. Netw.* 22, 573–587. doi: 10.1109/TNN.2011.2108315
- Yang, J., Yan, R., and Hauptmann, A. G. (2007). "Cross-domain video concept detection using adaptive svms," in *Proceedings of the 15th ACM International Conference on Multimedia*, (New York, NY: ACM), 188–197.
- Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019b). 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
- Zhang, Y., Chung, F. L., and Wang, S. (2019a). Takagi-sugeno-kang. Fuzzy systems with dynamic rule weights. *J. Intell. Fuzzy Syst.* 37, 8535–8550. doi: 10.1016/j.isatra.2017.10.012
- Zhang, Y., Li, J., Zhou, X., Zhou, T., Zhang, M., Ren, J., et al. (2019c). A view-reduction based multi-view TSK fuzzy system and its application for textile color classification. *J. Amb. Intellig. Human. Comput.* 9, 1–11. doi: 10.1007/s12652-019-01495-9
- 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. Imag. Health Inform.* 7, 1654–1659. doi: 10.1006/cbmr.1996.0023
- Zhang, Y., Wang, L., Wu, H., Geng, X., Yao, D., and Dong, J. (2016). A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation [J]. *J. Med. Imag. Health Inform.* 6, 1337–1344. doi: 10.1166/jmihi.2016.1923
- 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. Fus.* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002
- Zhao, Y., Liebgott, H., and Cachard, C. (2015). Comparison of the existing tool localization methods on two-dimensional ultrasound images and their tracking results. *Control Theor Appl.* 9, 1124–1134. doi: 10.1049/iet-cta.2014.0672
- Zhao, Y., Shen, Y., Bernard, A., Cachard, C., and Liebgott, H. (2016). Evaluation and comparison of current biopsy needle localization and tracking methods using 3d ultrasound. *Ultrasonics* 73, 206–220. doi: 10.1016/j.ultras.2016.09.006
- Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transact. Autono. Mental Dev.* 7, 162–175. doi: 10.1109/tamd.2015.2431497
- Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Transact. Affect. Comput.* 99:1. doi: 10.1109/taffc.2020.2994159
- Zhou, S.-H., Liu, X.-W., Zhu, C.-Z., Liu, Q., and Yin, J. (2014). "Spectral clustering-based local and global structure preservation for feature selection," in *Proceedings of the 2014 International Joint Conference on Neural Networks*, Beijing, 550–557.
- Zhou, X., Jin, K., Shang, Y., and Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial Im-ages. *IEEE Transact. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819
- Zhu, X.-J. (2008). *Semi-Supervised Learning Literature Survey [R]*. *Computer Science TR 1530*. Madison, WI: University of Wisconsin.
- Zhu, X.-J., and Goldberg, A. (2009). *Introduction to Semi-Supervised Learning*, Vol. 3. San Rafael, CA: Morgan & Claypool, 130.
- Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2017). automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Transact. Affect. Comput.* 9, 578–584. doi: 10.1109/TAFFC.2017.2650899
- Zu, B.-B., Xia, K.-W., Wen, L., Niu, W.-J., and Jiang, X.-Q. (2019). Semi-supervised classification application of remote sensing image based on block low rank images. *J. Front. Comput. Sci. Technol.* 13, 1217–1226.

Conflict of Interest: The 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.

Copyright © 2022 Dan, Tao and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.