2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)
Virtual Conference, May 4-6, 2021

Domain Adaptation for Cross-Subject Emotion Recognition by Subject Clustering

Jin Liu*, Xinke Shen*, Sen Song and Dan Zhang, *Member, IEEE*

2021 10th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-7281-4337-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/NER49283.2021.9441368

Abstract— **The high inter-subject variability in emotional EEG activities has posed great challenges for practical EEG-based affective computing applications. The recently popular domain adaptation strategy seemed to be a promising technique for addressing this issue, by minimizing the discrepancy of EEG data from different subjects. The present study proposed and implemented an extended Domain Adaptation method by introducing Subject Clustering (DASC). By clustering subjects based on the similarity of their emotion-specific EEG activities, the DASC method could make a flexible use of the available source domain information towards an optimized target domain application. Using the publicly available EEG dataset of DEAP, the DASC method achieved an average accuracy of 73.9±13.5% and 68.8±11.2% for binary classifications of the high or low levels of valence and arousal. Comparison with the state-of-the-art performance as well as the ablation experiments suggest the proposed DASC method as an effective extension to the conventional domain adaptation methods for EEG-based emotion recognition.**

I. INTRODUCTION

Affective computing aims to improve the computer's ability to understand and correctly respond to human's emotional state, which has developed rapidly [1] [2]. In recent years, more and more attention has been paid to the research of affective computing based on EEG [3] [4], for its portability and cost effectiveness as compared to other brain imaging techniques such as functional magnetic imaging (fMRI) and magnetoencephalography (MEG). Progress has been made mainly the last two decades, with promising performance towards practical use. For instance, using publicly available EEG dataset of DEAP [5], SEED [6] and DREAMER [7], etc., EEG-based emotional recognition accuracy of over 85% has been achieved.

However, most of the studies to date have focused on intra-subject classifications, with emerging efforts on cross-subject classification in recent years. Cross-subject emotion recognition is important for the generalization of affective computing by reducing the dependence on the data of a new user, which is expected to greatly improve the user experience of EEG-based affective computing systems. Nevertheless, it remains to be a challenging task due to large subject-to-subject variability of EEG responses, e.g. by dispositional emotional experiences [8]. Indeed, highly individualized EEG emotional responses were observed, resulting in overall lower cross-subject classification performances as compared to the intra-subject counterparts [4]. For instance, accuracies of ∼60% were reported for binary classification of high or low levels of arousal and valence by using the DEAP dataset [9]. In another study that has focused on emotional EEG responses during music listening, the performance of the model dropped by nearly 20% [10].

Aiming at improving cross-subject emotion recognition performances, efforts have been devoted to explore common and stable emotion-related EEG features that are invariant between subjects. Domain adaptation (DA) has been one of the most popular strategy, which seeks to minimize the discrepancy of source domain (training set) and target domain (testing set) by finding a latent, domain invariant subspace to project the data of two domains. Transfer component analysis (TCA) [11], Subspace alignment (SA) [12] and Maximum independence domain adaptation (MIDA) [13] are three representative DA methods. TCA attempts to mitigate the distribution mismatch by minimizing the Maximum Mean Discrepancy (MMD) in a reproducing kernel Hilbert space (RKHS). SA attempts to align the principal component analysis (PCA)-induced bases of the subspace of the source and the target domains. MIDA seeks to maximize the independence between the projected samples and their respective domain features measured by the Hilbert-Schmidt Independence Criterion (HSIC). As the results of [14] showed that, using the referred domain adaptation techniques on DEAP and SEED dataset can improve the positive-negative-neutral classification accuracy significantly by 7.25%-13.40% compared to the baseline accuracy where no domain adaptation technique was used. Other state-of-the-art studies [9] [4] [15] [16] which applied the feature extraction techniques such as variational mode decomposition (VMD), a pretrained CNN model, Sequential Backward Selection and flexible analytic wavelet transform (FAWT) to obtain the invariant feature sets across subjects, also improved the model performance on cross-subject emotion classification by 3.44%-20.91%.

However, incorporating all available data for domain adaptation may not be the optimal solution for cross-subject classification due to the large inter-subject variation. For instance, using the DEAP dataset, the cross-subject classification accuracy of two-class valence could be as low as 25% (see Fig. 2). Therefore, leveraging all source domains with brute force may undermine the potential performance improvement due to the possibly large discrepancy between

This work was supported by National Science Foundation of China (U1736220 and 61977041), and Tsinghua University Initiative Scientific Research Program (20197010009 and 20197010006).

*These authors contributed equally to this work

J. Liu, X. Shen and S. Song are with the Laboratory of Brain and Intelligence and Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China.

D. Zhang is with the Department of Psychology and the Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China (corresponding author D. Zhang, phone: +86-10-62796737, dzhang@tsinghua.edu.cn)

some source domains (e.g. data from some subjects) and the target domain, which is referred to as "negative transfer" [17]. To address this issue, it would be reasonable to perform a selection of the source domains prior to applying domain adaptation methods: only the source domains that are expected to have a "positive transfer" should be included. Such an idea has been indirectly supported by the psychological literature on individual differences that advocated the existence of different personality types [18], as well as recent affective computing studies suggesting a reliable link between personality and emotional EEG responses [19].

In the present study, we proposed an extended domain adaptation algorithm to reduce the impact of "negative transfer" by introducing subject clustering (Domain Adaptation with Subject Clustering, DASC). Cross-subject classification was based on subspace alignment (SA) method using the source domains with possibly "positive transfer" on the target domain. The publicly available dataset DEAP was used for validation. The obtained cross-subject emotion recognition performance suggests the proposed method as a promising technique towards practical EEG-based affective computing applications.

II. METHODS

The pipeline of the proposed DASC method is shown in Fig. 1. The method makes a flexible use of source domain information (i.e. the subjects' EEG data in the training set) for emotion recognition in the target domain by clustering the subjects in the training set according to the inter-subject similarity of their emotion-specific EEG activities. During the domain adaptation on the target (i.e. the test subject), only the source cluster that best matches the target is used, and optimal sources in this cluster with possibly "positive transfer" on the target are selected for the classification of emotional states of the target.

A. Source Clustering

The inter-subject similarity of their emotion-specific EEG activities was characterized by cross-subject classification accuracy. The cross-subject classification accuracy was computed by applying the classifier trained by one subject's EEG data to the EEG data of another subject, providing a straightforward yet comprehensive index for describing the similarity of the two datasets in their emotional EEG activity patterns as well as the separability among different EEG categories (high or low levels of valence or arousal, in the present study). Specifically, a linear discriminant analysis (LDA) classifier was trained using one subject's data for binary classification of valence or arousal, and applied to the data of the other subjects. For a total number of M subjects in the training set, M–1 cross-subject classification accuracies could be obtained per subject, representing the similarity between the subject and others. The M–1 cross-subject classification accuracies per subject were then considered as data feature for subject clustering. The k-means algorithm with squared Euclidean distance was employed, with the number of clusters determined by the Silhouette coefficient, which is a generic method to evaluate the clustering results (the higher value means the better clustering result). In this way, the source domains (i.e. the subjects in the training set) were separated into different source clusters, with similar subjects in the same source cluster.

After source clustering, we trained a 4-layer feedforward neural network on each source subject as the classifier for emotion recognition on the target subject. To tackle overfitting during training deep network, the source data augmentation was conducted. For each source, the subspace alignment (SA) [12] method was applied individually to align the data of other sources in the same cluster with the data of it. Subspace alignment as a kind of domain adaptation method aligns the bases of the subspace of the source domain X_S and the target domain X_T (i.e. two subjects' EEG data) through a linear transformation M to reduce the data discrepancy. The bases of the subspace of X_S and X_T are obtained by principal component analysis (PCA), denoted as V_S and V_T respectively. To align V_S with V_T, the desired M is defined as:

$$M = argmin_M(||V_S M - V_T||_F^2) \tag{1}$$

Where $||.||_F^2$ is the Frobenius norm and the closed-form solution of M is given by M = $V_S^\top V_T$. The source and target domain (X_S, X_T) can then be projected to the aligned subspaces (Z_S, Z_T), respectively, by Z_S = $X_S V_S V_S^\top V_T$ and Z_T = $X_T V_T$.

In this way, the data of each source subject were effectively augmented using the transformed data of similar source subjects with less discrepancy of cross-subject EEG data.

B. Cluster Selection

Based on the source clustering results, cluster selection was conducted to match the target with one source cluster, which was the basis of source selection. Compared with other clusters, the selected cluster (Cluster K in Fig. 1) should have the most number of source subjects whose emotional EEG activity patterns are similar to the target subject, therefore the classifiers in this cluster also achieve the most consistent results of cross-subject classification on the target. For each cluster, the proportion of major prediction result on each sample of the target was computed and averaged as the consistency of results in this cluster. The source cluster with the highest consistency of results was selected.

C. Source Selection

However, there may exist some sources in the selected cluster that produce "negative transfer" on the target, so the source selection based on the determined cluster is necessary. Source selection was conducted on a new calibration set A_L^C, which was obtained by projecting the data of each source in the selected cluster to the target by subspace alignment method. A_L^C resembled the target data with the labels of source data, so the performance of source classifier on A_L^C can be used to evaluate it on the target. We applied the classifiers in the selected cluster on A_L^C and located the top several classifiers with high accuracies. The corresponding

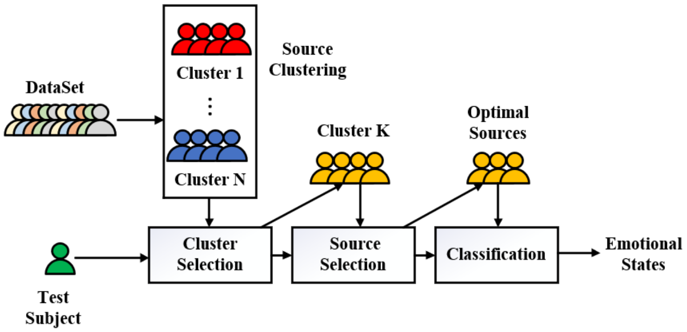

Fig. 1. The framework of the proposed DASC method

sources of the located classifiers were regarded as the sources with possibly "positive transfer" on the target and selected. For each selected source, we aligned the target with it by subspace alignment method and applied the classifier of the source to make cross-subject classification on the aligned target. Finally, the predictions from the classifiers of selected sources were ensembled with majority voting strategy as the emotional states of the target.

III. EXPERIMENTS

A. EEG Dataset

The proposed model was evaluated using the DEAP dataset. DEAP [5] is a well-known publicly available dataset for emotion recognition. This dataset contains 32-channel EEG recordings from 32 participants. Each participant watched 40 one-minute music videos with simultaneous EEG recordings at a sampling rate of 512 Hz. After watching each video, participants rated their subjective feelings of valence, arousal, dominance and liking, on a continuous scale from 1 to 9.

The preprocessed data provided by DEAP were used in the present study. The preprocessing pipeline was as follows: The raw data were first down-sampled to 128 Hz, with a bandpass filter of 4-45 Hz and then segmented into 60-s trials with the subtraction of a 3-s pre-trial baseline.

B. Feature Extraction

Differential entropy (DE) was extracted from the segmented EEG data as feature for classification, for its proven effectiveness in recent EEG-based affective computing studies [20]. DE is defined as [21]:

$$\begin{aligned} DE &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\right) dx \\ &= \frac{1}{2}\log\left(2\pi e\sigma^2\right) \end{aligned} \tag{2}$$

Where x is a one-channel EEG signal. The computed DE reflects the degree of "disorder" in EEG signals and is equivalent to the logarithm energy spectrum under the assumption that EEG signals follow the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ on each frequency band [21]. We computed DE on four classical frequency bands of each EEG segment: Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-30 Hz) and Gamma (30-45 Hz). The feature dimensions for the DE on each frequency band and the concatenated DE from all frequency bands were 32 and 128 respectively (32 channels×4 frequency bands).

C. Experiment Details

In this work, binary classifications of valence and arousal were conducted. The valence or arousal ratings larger than 5 were regarded as positive or active, while the ratings lower than 5 were taken as negative or passive respectively.

We adopted "one-subject-out" cross-validation tests and calculated the mean accuracy of 32 subjects to evaluate the performance of the proposed DASC method. Each time, one subject's data were excluded from the training set as the target domain. The data of each subject in the remaining training set were seen as an independent source domain. In source clustering, we compared the values of Silhouette coefficient under different number clusters (from 2 to 5), and divided the source domains into the number clusters with the highest value.

After source clustering and source data augmentation, a 4-layer feedforward neural network (the dimension of each layer was 32, 20, 10, 1) was trained on each source subject using the MATLAB Neural Network Toolbox as the source classifier. We ran 200 epochs for the network training using Bayesian-Regularization Algorithm with learning rate set to 0.001.

In source selection, to find the appropriate number of source domains, we compared the mean accuracies achieved with our model under different number of source domains (from 1 to 10), and selected the appropriate number sources for valence or arousal prediction.

The ablation experiments were also conducted to investigate the contribution of each key module in the proposed DASC method. We removed one module as source selection or cluster selection or source clustering in each ablation experiment. The baseline model directly used all source domains in training set for domain adaptation by subspace alignment and applied the linear discriminant analysis for classification on the target. It should be noted that without source clustering, the cluster selection was also removed.

IV. RESULTS

We first compared the performance of DASC method on DE from each frequency band (Theta, Alpha, Beta, Gamma) and the concatenated DE from all frequency bands. DE from Gamma band yielded the top performance out of others in both valence prediction (Theta: 66.7±15.8%, Alpha: 69.5±14.2%, Beta: 72.7±11.4%, Gamma: 73.9±13.5% and concatenated bands: 73.0±14.8%) and arousal prediction (Theta: 62.3±9.8%, Alpha: 64.5±12.1%, Beta: 68.5±10.8%, Gamma: 68.8±11.2% and concatenated bands: 65.9±13.7%), so we focused on Gamma band in the following analysis.

Fig. 2 shows the cross-subject classification accuracy matrix of Subject 02-32 (source domains) for valence, when Subject 01 was the target domain. The index of each row and each column represents one source domain and the corresponding LDA classifier, respectively. It should be noted that the indexes of source domains have been reordered according to the clusters for better visualization. The diagonal elements were the accuracies of intra-subject classification, which were obviously higher than the accuracies of cross-subject classification. The cluster number of two had the highest Silhouette Coefficient value (valence: 0.36, arousal: 0.59), compared with the number of cluster as three, four and five (valence: 0.24, 0.16 and 0.05. arousal: 0.46, 0.39 and 0.23), therefore the source domains were divided into two clusters by K-means algorithm. In Fig. 2, most subjects were well defined in two clusters, with higher cross-subject classification accuracies of intra-cluster than inter-cluster. Cluster 1 and Cluster 2 included 11 and 20 subjects (source domains) respectively.

Fig. 3 shows the mean accuracies of binary classification for valence and arousal achieved with different number of source domain. The peak accuracy of valence (73.9%) or arousal (68.8%) was achieved with the number of source domain as five or four, so five and four source domains were selected for the target domain in valence and arousal prediction respectively.

Table I reports the results of ablation experiments. For valence prediction, the improvements of each module on classification accuracy were varied from 5.8% (source selection) to 13.5% (source clustering). For arousal prediction, the improvements were varied from 2.1% (source selection) to 11.9% (source clustering). Compared with the baseline model, the total improvement of accuracy achieved by DASC method was up to 26.8% for valence and 15.6% for arousal.

For the "one-subject-out" cross-validation tests on DEAP by DASC method, we obtained the average accuracy as 73.9% (Std = 13.5%) of valence and as 68.8% (Std = 11.2%) of arousal. For valence prediction, the highest accuracy was 94.5% (Subject 01) and the lowest one was 43.1% (Subject 05). There were 21 subjects with accuracy over 70%. For arousal prediction, the highest accuracy was 90.3% (Subject 23) and the lowest one was 45.8% (Subject 29). The number of subject with accuracy over 70% was 18.

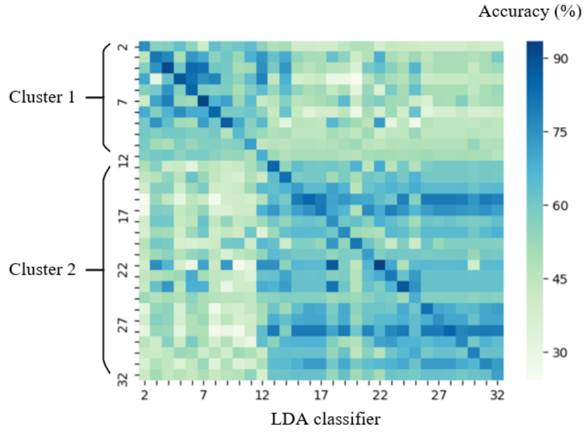

Fig. 2. The cross-subject classification accuracy matrix of Subject 02-32, when Subject 01 was the target (Pos-Neg valence classification).

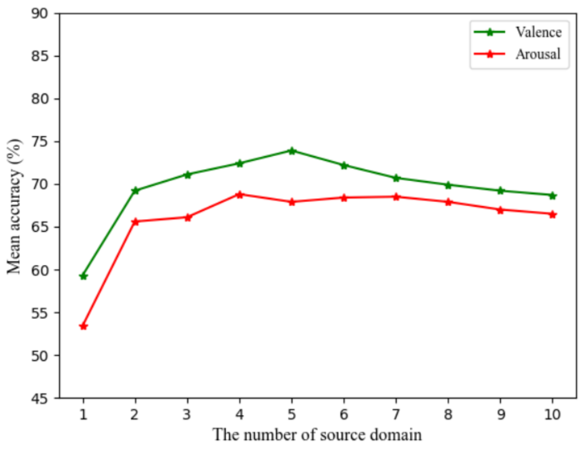

Fig. 3. Mean accuracy as a function of the number of source domain. The green and red line indicates the binary classification of valence and arousal, respectively.

V. DISCUSSION & CONCLUSION

The present study proposed and implemented an extended domain adaptation algorithm by introducing subject clustering for the purpose of cross-subject emotion recognition. The results of "one-subject-out" cross-validation tests based on the publicly available and widely used benchmarking DEAP dataset evaluated the feasibility of using DASC method for cross-subject EEG-based emotion classification in both valence (73.9±13.5%) and arousal (68.8±11.2%). The ablation experiments also verified the individual contribution of three modules (source clustering, cluster selection and source selection) in the proposed DASC method.

The results of Pos-Neg valence classification on DEAP dataset show better or comparable performance as compared with the reported results of seven representative methods in previous studies, ranging from traditional domain adaptation algorithms to the combination of various feature extraction techniques. As expected, the DASC method has significantly better performance than classical domain adaptation algorithms (MIDA [13] with 48.9%, TCA [11] with 47.2% and

TABLE I
ABLATION EXPERIMENTS FOR BINARY CLASSIFICATION OF VALENCE AND AROUSAL ON DEAP

Model	Accuracy of valence	Accuracy of Arousal
Proposed model	73.9	68.8
- Source selection	68.1	66.7
- Cluster selection	65.3	63.5
- Source clustering	60.4	56.9
Baseline model	47.1	53.2

SA [12] with 38.7%), which combined the data of all subjects in training set as a unified source domain. It should be noted that the inter-subject discrepancy of distributions in emotional EEG is so large that it is reasonable to regard the data of each subject as an independent domain. Compared with the state-of-the-art studies which focused on extracting common EEG features across subjects, our work achieved higher accuracy than a pretrained CNN [4] (72.8%), ST-SBSSVM [15] (72%) and VMD-DNN [9] (62.5%). The FAWT [16] produced slightly higher performance than our work as 79.9%, although it used the different EEG feature by wavelet transform. While previous results on the cross-subject classification of arousal have been limited, our ablation experiments demonstrated the effectiveness of our proposed method, as discussed below.

The ablation experiments illustrated the contribution of each module for the final classification performance. Most importantly, source clustering was shown to have the most prominent contribution (performance enhancement of 13.5% and 11.9% for valence and arousal, respectively), supporting our proposal of using subject clustering for cross-subject classification. Notably, the other modules also had their own contributions: cluster selection had the performance enhancement of 8.6% for valence and 5.3% for arousal, which was larger than the contribution of source selection (valence: 5.8%, arousal: 2.1%). It means that the result of cluster selection may have the greater effect than source selection on the classification of the target. As the data discrepancy inter cluster is larger than intra cluster, mismatched cluster selection (e.g. select Cluster 1 for the target who belongs to Cluster 2 in fact) may explain the extremely low accuracies in some subjects (valence: Subject 05 with 43.1%, arousal: Subject 29 with 45.8%).

The present work could be further improved in the following directions. First, better performance is expected with an increased number of subjects in the training set, for a more comprehensive coverage of possible subject sub-types. In addition, the psychological concept of personality and its quantitative measures may promote our understanding of individual differences in emotional responses and therefore could be used as an effective index in future subject clustering and matching. With these efforts, the proposed DASC method is expected to be a promising candidate towards practical EEG-based affective computing applications.

References

[1] R. W. Picard, *Affective Computing*. MIT press, 2000.

[2] X. Hu, J. Chen, F. Wang, and D. Zhang, "Ten challenges for eeg-based affective computing:," *Brain Science Advances*, vol. 5, no. 1, pp. 1–20, 2019.

[3] G. L. Ahern and G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: Eeg spectral analysis," *Neuropsychologia*, vol. 23, no. 6, pp. 745–755, 1985.

[4] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition," *Sensors*, vol. 20, no. 7, p. 2034, Apr 2020. [Online]. Available: http://dx.doi.org/10.3390/s20072034

[5] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[6] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[7] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018.

[8] G. Zhao, Y. Ge, B. Shen, X. Wei, and H. Wang, "Emotion analysis for personality inference from eeg signals," *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 362–371, 2018.

[9] P. Pandey and K. Seeja, "Subject independent emotion recognition from eeg using vmd and deep learning," *Journal of King Saud University - Computer and Information Sciences*, 2019.

[10] P. Keelawat, N. Thammasan, B. Kijsirikul, and M. Numao, "Subject-independent emotion recognition during music listening based on eeg using deep convolutional neural networks," in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2019, pp. 21–26.

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[12] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 2960–2967.

[13] K. Yan, L. Kou, and D. Zhang, "Learning domain-invariant subspace using domain features and independence maximization." *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 48, no. 1, pp. 288–299, 2018.

[14] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Muller-Putz, "Domain adaptation techniques for eeg-based emotion recognition: A comparative study on two public datasets," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 1, pp. 85–94, 2019.

[15] F. Yang, X. Zhao, W. Jiang, P. Gao, and G. Liu, "Multi-method fusion of cross-subject emotion recognition based on high-dimensional eeg features." *Frontiers in Computational Neuroscience*, vol. 13, p. 53, 2019.

[16] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2266–2274, 2019.

[17] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1855–1862.

[18] M. Gerlach, B. Farb, W. Revelle, and L. A. N. Amaral, "A robust data-driven approach identifies four personality types across four large data sets," *Nature Human Behaviour*, vol. 2, no. 10, pp. 735–742, 2018.

[19] W. Li, X. Hu, X. Long, L. Tang, J. Chen, F. Wang, and D. Zhang, "Eeg responses to emotional videos can quantitatively predict big-five personality traits," *Neurocomputing*, vol. 415, pp. 368–381, 2020.

[20] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 81–84.

[21] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for eeg-based vigilance estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2013, 2013, pp. 6627–6630.