Enhanced Subspace Alignment with Clustering and Weighting for Cross-Subject Multi-Session EEG-based Emotion Recognition

Mohsen Shirkarami
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
m.shirkarami@gmail.com

Hoda Mohammadzade*
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
hoda@sharif.edu

***Abstract*—The non-stationary nature of brain activity signals and their many inter-subject differences have created many challenges in the practical applications of electroencephalogram (EEG)-based emotion recognition, such as brain-computer interfaces. In such a way, using traditional classifiers in classifying these signals would significantly decrease accuracy when applying the classifier to a new subject. Domain Adaptation methods seem to be an effective way to solve this problem by minimizing the difference between the EEG signals of different subjects. But in the basic techniques for domain adaptation, looking at all subjects' data in the same look causes the loss of a part of the potential power of these methods. The present study proposed an extended version of the Subspace Alignment method for Domain Adaptation utilizing the clustering of source subjects. This method can make an effective choice between the available data from the source domain by clustering the source subjects based on the behavioral similarity of their EEG signals. This effective selection among the subjects and the idea of purposeful data weighting have caused the final model to represent the space of the target subject more effectively. To confirm the feasibility of the proposed method, we have conducted experiments on the SEED dataset, which is widely used in the emotion recognition EEG-based task, and it achieved an accuracy of 84.3%. The results show that the presented model has better classification accuracy than several state-of-the-art models.**

Keywords—Emotion Recognition, Multi-channel EEG, Domain Adaptation, Clustering, Subspace Alignment.

I. Introduction

Emotion plays an essential role in human behavior, decision-making, reasoning, and social relations [1]. Today, with the increasing development of brain-computer interfaces, emotion is an important factor in the interaction between humans and machines [2][3]. In order to make a decision to perform a task in the computer-brain interface systems and the need to know about a person's emotions, the problem of emotion recognition has become very important. Signals such as facial expressions, one's voice, or body movements can be used to detect emotions. However, despite the ease of investigation, these signals cannot be generalized to everyone. Because due to some differences in the appearance and behavior of humans, it is possible to make a mistake in recognizing emotions from these signals. For example, in some people with a mental disorder, the facial expression is very different from the person's emotional state, making it difficult for the doctor to monitor the person's treatment state [1].

Today, with the advancement of signal recording and medical imaging methods, as well as the use of new machine learning tools in classification, the problem of emotion recognition is done more accurately and differently. In recent years, it has been common to use signals of physiological origins, such as Electroencephalogram (EEG), Electrocardiogram (ECG), and Functional Magnetic Resonance Imaging (FMRI) signals to detect emotion [4][5]. In the meantime, EEG signals are considered a better option compared to other brain imaging methods due to their ease and economic advantages in recording, non-invasiveness, high temporal resolution, and portability [6][7][8][9][10].

In classification problems using traditional classifiers, the training and test sets are assumed to be generated from the same probability distribution. Nevertheless, in the case of brain signals, due to the non-stationary nature of these signals, this assumption is not correct, and the statistical characteristics of these signals are different from one subject to another and even at different times from a specific subject [11]. Therefore, if the classifier is trained using the available data from several subjects, the classification accuracy will drop by applying this classifier to another subject. The solution to this problem is Transfer Learning, or more specifically, Domain Adaptation (DA). In [11], the efficacy of classic domain adaptation methods, including Subspace Alignment (SA) [12], Transfer Component Analysis (TCA) [13], and Maximum Independence Domain Adaptation (MIDA) [14], is explored on the SEED and DEAP datasets. For example, SA attempts to align the Principal Component Analysis (PCA)-induced bases of the subspace of the source and the target domains. The results [11] indicate that using domain adaptation can improve the accuracy significantly by 7.25% - 13.40% compared to the baseline accuracy where no domain adaption technique is used. Other state-of-the-art studies [1][15][16][17] proposed different approaches to perform domain adaptation, utilizing deep neural network structures, and have achieved good results in improving classification accuracy.

Due to the great variety among the subjects, blindly using the data of all the subjects for classification does not seem to be an optimal solution. Because by more closely examining the accuracy of inter-subject classification, sometimes we come across very low accuracies in a classifier trained with one subject's data applied to another subject. For this reason, looking at the data of all subjects in the same way, due to the inherent difference between the subjects and, more precisely, the different probability distributions of each of them, can lead to the loss of the potential improvement of accuracy by DA.

* Corresponding Author

As a result, the final model cannot optimally represent the target subject subspace.

In [18], a method is proposed for emotion recognition by clustering subjects prior to applying DA methods. They implemented it on the DEEP dataset, which was able to improve the classification result compared to the case in which no adaptation was used, as well as the original case of subspace alignment.

In this paper, inspired by the work in [18], we extend the application of the method to multi-session EEG signals. In this selective look at the subjects, only those most similar to the target subject in emotional patterns are used. Also, during this process, several solutions are proposed and investigated to increase the accuracy of the final model. The SEED dataset, recorded in three sessions, will be used to test the feasibility of our proposed method.

The rest of this paper is organized as follows. Section II introduces the proposed methodology. Section III presents the dataset used in the experiment and other experimental details of the method. Section IV gives the experimental results. Section V concludes our study.

II. METHODOLOGY

In this section, we will introduce the proposed method in detail, as shown in Fig. 1. The steps are as follows:

A. Terminology

The *source subject(s)* refer to the subject(s) for which labeled data has been collected and used to train a model. The *target subject* is the subject to which we aim to adapt or generalize the model. In a multi-session framework, the *source session(s)* and *target session* can be defined similarly, representing the session(s) from which we have collected labeled data and used to train our model and the session to which we want to generalize our model, respectively.

B. Source Subjects Clustering

The aim of clustering is to put similar sources in the same clusters. Therefore, we need to define a similarity metric for clustering the source subjects. To do this, we can use cross-subject classification accuracy as a metric. We train a Linear Discriminant Analysis (LDA) classifier using each source and test it on other sources, as illustrated in Fig. 2.

We put the classification results in a matrix so that the elements in each row are the accuracies of applying a classifier trained on one subject and tested on other subjects. We use this matrix as the features for clustering the subjects (each row represents the features for one subject). The K-means method is used for clustering, and the optimum number of clusters is determined by the Silhouette score [19]. We define a range for the potential number of clusters and select the number which gives the maximum Silhouette score as the optimal one. For determining the upper bound of this range, it is crucial to consider that selecting a number greater than a certain threshold value may lead to the formation of clusters that contain only one source. The outcome can be considered inappropriate or undesirable. Therefore, when setting parameters, we should be cautious about choosing the upper bound value. Careful consideration and possibly experimentation with different threshold values are necessary

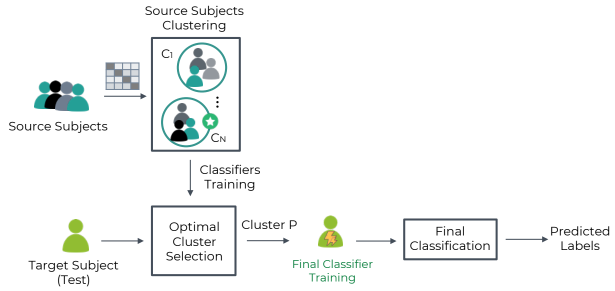

Fig. 1. The framework of the proposed method.

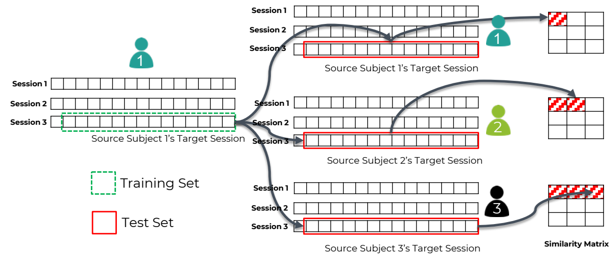

Fig. 2. The formation of the similarity matrix. Each row and each single square represent one recording session and one trial, respectively. The element at position (i, j) of the similarity matrix is derived from the classification accuracy obtained by training an LDA model on the data from subject i and testing it on subject j. In this simplified illustration, to help better understanding, this was done for only three subjects, but in the case of the SEED dataset, the number of source subjects will be 14, and as a result, the dimensions of the similarity matrix will be 14×14.

to achieve appropriate and insightful results from the clustering.

C. Classifiers Training

After clustering the sources, we must train one classifier for every source subject in each cluster. We will use these classifiers in the optimal cluster selection step. For each subject, we use data augmentation in its training process to avoid overfitting and then train a Multi-Layer Perceptron (MLP) using the augmented training set. Data augmentation is achieved by utilizing the Subspace Alignment (SA) method, which involves the alignment of the bases of the subspaces of the source domain (X_S) and the target domain (X_T) (i.e. EEG data from two different subjects). This alignment is carried out through a linear transformation denoted as M, with the objective of minimizing data discrepancies. The bases of the subspaces for X_S and X_T, referred to as V_S and V_T, respectively, are acquired by applying Principal Component Analysis (PCA). In this process, PCA is used to select d eigenvectors corresponding to the d largest eigenvalues, for each domain. To align V_S with V_T, the optimal M is defined as:

$$M = argmin_M(\|V_S M - V_T\|_F^2) \quad (1)$$

Here, $\|\cdot\|_F^2$ represents the Frobenius norm. It can be concluded that the optimal M is obtained as $M = V_S^{\top} V_T$. The source and target domains, X_S and X_T, are then projected into the aligned subspaces (Z_S and Z_T, respectively) using the following transformations: $Z_S = X_S V_S V_S^{\top} V_T$ and $Z_T = X_T V_T$ [12][18].

An aligned set is obtained by aligning the data of the subjects in the same cluster with the subject for which we intend to train the classifier. By combining this aligned set with the projected data of the subject itself in the subspace spanned by its principal components, the training set for that classifier is formed. The process is illustrated in Fig. 3, where sources 1, 2, and 3 are assumed to be clustered together. In this case, we need to train a total of three classifiers (MLP) to correspond to the number of subjects in the cluster. In the training of each classifier, to perform data augmentation, the data of two other subjects are aligned to the subspace of the subject for which we intend to train the classifier.

Fig. 4 depicts the procedure of forming the training set of each classifier in more detail. The data of each session of the subjects in the cluster and the source sessions of the subject for which we intend to train the classifier are separately aligned with the data of the target session of the subject. The training set of that classifier is built by aggregating these aligned data and the projection of the target session of the subject in the subspace spanned by its principal components.

D. *Weighting*

For the training set of each classifier to better represent the subspace of the desired subject, we perform weighting. In the stage of aggregating aligned data to form the training set of each classifier, we give weight to the available data from the target session of the subject corresponding to the classifier against the rest of the data. The number we choose as weight is three times the number of subjects in that cluster. The choice of the number three here is because the SEED dataset was recorded in three sessions. Clearly, for other datasets, this number will be equal to the number of recorded sessions. Fig. 5 provides a visual elucidation, offering a clear rationale for choosing weight.

The training set for each classifier is formed by aggregating three sets: firstly, the aligned data of the source subjects within the cluster, aligned with the target session of the desired subject, which is the subject we want to train a classifier corresponding to. Secondly, the aligned data of the desired subject's source sessions aligned with its target session data. Finally, the projection of the desired subject's target session data into the subspace spanned by its principal components. As a heuristic rule of thumb for the weighting factor, the product of the number of sessions times the number of subjects in that cluster is used, which is approximately equal to the number of aligned sets obtained from the first and second items constituting the classifier training set, explained above. By giving weight to the target session data and incorporating the aligned data from other subjects, the classifier can effectively learn from both subject-specific information and general patterns across subjects, enhancing its ability to generalize and perform well on new, unseen data.

E. *Optimal Cluster Selection*

In order to be able to achieve our goal in this method, i.e., optimal selection among subjects, we must select a cluster whose subjects are most similar to the target subject. To define this similarity and, as a result, to select the optimal cluster, we can have two different scenarios:

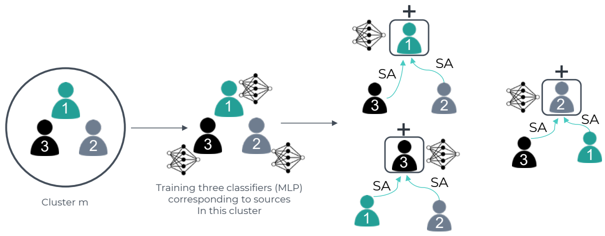

Fig. 3. Data augmentation in the formation of the training set of classifiers corresponding to the subjects in a cluster.

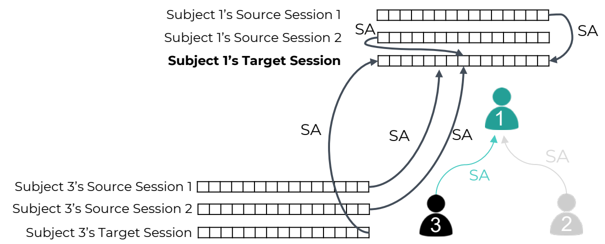

Fig. 4. Alignments during data augmentation related to a classifier. The data of different sessions from subjects in the same cluster with the subject corresponding to the classifier, as well as its own source sessions, are aligned separately with the target session of that subject and then aggregated. In the illustration, this is done for subject 3, which should be done similarly for subject 2.

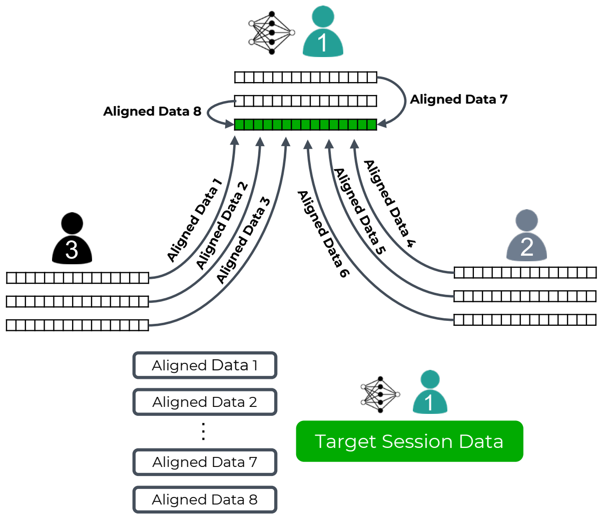

Fig. 5. Weighting in the process of formation of the training set of a classifier. Assuming we intend to train the classifier corresponding to subject 1, the aligned data form eight sets. In this illustrated example, we assign a weight of 9 to subject 1's target session data. This weight of 9 results from the product of three (the number of sessions) by three (the number of subjects in the cluster).

- Unsupervised Cluster Selection (Maximum Major Prediction Proportion): Assuming that we do not have any labeled data from the target subject, the optimal cluster is the cluster with the most source subjects having similar emotional behavior patterns to the target subject compared to other clusters. To achieve this goal, we choose the optimal cluster as one that has the highest agreement between their classifiers in the classification of samples of the target subject. For each cluster, the proportion of major prediction results for each sample of the target is calculated, and then on all samples, this ratio is averaged to represent the degree of agreement for that cluster; The cluster with the highest degree of agreement between its subjects (the cluster with subjects more unanimous) is selected as the optimal cluster [18].

- Supervised Cluster Selection: Assuming that we have some labeled data from the target subject, we have defined some criteria by which we can choose the optimal cluster:
    - Choosing by Clustering: The optimal cluster is determined by employing the clustering method on the labeled data available from the target subject. The cluster that best represents or aligns with the labeled data is chosen as the optimal cluster. This is done similarly to the source subjects clustering step and forming the similarity matrix; In such a way, the target subject is treated as a source subject, and the cluster to which the target subject is assigned is the optimal cluster.
    - Maximum Correctly Predicted Classifiers Proportion: For each cluster, according to the availability of labels of some target samples, the ratio of the number of classifiers, which predicted the correct label for each sample, is calculated and then averaged over all target samples. The optimal cluster is the cluster with the highest average ratio of classifiers that have predicted labels correctly among their classifiers.
    - Maximum Mean Accuracy: By applying the trained classifiers in each cluster to the labeled data available from the target subject, the optimal cluster is one that has the highest average accuracy of its classifiers.
    - Maximum Top Accuracy: By applying the classifiers to the labeled data from the target subject, we calculate the accuracy for each classifier. The optimal cluster is one containing the classifier that has the highest accuracy among all the classifiers in all the clusters.

F. Final Classification

After selecting the optimal cluster, we perform the final classification by training a new classifier. Prior to this, we apply a data augmentation process similar to what we did in the classifiers training step, with the difference that here, we intend to train the classifier corresponding to the target subject. The training set of this classifier consists of the aggregation of two sets: firstly, the aligned data of the source subjects in the optimal cluster aligned with the available samples of the target subject, and secondly, the projection of the target subject's available labeled data in the subspace spanned by its principal components. In this step, a Multi-Layer Perceptron (MLP) is trained with the same number of layers and characteristics of the layers as in the classifier training step. This classifier is the final classifier that by applying it to the test data, one can predict their labels and calculate the accuracy.

III. Experiments

A. Dataset

To test the performance of our method, we use the publicly available SEED dataset, one of the widely used datasets in EEG-based emotion recognition tasks. The SEED [20][21] dataset consists of EEG recordings from 15 participants, including both males and females. Each participant was exposed to visual stimuli designed to induce specific emotional states, including positive, negative, and neutral. Each type of emotion was assigned by five movie clips, each lasting approximately 4 minutes, which were well-edited to create coherent emotion elicitation and maximize emotional meanings. In total, there are 15 trials for each experiment. Each subject performed the experiment three times (sessions) with an interval of approximately one week between two successive recording sessions. The EEG signals were recorded by a 62-channel with the international 10-20 system at a sampling rate of 1000 Hz and then downsampled to 200 Hz.

B. Implementation Details

- Preprocessing: Due to the small amplitude of EEG signals (in the range of microvolts), these signals are strongly affected by noise. For this reason, feature extraction is used to have a better Signal-to-Noise Ratio (SNR) [1]. In various research, features such as DE [21], PSD, DASM, and RASM have been used in emotion recognition [21][22]. Among these features, Differential Entropy (DE) performs best in the problem of emotion recognition. In the SEED dataset, the feature data undergoes two initial steps: downsampling to a 200 Hz sampling rate and subsequent processing with a band-pass filter. Following this, the extraction of Differential Entropy (DE) features and other types of features takes place. The author applied conventional moving averages and Linear Dynamic Systems (LDS) to smooth all the features. For every sample, the features were calculated across five frequency bands (Delta: 1-3 Hz; Theta: 4-7 Hz; Alpha: 8-13 Hz; Beta: 14-30 Hz; Gamma: 31-50 Hz) for each channel. Previous studies show the performance of the Gamma band in emotion recognition tasks [18][23]. So, we use the DE features in the Gamma band for our model. According to [23], the best classification accuracy in each video clip is obtained for the samples in the interval of 140 seconds until the end of the signal. In this period, on average, the performance of classification of people's emotions using EEG signals was higher because the person was probably at the peak of emotional stimulation, and their emotions were expressed well. For this reason, in this study, we consider the samples in the mentioned range.

- Subspace Dimension Parameter of SA Method: The subspace dimension is assumed to be 40 in the Subspace Alignment method.

- Multi-layer Perceptron Details: In the classifier training stage, a multi-layer perceptron consisting of 3 hidden layers with 32, 20, and 10 neurons and an output layer with three neurons, as many as the classes of the SEED dataset, has been used. The selection of the number of layers, their dimensions, and other network hyperparameters is made so that there is no overfitting.

- Division of training/test sets: Leave-one-trial-out cross-validation is implemented.
- Number of clusters: In the clustering stage, according to the number of subjects in the SEED dataset, we calculated the Silhouette score for the number of clusters from 2 to 4. Then we considered the number for which the highest score was obtained as the optimal number of clusters.

IV. Results

The results are summarized in Table I. Among the different criteria defined for choosing the optimal cluster, Maximum Correctly Predicted Classifiers Proportion and Maximum Mean Accuracy have the best average accuracy. The interesting point is that these two methods have led to the selection of the same optimal clusters in each iteration. The classification results by selecting these criteria, by subject and target session, are shown in Fig. 6. Among the sessions, on average, the best performance corresponds to the target session 1. Also, on average, over all sessions, subjects 15, 13, and 12 have the best performance, and subjects 7, 5, and 1 have the worst performance among all subjects.

One of the steps of the method is weighting the data, which helps the model to better express the subspace of the subject for which we intend to train the classifier. In order to determine the effect of the weighting, we have once again performed the tests without weighting. Table II reports the results of the ablation experiment for each optimal cluster selection criterion. As we can see in Table II, weighting the data has improved the accuracy, regardless of the optimal cluster selection criteria.

Experimental results have indicated that our proposed method performs better than the basic domain adaptation methods (about an 18.6% improvement compared to the baseline Subspace Alignment method, according to the results of [11]). Also, compared to state-of-the-art methods, a comparable result has been reached. Our proposed method has performed better than the method of some other research, such as DAN [16] (with an accuracy of 79.93%), DCORAL [24] (with an accuracy of 76.89%), and DANN [25] (with an accuracy of 79.19%), which all have been evaluated on the SEED dataset.

Table I. AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF DIFFERENT CRITERIA FOR CHOOSING THE OPTIMAL CLUSTER

Criterion for choosing optimal cluster	Accuracy
Maximum Major Prediction Proportion	83.8±6
Choosing by Clustering	83.3±6
Maximum Correctly Predicted Classifiers Proportion	**84.3±6.8**
Maximum Mean Accuracy	**84.3±6.8**
Maximum Top Accuracy	83.9±6.6

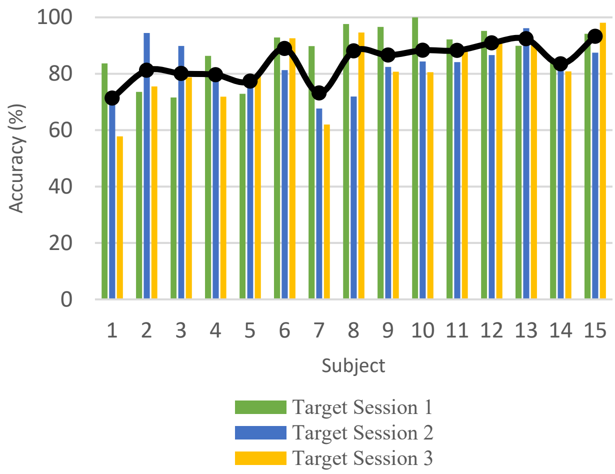

Fig. 6. Accuracy for each subject in each target session based on the Maximum Correctly Predicted Classifiers Proportion or Maximum Mean Accuracy as optimal cluster selection criterion. The mean accuracy of all sessions is plotted in the black continuous line.

Table II. AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF NON-WEIGHTED AND WEIGHTED APPROACHES FOR DIFFERENT OPTIMAL CLUSTER SELECTION CRITERIA

Criterion for choosing optimal cluster	Method	
	Non-weighted	*Weighted*
Maximum Major Prediction Proportion	81.1±7.4	**83.8±6**
Choosing by Clustering	81.8±6.7	**83.3±6**
Maximum Correctly Predicted Classifiers Proportion	82.9±7.3	**84.3±6.8**
Maximum Mean Accuracy	82.9±7.3	**84.3±6.8**
Maximum Top Accuracy	83.6±7.5	**83.9±6.6**

V. Conclusion

In this paper, we proposed an extended version of the Subspace Alignment method for emotion recognition using multi-session EEG signals. Using clustering and having a selective look at the subjects, we implemented and analyzed the optimal cluster selection criteria as well as the data weighting process. Experimental results have shown that our proposed method has achieved a better classification accuracy than several state-of-the-art methods.

The current study has the potential for further enhancement in the following directions. In the proposed method, we used Multi-Layer Perceptron as our classifiers. Constructing 2D frame sequences by considering the spatial position relationship across channels [26], the structure can be developed using Convolutional Neural Networks (CNNs). In selecting the samples, according to the previous research [23], we considered the samples placed within 140 seconds to the end of each trial as the samples that we were sure of their label. Designing a comprehensive solution that can be used to select samples in which the emotion of purpose is fully evoked and can be assigned a label with high confidence can also be considered as the future direction of this research.

REFERENCES

[1] Li, J., Chen, H. and Cai, T., 2020, December. Foit: Fast online instance transfer for improved eeg emotion recognition. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2618-2625). IEEE.

[2] Wang, F., Zhong, S.H., Peng, J., Jiang, J. and Liu, Y., 2018. Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II 24* (pp. 82-93). Springer International Publishing.

[3] Moon, S.E., Jang, S. and Lee, J.S., 2018, April. Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2556-2560). IEEE.

[4] Nikolova, D., Petkova, P., Manolova, A. and Georgieva, P., 2018. ECG-based emotion recognition: Overview of methods and applications. *ANNA'18; Advances in Neural Networks and Applications 2018*, pp.1-5.

[5] Wabnegger, A., Ille, R., Schwingenschuh, P., Katschnig-Winter, P., Kögl-Wallner, M., Wenzel, K. and Schienle, A., 2015. Facial emotion recognition in Parkinson's disease: An fMRI investigation. *PLoS One*, *10*(8), p.e0136110.

[6] Song, T., Zheng, W., Song, P. and Cui, Z., 2018. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *11*(3), pp.532-541.

[7] Ahern, G.L. and Schwartz, G.E., 1985. Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia*, *23*(6), pp.745-755.

[8] Cimtay, Y. and Ekmekcioglu, E., 2020. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors*, *20*(7), p.2034.

[9] Li, C., Tao, W., Cheng, J., Liu, Y. and Chen, X., 2019. Robust multichannel EEG compressed sensing in the presence of mixed noise. *IEEE Sensors Journal*, *19*(22), pp.10574-10583.

[10] Liu, Y.J., Yu, M., Zhao, G., Song, J., Ge, Y. and Shi, Y., 2017. Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Transactions on Affective Computing*, *9*(4), pp.550-562.

[11] 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 Transactions on Cognitive and Developmental Systems*, *11*(1), pp.85-94.

[12] Fernando, B., Habrard, A., Sebban, M. and Tuytelaars, T., 2013. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision* (pp. 2960-2967).

[13] Pan, S.J., Tsang, I.W., Kwok, J.T. and Yang, Q., 2010. Domain adaptation via transfer component analysis. *IEEE transactions on neural networks*, *22*(2), pp.199-210.

[14] Yan, K., Kou, L. and Zhang, D., 2017. Learning domain-invariant subspace using domain features and independence maximization. *IEEE transactions on cybernetics*, *48*(1), pp.288-299.

[15] Cao, J., He, X., Yang, C., Chen, S., Li, Z. and Wang, Z., 2022. Multi-source and multi-representation adaptation for cross-domain electroencephalography emotion recognition. *Frontiers in Psychology*, *12*, p.809459.

[16] Li, H., Jin, Y.M., Zheng, W.L. and Lu, B.L., 2018. Cross-subject emotion recognition using deep adaptation networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25* (pp. 403-413). Springer International Publishing.

[17] Chen, H., Jin, M., Li, Z., Fan, C., Li, J. and He, H., 2021. MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition. *Frontiers in Neuroscience*, *15*, p.778488.

[18] Liu, J., Shen, X., Song, S. and Zhang, D., 2021, May. Domain adaptation for cross-subject emotion recognition by subject clustering. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 904-908). IEEE.

[19] Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of computational and applied mathematics* 20 (1987): 53-65.

[20] Zheng, W.L. and Lu, B.L., 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3), pp.162-175.

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

[22] Saa, J.F.D. and Gutierrez, M.S., 2010, June. EEG signal classification using power spectral features and linear discriminant analysis: a brain computer interface application. *In Eighth Latin American and Caribbean Conference for Engineering and Technology* (pp. 1-7). Arequipa: LACCEI.

[23] Khazaei, E. and Mohammadzade, H., 2021. Temporal analysis of functional brain connectivity for eeg-based emotion recognition. *arXiv preprint arXiv*:2112.12380.

[24] Sun, B. and Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14 (pp. 443-450). *Springer International Publishing.*

[25] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V., 2016. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1), pp.2096-2030.

[26] Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A. and Chen, X., 2020. Emotion recognition from multi-channel EEG via deep forest. *IEEE Journal of Biomedical and Health Informatics*, 25(2), pp.453-464.