JOINT TEMPORAL CONVOLUTIONAL NETWORKS AND ADVERSARIAL DISCRIMINATIVE DOMAIN ADAPTATION FOR EEG-BASED CROSS-SUBJECT EMOTION RECOGNITION

Zhipeng He⋆ *Yongshi Zhong*⋆ *Jiahui Pan*⋆✉

⋆ School of Software, South China Normal University, Guangzhou, China

ABSTRACT

Cross-subject emotion recognition is one of the most challenging tasks in electroencephalogram (EEG)-based emotion recognition. To guarantee the constancy of feature representations across domains and to eliminate differences between domains, we explored the feasibility of combining temporal convolutional networks (TCNs) and adversarial discriminative domain adaptation (ADDA) algorithms in solving the problem of domain shift in EEG-based cross-subject emotion recognition. In light of EEG signals that have specific temporal properties, we chose the temporal model TCN as the feature encoder. To verify the validity of the proposed method, we conducted experiments on two public datasets: DEAP and DREAMER. The experimental results show that for the leave-one-subject-out evaluation, average accuracies of 64.33% (valence) and 63.25% (arousal) were obtained on the DEAP dataset, and average accuracies of 66.56% (valence) and 63.69% (arousal) were achieved on the DREAMER dataset. Extensive experiments demonstrate that our method for EEG-based cross-subject emotion recognition is effective.

Index Terms— Emotion recognition, EEG, Temporal convolutional network (TCN), Adversarial discriminative domain adaptation (ADDA)

1. INTRODUCTION

Emotion is crucial in human-computer interaction. Having the attribute of emotional recognition is an important component of machine humanization. The behavior and psychological performance in a certain context are referred to as one person's emotional state. We can recognize emotional states based on their behavior, language, expression and so on. Because electroencephalogram (EEG) is difficult to conceal, it can more objectively reveal reaction emotional states than behavioral modalities [1, 2].

✉ Corresponding Author

This work was supported by the National Natural Science Foundation of China under grant 62076103, the Guangdong Natural Science Foundation under grant 2019A1515011375, and the Key Realm R and D Program of Guangzhou under grant 202007030005.

However, due to the inherent diversity of psychological states, the fact that different people respond to the same stimuli to different degrees, etc., EEG signals are non-stationary and have individual variability[3]. An important challenge of implementing a high-fidelity emotion recognition system is that EEGs are quite different among subjects [4].

At present, many EEG-based studies [2, 5, 6] are subject-dependent experiments in which both the training set and the test set come from the same individual. This kind of research needs to spend much time on system calibration for each subject, which limits the application in real scenes [7]. Therefore, in terms of human-computer interface applications, EEG-based cross-subject emotion recognition is critical.

To achieve EEG-based cross-subject emotion recognition, many studies [8-10] have exploited the transferability of deep learning to seek shared feature spaces across diverse domains. However, neural networks, such as convolutional neural networks (CNNs), have trouble ensuring consistency of feature representation across different data domains in a high-dimensional space. Domain adaptation (DA) [11] is a promising strategy for addressing this problem, with the goal of removing the influence of distribution disparities between the source and target domains. A domain adaptation framework called adversarial discriminative domain adaptation (ADDA) [12] was originally proposed to achieve cross-domain image classification tasks and achieved good performance. Furthermore, Bai et al. [13] were the first to mention temporal convolutional networks (TCNs). Their research revealed that TCNs outperform recurrent neural networks (RNNs) in natural language processing and computer vision.

Motivated by the ability to minimize differences in different data domains of ADDA and the temporal series modeling capabilities of TCNs, we combine TCNs and ADDA as a novel method for EEG-based cross-subject emotion recognition. Our contributions include the following: (1) TCN is used to learn more transferable and differentiated intrinsic EEG features to eliminate domain transfer between different subjects and to learn dynamic temporal information. (2) The TCNs are incorporated into the ADDA, which combines untied weight sharing, discriminative modeling, and adversarial loss.

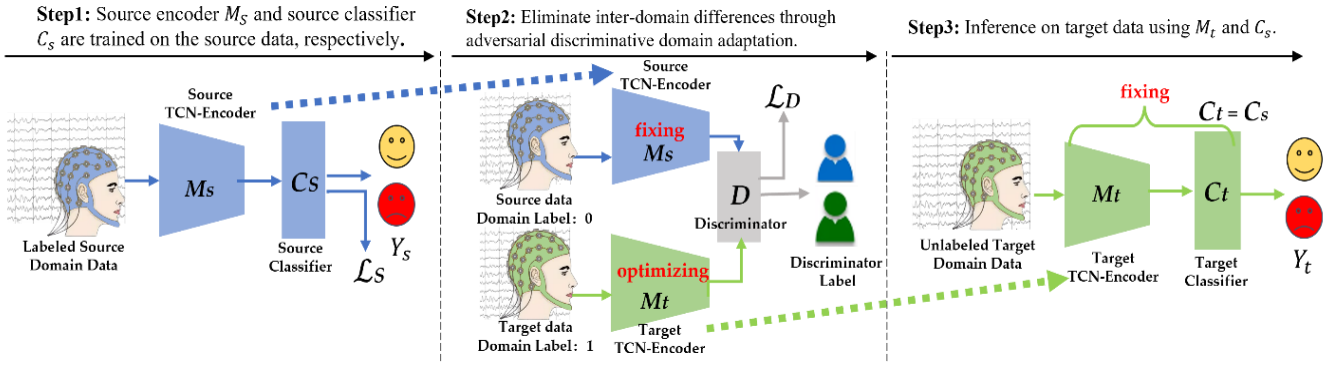

Fig. 1. The architecture of ADDA-TCN in this work

2. METHODS AND MATERIALS

2.1. Problem formulation and framework

It is assumed that the source domain D_S consists of labeled EEG X_S and marginal probability distribution $P_s(X_s)$, while the target domain D_t consists of unlabeled dataset and marginal probability distributions $P_t(X_t)$. The marginal probability distribution of the source domain is different from that of the target domain, that is, $P_t(X_t) \neq P_s(X_s)$.

The focus of our research is to minimize the distance between source mapping distribution $M_S(X_s)$ and target mapping distribution $M_t(X_t)$ by means of adversarial learning between source encoder M_S and target encoder M_t. In this paper, we adopt a three-step framework, as shown in Fig. 1. (1) The source encoder M_S and source classifier C_s are trained on the source data. M_S and C_s use labeled data from the source domain for pretraining. The source domain model obtains the characteristic attributes of emotional states through pretraining. (2) Eliminate interdomain differences through adversarial discriminative domain adaptation. The target encoder is then trained adversarially for the features extracted from M_t to have roughly the same distribution as the features extracted from M_S. (3) Inference on target data using M_t and C_s. Performance is verified using M_t and.C_s. For clarity, the algorithm is presented as an overview in pseudo code (Algorithm 1).

Algorithm 1: Training and Optimization Procedures of ADDA based on TCN-encoder

Inputs: labeled EEG data from the source domain $D_s = \{X_s, Y_s\}$; unlabeled EEG data from the target domain $D_t = \{X_t, \}$;

Outputs: Trained target TCN-based encoder M_t; Predicted target domain labels Y_t;

Initializeθ_s

for X_s, Y_s in $S_{(x,y)}$ do

  pretrain M_S and C_s based on loss below:

$min_{M_S,C}\mathcal{L}_{cls}(X_S, Y_S)$

$= -\mathbb{E}_{(x_s,y_s)\sim(X_s,Y_s)} \sum_{k=1}^{K} \mathbb{I}_{[k=y_s]} \log C(M_s(X_s))$

end for

Set $M_S = M_t$, $C_s = C_t$

repeat

  for iteration-number do

    updates θ_D, θ_t using Adam gradient descent

$\nabla_{\theta_D}(-\mathbb{E}_{x_s\sim X_S}[\log D(M_S(x_s))] - \mathbb{E}_{x_t\sim X_t}[\log(1 - D(M_t(x_t)))])$

$\nabla_{\theta_t}(-\mathbb{E}_{x_t\sim X_t}[\log D(M_t(x_t))])$

  end for

until convergence

return predicted target domain labels Y_t, trained target TCN-based encoder M_t

2.2. Feature encoder-TCN

In this paper, we choose the TCN as the feature encoder for ADDA. TCN can be used iteratively to capture long-term associations across multiple layers. The critical component is causal dilated convolution. Fig. 2 displays a graphical representation of casual dilated convolution. In this paper, it has a kernel size k of 3 and dilation factors d of 1, 2 and 4.

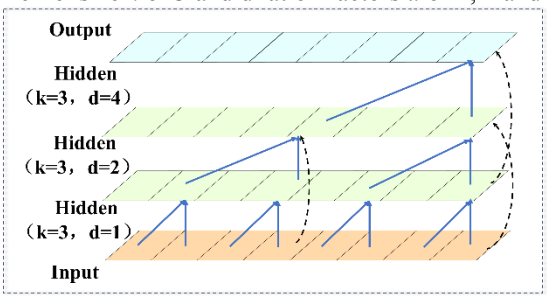

Fig. 2. Causal dilated convolution

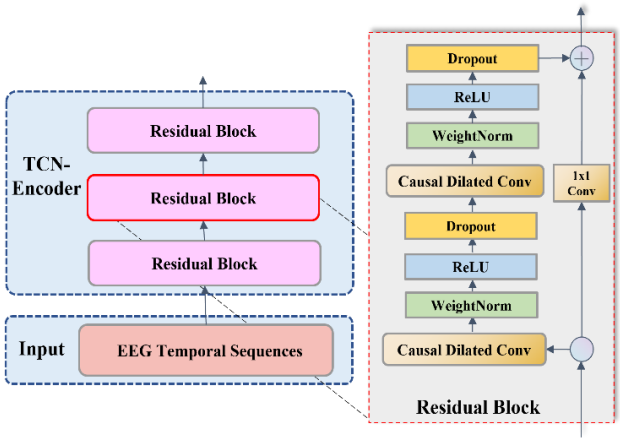

Fig. 3. The feature encoder structure in this work

A causal convolution is a class of convolutional models that handles sequential issues, where each node's operations can be considered with respect to the nodes that preceded it. When the number of convolutional layers grows, negative impacts such as gradient vanishing and overfitting often occur. The TCN incorporates a dilation convolution to address these issues. A dilated convolution expresses the size of the dilation layer using the dilation rate, which is useful because it can broaden the receptive field. The one-dimensional dilated convolution calculation is given as:

$$g(x) = \sum_{l=0}^{k-1} h(l) f(x - d \cdot l) \quad (1)$$

where $h(l)$ represents a filter length k, $f(*)$ serves as the input, $g(*)$ stands for the output of the dilated convolution, d is the dilation factor, and $x - d \cdot l$ denotes the pre-existing orientation of factor x.

As shown in Fig. 3, three residual blocks are utilized in the TCN in this work. There are two kinds of layers within the residual block: a dilated causal convolutional layer and a ReLU layer. The weights are normalized and applied to the convolution filter.

2.3. Adversarial discriminative adaptation

The trained source encoder M_S is used to initialize the target feature encoder M_t, and we freeze the source encoder M_S during the adversarial discriminative training phase. In the second stage, M_t is initialized to the weights of M_S, as shown in Fig. 1. Sharing the weights of M_S with M_t at the beginning makes the domain discriminator D unable to differentiate which data domain the features come from, thereby preventing the unbalanced training issue caused by overpowered discriminators in the early phase of adversarial discriminative learning.

With the weights of M_S and M_t being shared, the weights of M_t will be iteratively optimized in adversarial discriminative learning. The features of M_S and M_t are extracted from the source and target domains individually, and then the discriminator D differentiates which domain those extracted features come from. In addition, the discriminator outputs a label of 0 or 1 to denote that the feature is from the source or target domain.

Concretely, we first optimize $\mathcal{L}_{cls}$ on M_S and C by training with X_S, Y_S. The following optimization formulae are used.

$$min_{M_S,C} \mathcal{L}_{cls}(\mathrm{X}_S, \mathrm{Y}_S) = -\mathbb{E}_{(x_s \sim y_s) \sim (\mathrm{X}_s, \mathrm{Y}_s)} \sum_{k=1}^{K} \mathbb{I}_{[k=y_s]} \log C\left(M_S(\mathrm{X}_s)\right) \quad (2)$$

Since we choose to let M_S be fixed while training M_t, we could optimize $\mathcal{L}_{adv_D}$ and $\mathcal{L}_{adv_M}$ without reconsidering the previous target term. This corresponds to the following constraint-free optimization:

$$min_D \mathcal{L}_{adv_D}(\mathrm{X}_s, \mathrm{X}_t, M_s, M_t) = -\mathbb{E}_{x_s \sim X_s}\left[\log D\left(M_s(x_s)\right)\right] - \mathbb{E}_{x_t \sim X_t}\left[\log\left(1 - D\left(M_t(x_t)\right)\right)\right] \quad (3)$$

$$min_{M_t} \mathcal{L}_{adv_M}(\mathrm{X}_s, \mathrm{X}_t, D) = -\mathbb{E}_{x_t \sim X_t}\left[\log D\left(M_t(x_t)\right)\right] \quad (4)$$

where D represents the discriminator, which is applied to discriminate whether the extracted features belong to the source encoder M_S or the target encoder M_t. The difference in the distribution of features extracted from the source and target domains is reduced by adversarial training. Meanwhile, the discriminator D is confused and cannot identify the origin of the data. This adversarial discriminative learning approach results in a similar distribution of feature data extracted by M_t and M_S. During the training phase, the parameters of M_t are fixed before the discriminator D is optimized, and then the discriminator D is fixed to optimize M_t.

2.4. Materials for EEG-based emotion recognition

Dataset description. In this work, we used two publicly accessible datasets, DEAP[5] and DREAMER[6], for our analysis. The DEAP dataset includes thirty-two healthy people (sixteen males and sixteen females) who took part in the study. Each subject watched 40 one-minute music videos while the signal was being acquired. After watching each movie, subjects scored the self-assessment manikins (SAM) for arousal, valence, dominance, and liking on a continuous range of 1 to 9. The EEG data from fourteen EEG electrodes of twenty-three subjects make up the DREAMER dataset (fourteen males and nine females). The researchers used eighteen movie snippets to create the dataset, each of which evoked a different emotion. Each movie clip can last between 65 and 393 seconds. The SAM was used to assess subjectively the degree of arousal, valence, and dominance.

Data processing. Frequency features were extracted in four bands (theta, alpha, beta, and gamma). For each of the four frequency bands, we extracted differential entropy (DE) features per second. The following is the DE feature extraction formula:

$$h(x) - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2}\left(\log 2\pi e\sigma^2\right) \quad (5)$$

where x is a given segment of the EEG signal with Gaussian distribution $N(\mu, \sigma^2)$, σ is the variance of x, and e is the Euler constant.

Emotional categories are frequently dichotomized based on a statically fixed threshold in some studies [5, 6, 14, 15] on the DEAP and DREAMER datasets, and rated data can be directly generally separated into high and low arousal (or valence) states. However, the self-assessment is subjective[2]. To produce reasonable emotional labels for the high and low categories, we use the midpoint of k-means clustering [16] on each subject's subjective rated data, which is consistent with previous studies [17, 18].

3. EXPERIMENTS AND RESULTS

The DEAP and DREAMER datasets were used to test the effectiveness of leave-one-subject-out cross-validation for

cross-subject emotion classification. Specifically, we employed one subject as the target data, while other relevant subjects were used to construct the source data.

Comparison with baselines: Various domain adaptation methods have been developed to find the common feature space for source and target subjects, including transfer component analysis (TCA) [19], kernel principal component analysis (KPCA) [20], transudative support vector machines (TSVMs) [21], and transudative parameter transfer (TPT) [22]. Table 1 shows the mean accuracies and standard deviations of different baselines (TCA, KPCA, TSVM, TPT) and our method. The results demonstrate that our method can achieve better results.

Table 1. Average accuracy (standard deviation) for the baselines and our methods on the DREAMER and DEAP datasets. Significant differences between our work and other methods are indicated by sign (paired t test: ~nonsignificant, *p<0.05, **p<0.01).

Methods	DREAMER		DEAP	
	Valence	Arousal	Valence	Arousal
TSVM[21]	60.76* (9.77)	55.67** (12.07)	61.77* (8.93)	56.59** (11.98)
TPT [22]	59.22* (15.01)	61.89~ (13.18)	57.43* 14.54	54.76 * (12.48)
TCA [19]	55.85** (6.45)	54.37* (8.56)	56.23** 14.33	51.81** (15.03)
KPCA[20]	53.74** (8.47)	60.03* (11.24)	54.35** (10.22)	58.15* (14.96)
Our work	**66.56 (10.04)**	**63.69 (6.57)**	**64.33 (7.06)**	**63.25 (4.62)**

Comparison with prior works: To the best of our knowledge, there are few cross-subject studies on emotion recognition for DREAMER datasets. In this part, Table 2 shows the accuracy comparison of our work against existing studies on the DEAP dataset. As illustrated in Table 2, compared with previous studies, our method achieved higher average accuracy in the DEAP dataset, that is, 64.33% for valence and 63.25% for arousal.

Table 2. Accuracy (%) comparison with existing works on the DEAP dataset.

Study	Valence	Arousal
Rayatdoost et al.[8]	59.22	55.70
Lew et al.[23]	56.78	56.60
Seeja et al. [9]	61.50	58.50
Pandey et al.[24]	62.50	61.25
Li et al.[10]	64.20	58.40
Miguel et al. [15]	64.00	59.00
Our work	**64.33**	**63.25**

Ablation study: In this part, to further verify the effectiveness of each module in our method, we performed an ablation study on the DEAP dataset. We provide the results of TCN without adversarial discriminative training and use other classical models as feature encoders with adversarial discriminative training, which include MLP (multilayer perception), RNN, and CNN.

The ablation study's results are presented in Table 3. It is noteworthy that the performance improvements with our proposed method are more statistically significant than the counterpart models.

Table 3. Comparison results (%) in an ablation study for cross-subject experiments on the dataset DEAP (paired t test: *p<0.05, **p<0.01).

Method	Valence	Arousal
TCN	58.24±8.23*	53.03±6.25**
ADDA-MLP	56.75±10.35*	57.18±7.21*
ADDA-CNN	58.47±8.74*	56.84±7.80*
ADDA-RNN	60.55±7.68*	57.52±3.89*
ADDA-TCN	**64.33±7.06**	**63.25±4.62**

Parameter sensitivity analysis: Because our method is based on EEG signals that are made of temporal sequences, it is crucial to monitor emotion across numerous window lengths and shifts. We analyzed the parameter sensitivity on the DEAP. We selected window lengths of 3 s, 9 s, 6 s, and 12 s and altered the shifts between 1 and 2 s to decide on the suitable window lengths and shifts. According to the results in Table 4, our method performs well with the window-length and shift setup at 9 s and 1 s. Therefore, we chose the window-length and shift of 9 s and 1 s in this paper.

Table 4. Comparison of the accuracy (%) of our method at various window lengths and shifts.

Shift	Window-length	Valence	Arousal
1 s	3 s	57.24	58.78
	6 s	61.64	60.12
	9 s	**64.33**	**63.25**
	12 s	62.87	62.38
2 s	3 s	54.47	57.41
	6 s	58.02	62.74
	9 s	60.43	60.17
	12 s	59.33	60.04

4. CONCLUSION

In this paper, ADDA and TCNs are combined as a novel method for EEG-based cross-subject emotion recognition. The approach achieves competitive performance for cross-subject emotion recognition by using encoder-TCN and ADDA. Our method does not require the labels of new subjects and only unlabeled new subject data to accomplish the creation of an emotion recognition model for new subjects. Thus, it facilitates the generalization of EEG emotion recognition models to applications.

5. REFERENCES

[1] W. Zhang and Z. Yin, "EEG Feature Selection for Emotion Recognition Based on Cross-subject Recursive Feature Elimination," in *2020 39th Chinese Control Conference*, pp. 6256-6261, 2020.

[2] 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, pp. 162-175, 2015.

[3] H. Morioka, A Kanemura, J Hirayama, M Shikauchi, T Ogawa, S Ikeda, M Kawanabe, and S Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage,* vol. 111, pp. 167-178, 2015.

[4] Z. He, Z. Li, F. Yang, L. Wang, J. Li, C. Zhou, J. Pan, "Advances in multimodal emotion recognition based on brain–computer interfaces," *Brain Sci*. vol. 10, pp. 687, 2020.

[5] S. Koelstra, S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing,* vol. 3, pp. 18-31, 2011.

[6] 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 Health Informatics,* vol. 22, pp. 98-107, 2017.

[7] W.-L. Zheng and B.-L. Lu, "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the Twenty-fifth International Joint Conference on Artificial Intelligence*, pp. 2732-2738, 2016.

[8] S. Rayatdoost and M. Soleymani, "Cross-corpus EEG-based emotion recognition," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing*, pp. 11-17, 2018.

[9] P. Pandey and K. R. Seeja, "Subject independent emotion recognition system for people with facial deformity: an EEG based approach," *Journal of Ambient Intelligence and Humanized Computing,* vol. 12, pp. 2311–2320, 2021.

[10] X. Li, P. Zhang, D. Song, G. Yu, Y. Hou, and B. Hu, "EEG based emotion identification using unsupervised deep feature learning," in *Workshop on Neuro-Physiological Methods*, pp. 1-8, 2015.

[11] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *2015 5th International Conference on Information Science and Technology*, pp. 315-322, 2015.

[12] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167-7176, 2017.

[13] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271,* 2018.

[14] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications,* vol. 47, pp. 35-41, 2016.

[15] M. Arevalillo-Herráez, M. Cobos, S. Roger, and M. G. rcía-Pineda, "Combining Inter-Subject Modeling with a Subject-Based Data Transformation to Improve Affect Recognition from EEG Signals," *Sensors (Basel, Switzerland),* vol. 19, pp. 2999, 2019.

[16] J. A. Hartigan and M. A. Wong, "A K-Means Clustering Algorithm," *Applied Statistics,* vol. 28, pp. 100-108, 1979.

[17] L. Yang and J. Liu, "EEG-Based Emotion Recognition Using Temporal Convolutional Network," in *IEEE 8th Data Driven Control and Learning Systems Conference*, pp. 437-442, 2019.

[18] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers in Neurorobotics,* vol. 11, pp. 19-30, 2017.

[19] S. J. Pan, Tsang, I. W., Kwok, J. T., Yang, Q, "Domain Adaptation via Transfer Component Analysis," *IEEE Transactions on Neural Networks,* vol. 22, pp. 199-210, 2011.

[20] K. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks,* vol. 12, pp. 181-201, 2001.

[21] R. Collobert, F. H. Sinz, J. Weston, and L. Bottou, "Large Scale Transductive SVMs," *Journal of Machine Learning Research,* vol. 7, pp. 1687–1712, 2006.

[22] E. Sangineto, G. Zen, E. Ricci, and N. Sebe, "We are not All Equal: Personalizing Models for Facial Expression Analysis with Transductive Parameter Transfer," in *ACM International Conference onMultimedia*, pp. 357–366, 2014.

[23] W. Lew, D. Wang, K. Shylouskaya, Z. Zhang, and A. H. Tan, "EEG-based Emotion Recognition Using Spatial-Temporal Representation via Bi-GRU," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.116-121, 2020.

[24] P. Pandey and K. Seeja, "Subject independent emotion recognition from EEG using VMD and deep learning," *Journal of King Saud University-Computer Information Sciences,* vol. 11, pp.1-9, 2019.