

Cross-subject EEG Emotion Classification on Datasets with Limited Channels

Jinhao Chen^{1*†}, Zixin Chen^{2†}, Jianlin Pan^{3†}

¹*Zhejiang University, Hangzhou, China*

²*University of Nottingham Ningbo China, Ningbo, China*

³*Shandong Experimental High School, Jinan, China*

**Corresponding Author. Email: 3230103697@zju.edu.cn*

†These authors contributed equally to this work and should be considered as co-first author.

Abstract: To solve the challenge of achieving strong and accurate EEG datasets, while solving the domain shift problem, this research focused on integrating the Few-Label Adversarial Domain Adaptation (FLADA) method with attention mechanism to see if the addition of the attention mechanism improves the performance over FLADA alone on EEG datasets with a small number of channels (specifically, 14 channels). We first apply the FLADA method to the research by solving the "domain shift" problem. Next, we added channel-wise-attention in an attempt to improve the algorithm. However, we found that the addition of channel-wise-attention failed to improve accuracy over the FLADA method alone, and in some cases made it even worse.

Keywords: EEG, emotion recognition, fewer channel, FLADA, attention mechanism

1. Introduction

EEG-based emotion recognition utilizing Artificial Intelligence methods is a potential technique that enables machines to understand and respond to humans' emotions and needs, thus providing a more personalized experience and more mental health support for users. Right now, EEG-based emotion recognition is being widely researched for applications in many fields. For psychology, scholars have explored it to help resolve PTSD patients' problems [1]; In the field of human-computer interaction, it has been used for improving the user experience, enabling machines, games, or virtual environments to dynamically adjust based on the user's emotional reactions. For the automobile industry, scholars have analyzed emotion fluctuation using EEG and tried to build a car that can avoid car accidents.

However, studies focusing on EEG datasets with a small number of EEG electrode channels—for example, 14 channels—remain limited, posing challenges in achieving both strong cross-subject generalization and high accuracy. The main problem is the so-called 'domain shift' problem, where differences in individual brain responses lead to variations in data distribution, making it difficult to develop a universal model [2].

Researchers are beginning to conduct research to tackle this problem. There are three main ways that have been proposed to solve it. First, calibrating or aligning the physiological signals instead of using the original one is widely used. Mohammad and Nishida deducted the baseline to avoid interference thus provided a stable basis for the following training [3]. Second, establishing specific

models targeted to the factors that cause domain shift is also useful. Bailenson introduced a characteristic model and a gender model showing that they perform better [4]. Zeng et al. applied the robust feature selection strategy. Third, transfer learning can be applied to improve the performance of the target domain [5]. The basis of the above methods is the Few-Label Adversarial Domain Adaptation (FLADA) method, which solves the ‘domain shift’ problem with few-labeled data. As the application of FLADA methods on high-channel datasets has proven efficient, we attempt to apply it to low-channel datasets to address the domain shift problem.

To achieve high accuracy in low-channel datasets, the attention mechanism has been employed effectively in prior work to increase the accuracy of 14-channel DREAMER datasets [6]. As the attention mechanism alone does not explicitly handle domain shift, the strong performance of the FLADA method in this area serves as a valuable complement.

Therefore, this research aimed to assess whether the performance of deep learning model on low-channel datasets (specifically the DREAMER datasets with only 14 channels) could be improved by integrating the FLADA method with attention mechanism. We hypothesized that this might achieve higher accuracy while enhancing the model’s adaptability to cross-subject variations. After implementing the combined FLADA method with attention mechanism, we found that the combined technique did not in fact achieve higher accuracy than the FLADA model alone, and in some cases performed even more poorly.

2. Methodology

Our method is composed of four steps: a) Processing raw data from the DREAMER datasets, which is described in section 2.1, into appropriate format, and differential entropy (DE) is calculated for training the model. b) In order to select top-K domains, data calibration test is done on every source domain. c) Data from the source and target domain are used to form six groups, as shown in Fig. 2. d) Adversarial training to the feature extractor, classifier and discriminator on the six groups. The whole process of training is presented in Figure. 1.

2.1. Datasets

To evaluate the applicability of the FLADA method on the low-channel datasets as well as that of our optimized network, the DREAMER [7] and SEED [8] databases were used in our research.

In the DREAMER database, data of 23 participants were kept in the database. In the experiment, participants were required to watch film clips while recording their EEG signals throughout the process. And self-assessment was required to score the emotion from three dimensions after each film clip. In the published database, there are 36 14-channel EEG data segments with a sampling frequency of 128Hz considering that there are 18 film clips and both the baseline and stimuli signals were recorded.

In the SEED database, there are 15 participants, and each film clip were labeled as negative, positive or neutral. The published EEG has 62-channel data with a sampling frequency of 200Hz.

2.2. Data processing

Before training the model for emotion classification task, it is important to choose a proper window size [9], which is a significant parameter for calculating differential entropy feature. In our research, it is set to 1s, and the data is filtered into five bands, including alpha, beta, delta, theta and gamma waves, which are the common settings for implementing emotion classification tasks [10].

Then the DE-feature, which is proved more effective on the emotion recognition task by [11], of each band is computed to facilitate EEG emotion classification, using the probability density function of signal, which is represented by $p(x)$ in the following arithmetic formula:

$$DE = - \int_{-\infty}^{+\infty} p(x) \log p(x) dx \quad (1)$$

Then a linear dynamic system method, which can eliminate the impart of unrelated components while making full use of the characteristic emotion changes in time domain, is applied to smooth the DE feature [12].

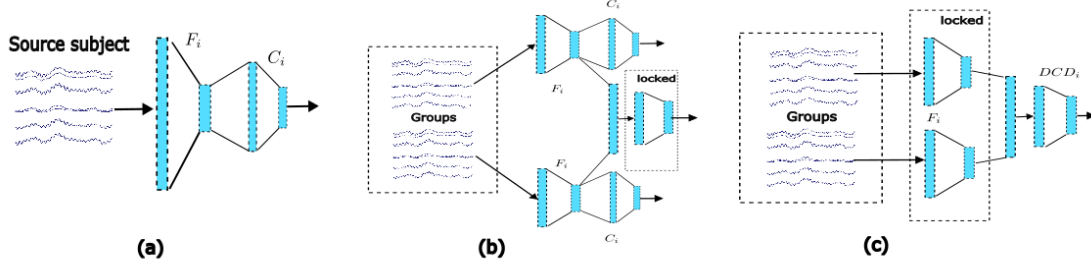


Figure 1: The structure of FLADA method. (a) Obtaining the feature extractor F_i and classifier C_i of each source model. (b) locking the weight of F_i model, the domain-class discriminator DCD_i model are trained for top-k source domain. (c) freezing the DCD_i model and training the F_i and C_i model

2.3. Top-K source domain

To perform the FLADA method, we chose the top-K source domains, which is more similar to the target domain. It should be mentioned that the results of transfer learning have a obvious correlation with the source only results, meaning that if the model trained on a source domain performs better than that trained on another source domain, the results after applying transfer learning on each domain would yield the same correlation. Therefore the source list can be selected from the results according to the cross-subject classification accuracy from high to low [5].

To verify the impact of selecting the top-K source domains as well as determine the optimal parameter K for effective domain adaptation, we conducted an ablation study comparing FLADA with different values of K. Specifically, the classification accuracy was analyzed under varying values of K. The value of K ranges from 1 to 22 for there are 22 subjects left in DREAMER datasets after selecting one subject as the target.

2.4. Attention-mechanism

The main idea of the attention mechanism is to apply different weights to every channel of raw inputs to get sensory inputs based on the query, which is implemented in imitation of human voluntary attention. In our research, we implement the FLADA method with channel-wise attention to achieve better results in emotion recognition tasks.

2.5. Building model

To implement the FLADA method on the DREAMER database, the neural network was modified. There are 14 channels recorded in the database and the signal of each channel is filtered into five bands, so the number of input data is 70. The main structure of the neural network designed for the DREAMER database is 70-64-32-16 (feature extractor), 16-2 (classifier) and 32-16-6 (DCD network). Relu function is applied behind each layer except the last layer. Then, to get an optimized result, channel-wise attention layer is constructed. Each of the 70 input channels is assigned a weight, which can be trained to adapt to the subject domain. The structure of attention layers is 70-16-70. After the

first layer, the Relu function is applied, and a sigmoid function is applied after the second layer to get weights parameter ranging from 0 to 1.

2.6. Adversarial training

After obtaining the top-K source domains and their own feature extractor and classifier, a more general feature extractor and classifier was trained for cross-subject emotion classification tasks.

Thus, considering the Siamese network in meta-learning, positive and negative samples can be formed to perform adversarial learning. Each sample is composed of paired EEG signals and labels. Those pairs with different labels are negative samples, and those with the same labels are positive samples. A discriminator is trained to distinguish whether the paired input is a positive sample or not.

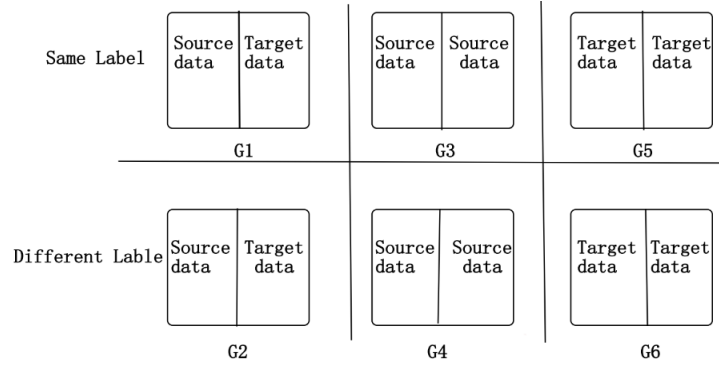


Figure 2: The data is divided into six groups to do the adversarial training, as illustrated here

Figure. 2 illustrates how we divided the samples into six groups. Then a domain discriminator is trained based on the six groups above. In this process, the classifier and feature extractor are frozen. After the domain discriminator is obtained, the weights of the discriminator are locked, while those of the feature extractor and classifier are unlocked. Then they are further trained to make the discriminator no longer work, which means that the boundary between the source and the subject domain is blurred. In other words, the feature extractor can obtain more general features that are suitable for both two domains.

2.7. Final model

Next we get trained F_i and C_i for each source domain. To fully take advantage of each source domain, we sum all the features and the result of possibility through Eq. (2), where $w_i = w_j = \frac{1}{k}$ while C_i and F_j representing the classifier and feature extractor of each subject domain..

$$\text{pre} = \text{Sigmoid}(\sum_k^i w_i \times C_i(\sum_k^j w_j F_j)) \quad (2)$$

2.8. Evaluation of datasets

To test the performance of the proposed method in datasets with limited channels, we utilize accuracy (ACC) and the area under the receiver operating characteristic curve (ROC-AUC) as our primary metrics, which provide a comprehensive measure of the performance.

The evaluation is conducted on two benchmark datasets: SEED and DREAMER. The result of SEED serves as a comparative baseline to assess the effectiveness of our approach against existing methods in a well-established datasets. Meanwhile, the result of DREAMER proves the viability of proposed method in few-channel datasets.

3. Results

3.1. Effect of the top-K source domain selection

Table 1: ACC and ROC-AUC between databases

datasets	ACC	ROC-AUC
SEED	51.2(0.03)	69.6(0.04)
DREAMER	68.9(0.05)	57.9(0.09)

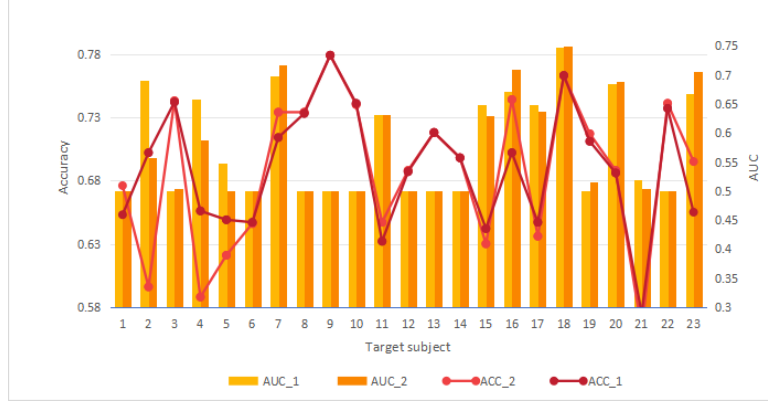


Figure 3: Mean accuracy of DREAMER with and without channel-wise attention. attention mechism doesn't improve the performance of FLADA method

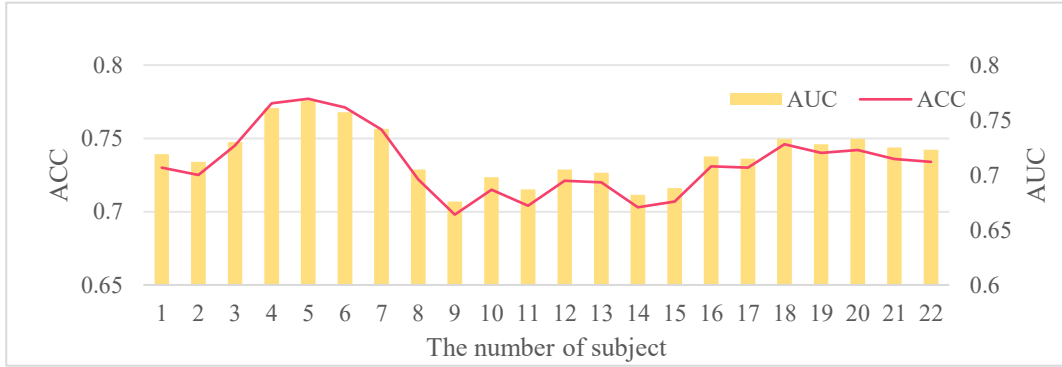


Figure 4: Accuracy vs. number of selected source domains (K). K = 5 is the best parameter for training

In Figure 4, we observe that selecting too few or too many closer-to-target source domains leads to suboptimal performance. The increase from $k=1$ to $k=4$, and when $K = 5$ the accuracy reach the peak, indicating that an optimal number of source domains is crucial for effective domain adaptation.

3.2. Results between SEED and DREAMER datasets

As shown in the Table 2. the accuracy of the results on DREAMER is 68.9%, which is better than that on SEED. This indicates the fact that FLADA method works better on datasets with a small number of channels.

3.3. Evaluation of model generalization

As illustrated in Figure 3, the relevant data from the FLADA model without the attention mechanism exhibits nearly the same performance compared to the proposed method. Channel-wise attention

shows improvement when subjects 1, 7, 11, 16 and 23 are selected as the target, while it shows a decrease in accuracy when subjects 2, 4, and 5 are selected. For the other subject, it has no effect. This indicates that channel-wise attention mechanism actually doesn't show stability and does not generally improve accuracy when applied to the FLADA method, contrary to our hypothesis.

4. Discussion

The reason why the accuracy of our proposed model seems to perform better in DREAMER datasets may be because the low-channel region better captures the core aspects of emotion detection, while other brain regions contribute less significantly and may even introduce noise. This suggests that data collection should focus on truly relevant brain regions, rather than incorporating unnecessary signals that could obscure meaningful patterns and reduce overall model performance.

The FLADA method works well on the DREAMER database, and attention mechanism has been proven effective on non-cross-subject tasks. However, our attempt to improve the performance of the FLADA method with channel-wise attention turned out not to work as well as we had expected. Attention mechanism may work well when model trained on a database consisting of different subjects. However, in the FLADA method, the model is only trained on one subject domain. The model learned different channel weights between subjects, which may even enlarge the bias. For some subjects, the features learned share more similarity with other subjects, so the results show improvement. Moreover, to make the results between datasets comparable, we adopt the FC model, the same model in the original FLADA method, in experiment. Thus, the attention method is also applied to the FC model while the attention method usually works with CNN. This may also be a reason why the channel-wise attention doesn't work as well as expected.

Furthermore, though the FLADA method deals with the 'domain shift' problem to some extent, we find that it is not practical in daily life due to its training process, and it should be modified to meet the real need. On the one hand, users can't be trained to fully understand how to measure their emotions correctly, and the misunderstood label may lead to a faulty model. On the other hand, the training cost is not affordable when the model has to be trained again for every new user, let alone the fact that you have to train the model for each source subject to find the k-domain of best fit.

5. Conclusion

In summary, the FLADA method showcases its effectiveness on a database with fewer channels, which indicates that future improvement based on the FLADA method can be done on a low-channel datasets other than the typical 64 or 32-channel datasets, thus reducing the training cost and time. However, the failure of channel-wise attention applied to FLADA and the weakness of the transfer learning indicate that using the CNN model with attention mechanism or modifying the FLADA method to train one universal model may show more potential in future research.

Acknowledgement

Jinhao Chen, Zixin Chen, and Jianlin Pan contributed equally to this work and should be considered co-first authors.

References

- [1] M. Augsburger and I. R. Galatzer-Levy, "Utilization of machine learning to test the impact of cognitive processing and emotion recognition on the development of PTSD following trauma exposure," *BMC Psychiatry*, vol. 20, no. 1, Jun. 2020, doi: 10.1186/s12888-020-02728-4.
- [2] X. Li et al., "EEG Based Emotion Recognition: A Tutorial and Review," Apr. 30, 2022, Association for Computing Machinery. doi: 10.1145/3524499.

- [3] Y. Mohammad and T. Nishida, "Using physiological signals to detect natural interactive behavior," in *Applied Intelligence*, Aug. 2010, pp. 79–92. doi: 10.1007/s10489-010-0241-4.
- [4] J. N. Bailenson et al., "Real-time classification of evoked emotions using facial feature tracking and physiological responses," *International Journal of Human Computer Studies*, vol. 66, no. 5, pp. 303–317, May 2008, doi: 10.1016/j.ijhcs.2007.10.011.
- [5] Y. Wang, J. Liu, Q. Ruan, S. Wang, and C. Wang, "Cross-subject EEG emotion classification based on few-label adversarial domain adaption[Formula presented]," *Expert Syst Appl*, vol. 185, Dec. 2021, doi: 10.1016/j.eswa.2021.115581.
- [6] W. Tao et al., "EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention," *IEEE Trans Affect Comput*, vol. 14, no. 1, pp. 382–393, Jan. 2023, doi: 10.1109/TAFFC.2020.3025777.
- [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 J Biomed Health Inform*, vol. 22, no. 1, pp. 98–107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.
- [8] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Trans Auton Ment Dev*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.
- [9] P. Keelawat, N. Thammasan, M. Numao, and B. Kijirikul, "A comparative study of window size and channel arrangement on eeg-emotion recognition using deep cnn," *Sensors*, vol. 21, no. 5, pp. 1–19, Mar. 2021, doi: 10.3390/s21051678.
- [10] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Information Fusion*, vol. 59, pp. 103–126, Jul. 2020, doi: 10.1016/j.inffus.2020.01.011.
- [11] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification," in *International IEEE/EMBS Conference on Neural Engineering, NER*, 2013, pp. 81–84. doi: 10.1109/NER.2013.6695876.
- [12] L.-C. Shi and B.-L. Lu, "Off-Line and On-Line Vigilance Estimation Based on Linear Dynamical System and Manifold Learning," 2010.