# Supplementary Material

June 9, 2025

## Contents



## 1 Appendix A: Discussion on the relationship between CNN and Transformer in cross-attention

The results in Table 1 indicate that the CNN branch not only contributes to the extraction of regional features, but also plays a greater role in conjunction with the Transformer branch, effectively integrating local and global features through cross-attention to enhance the representation of brain region features. However, we observed that removing the CNN only resulted in a 0.85% decrease in accuracy. This initially raised concerns about whether the CNN branch played a limited role. However, further investigation revealed that this phenomenon is primarily due to the residual connection within the cross-attention structure. In this appendix, we provide an in-depth analysis of the relationship between residual connection and feature contributions from the CNN and Transformer branches.

We observed the model structure and noticed that we introduced residual connection (one of the commonly used operations in Transformers) in the cross-attention structure, as shown in Formula (1).

$$F = MHA(F_{trans}, F_{cnn}, F_{cnn}) + F_{trans} \tag{1}$$

Here, $F_{\text{trans}}$ represents the features from the Transformer branch, and $F_{\text{cnn}}$ represents features from the CNN branch. The Multi-Head Attention (MHA) treats the Transformer features as the

Table 1: The verification of the validity of the relevant modules of DFRNet on the SEED-IV dataset.

| No structure | CNN | Transformer | Cross-attention | RFM | ACC / Std(%) | F1 / Std(%) |
|---|---|---|---|---|---|---|
| **No CNN** | ✗ | ✓ | ✓ | ✓ | **84.43 / 13.42** | **80.89 / 16.70** |
| **No Transformer** | ✓ | ✗ | ✓ | ✓ | **80.32 / 13.75** | **76.59 / 16.53** |
| No CNN branch | ✗ | ✓ | ✗ | ✓ | 83.25 / 12.78 | 79.22 / 16.82 |
| No Transformer branch | ✓ | ✗ | ✗ | ✓ | 83.53 / 13.13 | 81.14 / 15.37 |
| No cross-attention | ✓ | ✓ | ✗ | ✓ | 83.95 / 13.10 | 80.24 / 16.82 |
| No REM | ✓ | ✓ | ✓ | ✗ | 82.79 / 14.38 | 79.51 / 17.07 |
| Completed structure | ✓ | ✓ | ✓ | ✓ | 85.28 / 14.01 | 81.81 / 17.24 |

*A cross (✗) indicates that DRFNet does not contain the structure, while a check mark (✓) indicates that it does.*

query (Q) and CNN features as the key and value (K and V). The result is then added back to the original Transformer features through a residual connection.

We speculate that the residual connection allows feature $F$ to retain a significant amount of features from the Transformer branch after cross-attention. When CNN features are replaced by Transformer, the overall feature changes within the model are not particularly significant, and the model still learns relatively good content, resulting in only a 0.85% decrease. To validate our hypothesis, we conducted additional experiments by swapping the input positions of the CNN and Transformer branches in the cross-attention and repeating the ablation in Table 1. The experimental results are shown in Table 2. When the features from the CNN branch are used as Q, and the features from the Transformer branch are used as K and V, the accuracy is 81.74%. When we remove the Transformer, the result is 80.32%, a decrease of only 1.42%. Therefore, it is due to the existence of residual connection that replacing another input in cross-attention does not have as large an impact on overall features as initially imagined.

Table 2: The verification of the validity of the relevant modules of DFRNet on the SEED-IV dataset.

| | Cross-attention | ACC / Std(%) | F1 / Std(%) |
|---|---|---|---|
| Swap the input | $F = MHA(F_{cnn}, F_{trans}, F_{trans}) + F_{cnn}$ | 81.74 / 11.83 | 78.87 / 13.63 |
| No trans | $F = MHA(F_{cnn}, F_{cnn}, F_{cnn}) + F_{cnn}$ | 80.32 / 13.75 | 76.59 / 16.53 |
| Original input | $F = MHA(F_{trans}, F_{cnn}, F_{cnn}) + F_{trans}$ | 85.28 / 14.01 | 81.81 / 17.24 |
| No CNN | $F = MHA(F_{trans}, F_{trans}, F_{trans}) + F_{trans}$ | 84.43 / 13.42 | 80.89 / 16.70 |

We realize that two additional questions need to be explained here: "Why introduce residual connection in cross-attention?" This is because, in our preliminary experiments and model exploration, We found that introducing residual connection would yield better results. Furthermore, the self-attention mechanism in standard Transformers also includes this residual connection, so we incorporated it as well. Another question is, "Is the CNN branch necessary?" It is important to note that the results in Tables 1 and 2 are the average results from the three sessions of SEED-IV. After adding CNN, there was a stable improvement of 0.85% on average in the dataset, which is still considerable. This indicates that the CNN structure plays a significant role.

Furthermore, by comparing the results of "No CNN" and "No Transformer" in Tables 1 and 2, we find that the Transformer branch is suitable for the cross-attention structure, while the features of the CNN branch may not be suitable for input into the cross-attention alone. We speculate that in our cross-attention structure, the fusion relationship between local features and global features is more likely to be the former supplementing and assisting the latter.

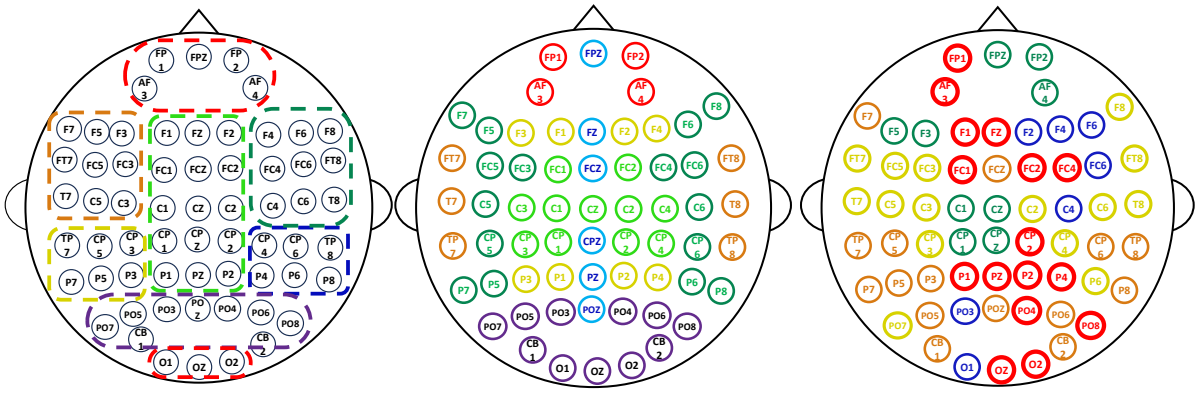
# 2 Appendix B: Detailed regional classification method

To make our research more complete, we added three brain region divisions as shown in Figure 1 to explore the impact of different brain region division methods on the accuracy of DRFNet.

**(1) The detailed introduction to other three ways of brain region division:**

First, we set up a non-overlapping brain region division, as shown in Figure 1 (a). This division retains the division of regions 1/2/4/7, while compressing regions 3/5/6, thereby preserving the original division as much as possible.

Second, we designed a brain region division similar to Yeo's 7-network, as shown in Figure 1(b). Yeo et al. [1] conducted a large-scale resting-state functional connectivity study, performing data-driven clustering analysis on cortical surface data from 1,000 healthy young adults and dividing the brain into seven major functional networks at the coarsest resolution. Yeo's 7-network achieved a consistency of up to 97.4% across different subsamples. Due to the inability to determine the exact electrode locations and the complexity of Yeo's 7-network, we were unable to achieve complete alignment with its division; however, we have made every effort to align as closely as possible with this standard.



(a) Non-overlapping division (b) Division based on Yeo's 7-network (c) Automatically learned division

Figure 1: The division methods of the brain regions

At last, we adopted a data-driven brain region segmentation method called Automatically learned division. Previous studies have used GCN[2] methods to capture spatial information between electrodes. A common single-layer GCN formula is shown below:

$$H^{l+1} = \sigma(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}H^{l}W^{l}) \tag{2}$$

where $A$ is the adjacency matrix, $\hat{A} = A + I$, $\hat{D}$ is the degree matrix of $\hat{A}$, $H$ is the feature of the graph nodes, and $W$ is the learnable parameter matrix. The GCN method utilizes an adjacency matrix to introduce "structural or functional" connections between electrodes into the feature aggregation and learning process, thereby enabling the neural network to capture the relationships between electrodes. Among them, RGNN[3] is a well-known GCN method in the field of EEG emotion recognition. However, it sets the adjacency matrix as learnable, and after feeding it with a large amount of data, the adjacency matrix $A$ learns the potential, deep-level relationships between EEG electrodes.

The adjacency matrix reflects the correlations between electrodes, and we consider dividing the brain into regions based on the automatically learned $A$. Therefore, we replicated the RGNN method, using three session data to train the network on the SEED-IV dataset, and then weighted the adjacency matrices of 15 subjects based on the accuracy on the test set to obtain a comprehensive adjacency matrix $A$. For the obtained $A$, we first used thresholding to retain only the top 10% of the strongest edges,i.e., the most significant connections. The Leiden community detection algorithm [4] can partition nodes in a graph into several "communities," ensuring that nodes within the same community are as closely connected as possible, while connections between different communities are as sparse as possible. Therefore, we used the Leiden algorithm to divide the brain into irregular regions as shown in Figure 1 (c). The code for this part has been made publicly available on GitHub to ensure the reproducibility of the experiment and facilitate communication among scholars. (https://github.com/braverSheep/DRFNet/tree/main/brain_region_division)

Table 3: The results of four different brain region divisions on the SEED and SEED-IV datasets.

| Datasets | Fusion method | ACC (%) | Std (%) | F1-score (%) | Std (%) |
|---|---|---|---|---|---|
| SEED | Division based on Yeo's 7-network | 96.36 | 07.13 | 96.03 | 07.65 |
| | Automatically learned division | 93.84 | 08.85 | 93.36 | 09.46 |
| | Non-overlapping division | 94.66 | 08.55 | 94.18 | 09.08 |
| | Original division | 97.27 | 06.15 | 97.00 | 06.76 |
| SEED-IV | Division based on Yeo's 7-network | 84.09 | 13.63 | 80.53 | 17.27 |
| | Automatically learned division | 82.54 | 13.72 | 79.16 | 16.86 |
| | Non-overlapping division | 83.83 | 12.44 | 82.19 | 14.91 |
| | Original division | 85.28 | 14.01 | 81.81 | 17.24 |

**(2) The analysis of experimental results:**

Table 3 shows the results of the four brain region partitioning methods on the SEED and SEED-IV datasets. Based on the experimental results, we can summarize the following points:

1. The original brain region segmentation method is more suitable for our approach.

2. Our model still achieves acceptable results under Yeo's 7-network brain region segmentation method, indicating that the model has a certain degree of robustness for brain regions.

3. Under the original division method, the accuracy of non overlapping is lower, which may be because the non overlapping method deprives electrodes located at the edge of brain regions of the ability to exchange information in multiple areas.

4. As can be clearly seen from the table, the accuracy of the Automatically learned division method has significantly decreased on both the SEED and SEED-IV datasets. First, this indicates that DRFNet is not without requirements for brain region division; the more reasonable the brain region division, the better DRFNet can perform. Second, it reflects that there are certain difficulties in automatically dividing brain regions through data, and we speculate that this may be more influenced by issues such as insufficient data volume and poor data quality. The actual situation needs to be verified through more experiments, and we plan to explore this as one of our next research directions.

# 3 Appendix C: Analysis of the SEED-VII dataset

Compared to other datasets, DRFNet's performance on SEED-VII has declined significantly. SEED-VII is the latest emotion dataset released by Shanghai Jiao Tong University's Brain-like Computing and Machine Intelligence research group. As shown in the Figure 2, the overall accuracy of the models reported by them on this dataset is not high. Most models perform poorly, which in turn suggests that this is likely due to the complexity of the categories and the characteristics of the dataset itself.

The accuracies and F1 scores (Avg./Std., %) achieved by different methods using EEG or eye movement signals.

| | Method | EEG signals | | | | | | Eye movements |
|---|---|---|---|---|---|---|---|---|
| | | delta band | theta band | alpha band | beta band | gamma band | all bands | |
| Accuracy | KNN [70] | 28.21/5.76 | 28.47/7.02 | 31.92/6.51 | 34.83/5.77 | 37.20/6.06 | 36.43/**5.38** | 36.01/**6.30** |
| | HCNN [36] | 42.33/6.36 | 42.63/5.28 | 43.62/**4.82** | 47.79/6.77 | 48.18/6.10 | 52.42/6.47 | - |
| | RGNN [37] | 42.31/5.14 | 41.73/5.53 | 43.93/4.98 | 44.68/**5.51** | 45.49/**4.73** | 48.50/6.83 | - |
| | Transformer [42] | 46.87/**4.10** | 47.01/**4.23** | 48.78/5.39 | 53.10/6.34 | 53.96/6.60 | 56.04/7.82 | - |
| | GCNCA [38] | **49.97**/4.94 | **50.46**/4.29 | **53.26**/5.45 | **58.36**/6.69 | **59.50**/5.77 | 58.04/7.78 | - |
| | MAET | - | - | - | - | - | **58.11**/8.78 | **50.31**/7.14 |
| F1 score | KNN [70] | 26.00/5.98 | 26.12/7.25 | 29.67/6.64 | 33.22/**5.59** | 34.98/5.62 | 34.08/**5.79** | 34.74/**6.33** |
| | HCNN [36] | 37.85/7.78 | 38.30/5.72 | 39.66/**5.23** | 44.05/7.55 | 44.33/6.68 | 49.02/6.80 | - |
| | RGNN [37] | 37.85/5.56 | 36.13/5.79 | 39.18/5.43 | 40.68/06.61 | 41.77/**5.50** | 45.32/7.20 | - |
| | Transformer [42] | 43.20/**4.26** | 43.85/**4.74** | 45.48/**5.23** | 49.81/6.70 | 51.48/6.96 | 53.35/8.30 | - |
| | GCNCA [38] | **47.61**/5.14 | **47.92**/5.01 | **50.99**/5.71 | **55.97**/7.12 | **57.54**/5.90 | **55.48**/8.30 | - |
| | MAET | - | - | - | - | - | 54.98/9.45 | **47.10**/7.84 |

Figure 2: The performance results of various models reported in the original paper[5] on the SEED-VII dataset.

First, increasing the number of emotion categories will greatly reduce the spatial intervals between various emotion signals, making it difficult for the model to distinguish between similar emotion categories. As shown in Figure 3, DRFNet misclassified a large number of happy and fearful samples as surprised, because happiness, fear, and surprise all belong to high arousal states. Additionally, sadness is typically characterized by low arousal and negative valence.

If the sadness is not sufficiently intense, and the participant is in a “mild sadness” state, EEG signals may resemble neutral states. As shown in the confusion matrix, DRFNet incorrectly classified sadness samples as neutral emotions.

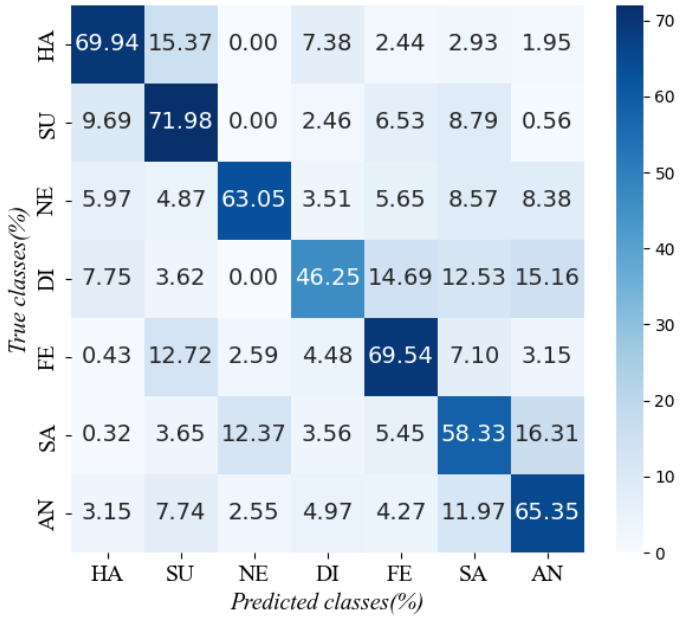



Figure 3: Confusion matrix of DRFNet’s results on the SEED-VII dataset.

Secondly, when collecting data, if the participants cannot be effectively stimulated to produce the corresponding emotions, the quality of the data will be low. As shown in the Figure 4, this is the "subjectID_date_sessionID_save_info.csv" file included with the publicly available SEED-VII, which contains the feedback from subject 1 for each movie clip. From this, we can see that some trials failed to effectively trigger the participants to produce the corresponding emotions, with many participants rating their level of successful induction below 0.5 (on a scale from 0 to 1). The SEED-VII data may contain a significant number of false labels, making it quite challenging.

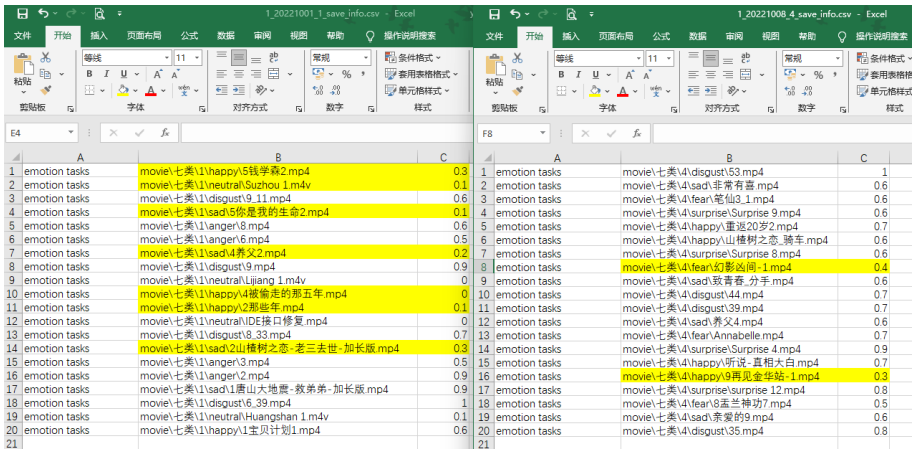



Figure 4: Confusion matrix of our model results on the SEED-VII dataset.

# 4 Appendix D: Cross-validation on the SEED-IV dataset

To facilitate fair future comparisons and provide a reproducible benchmark, we supplement our work with a subject-dependent k-fold cross-validation evaluation on the SEED-IV dataset. This complements the main experimental setup based on commonly used fixed splits in the literature and serves as a reference for future researchers interested in robust evaluation under multiple splits.

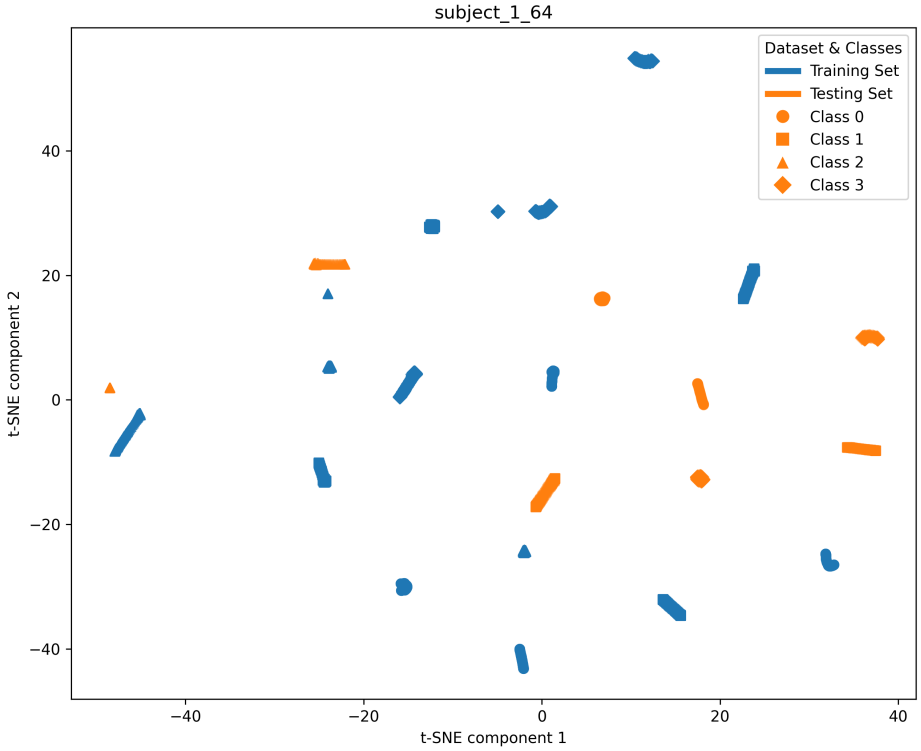



Figure 5: t-SNE visualization of subject1 data in the SEED dataset. There are a total of 24 data clusters, corresponding to the 24 trial experiments collected for subject 1.

While existing works such as DGCNN[6], RGNN[3], and PGCN[7] typically follow subject-dependent evaluation protocols with fixed training/test trials per subject, some studies have employed n-fold cross-validation. However, we have not yet seen a unified cross-validation method. In particular, certain cross-validation methods divide EEG samples randomly without preserving trial integrity, which may lead to data leakage—samples from the same trial appearing in both training and test sets—thus overestimating performance. As shown in Figure 5, samples from the same trail are highly similar. We have conducted experiments in which all data were randomly divided into n folds, which makes the final experimental results easily approach 1. Therefore, samples from the same trial should not be separated, as this could potentially lead to label leakage. To avoid this issue, we adopt a trial-level fold splitting strategy: in each fold, entire trials are kept intact and not split between training and test sets.

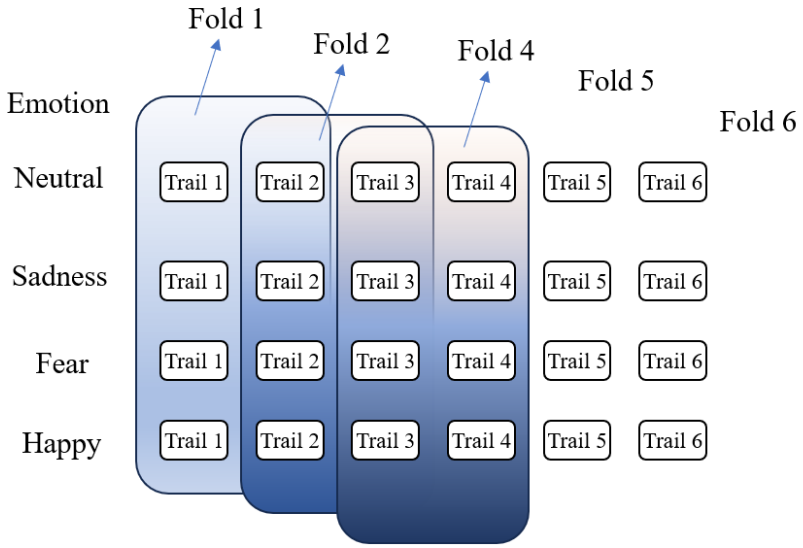



Figure 6: 6-fold cross-validation on the SEED-IV dataset. In the SEED-IV dataset, each person has 24 trial data points, with 6 trials for each emotion category.

For the convenience of future researchers who wish to use cross-validation to fairly compare our model, we will supplement our experiments with cross-validation and publish the fold division method and results on GitHub (https://github.com/braverSheep/DRFNet/tree/main/SEED-IV_cross_validation). We have temporarily conducted cross-validation on the SEED-IV dataset, following the original test set splitting ratio, where two trials per emotion are used as the test set. The data is then split into 6 folds as shown in the Figure 6. Table 4 presents our experimental results on the SEED-IV dataset.

The average accuracy across all folds and sessions is 87.12%, demonstrating strong and stable performance of DRFNet under multi-fold subject-dependent settings. We plan to extend this cross-validation framework to other datasets (e.g., SEED, SEED-V) and make all fold splits publicly available to promote fairness and reproducibility in EEG emotion recognition research.

Table 4: Fold-wise ACC and Std for Each Session, Their Three-Session Averages, and Overall Averages. (%)

| fold | session1 | | session2 | | session3 | | Average of 3 sessions | | All averages | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | Std | ACC | Std | ACC | Std | ACC | Std | ACC | Std |
| 1 | 78.42 | 12.51 | 86.77 | 10.16 | 89.39 | 9.79 | 84.86 | 10.82 | 87.12 | 11.41 |
| 2 | 80.48 | 12.11 | 86.13 | 8.24 | 90.00 | 8.72 | 85.53 | 9.69 | | |
| 3 | 87.29 | 10.32 | 91.76 | 8.40 | 91.81 | 7.71 | 90.28 | 8.81 | | |
| 4 | 86.67 | 12.82 | 84.59 | 12.51 | 91.22 | 13.15 | 87.49 | 12.82 | | |
| 5 | 83.13 | 14.54 | 83.75 | 13.55 | 84.24 | 16.64 | 83.70 | 14.91 | | |
| 6 | 83.19 | 12.46 | 86.60 | 10.57 | 89.33 | 11.20 | 86.37 | 11.41 | | |

# References


[1] B. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[2] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[3] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2022.

[4] V. A. Traag, L. Waltman, and N. J. Van Eck, "From louvain to leiden: guaranteeing well-connected communities," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[5] W.-B. Jiang, X.-H. Liu, W.-L. Zheng, and B.-L. Lu, "Seed-vii: A multimodal dataset of six basic emotions with continuous labels for emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–16, 2024.

[6] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.

[7] M. Jin, C. Du, H. He, T. Cai, and J. Li, "Pgcn: Pyramidal graph convolutional network for eeg emotion recognition," *IEEE Transactions on Multimedia*, 2024.