Multi-expert collaboration and knowledge enhancement network for multimodal emotion recognition

Kun Wang, Junyong Zhao, Liying Zhang, Qi Zhu (✉), and Daoqiang Zhang (✉)

College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics,
Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education,
Nanjing 211106, China.
zhuqinuaa@163.com, dqzhang@nuaa.edu.cn

Abstract. Emotion recognition leveraging multimodal data plays a pivotal role in human-computer interaction and clinical applications, such as depression, mania, Parkinson's Disease, etc. However, existing emotion recognition methods are susceptible to heterogeneous feature representations across modalities. Additionally, complex emotions involve multiple dimensions, which presents challenges for achieving highly trustworthy decisions. To address these challenges, in this paper, we propose a novel multi-expert collaboration and knowledge enhancement network for multimodal emotion recognition. First, we devise a cross-modal fusion module to dynamically aggregate complementary features from EEG and facial expressions through attention-guided. Second, our approach incorporates a feature prototype alignment module to enhance the consistency of multimodal feature representations. Then, we design a prior knowledge enhancement module that injects original dynamic brain networks into feature learning to enhance the feature representation. Finally, we introduce a multi-expert collaborative decision module designed to refine predictions, enhancing the robustness of classification results. Experimental results on the DEAP dataset demonstrate that our proposed method surpasses several state-of-the-art emotion recognition techniques.

Keywords: Emotion recognition · EEG · Multimodal fusion · Knowledge enhancement · Multi-expert collaboration.

1 Introduction

Affective computing shows significant potential in healthcare applications, especially for assisting in diagnosing emotional disorders and psychiatric conditions [13,5]. For example, quantitative emotional assessment can help in the early diagnosis of Parkinson's Disease [12]. However, mental health disorders often exhibit dynamic emotional changes in complex environments, which can be reflected through variations in facial expressions or EEG signals. Therefore, how to develop a multimodal data emotion recognition system to improve the accuracy of disease diagnosis is the key.

As affective states and distinct neurophysiological patterns are established associations, particularly in frequency bands (alpha, beta, theta) and spatial characteristics like frontal asymmetry [16,19]. Emotion recognition based on EEG signals has gained wide attention in recent years [21,9]. For example, Wu *et al.* [21] proposed a graph orthogonal purification network designed to capture both emotion-relevant and emotion-irrelevant features, which address distribution discrepancies across different emotion feature spaces. Li *et al.* [9] proposed a cross-attention-based dilated causal convolutional neural network to extract more discriminative features related to emotions. This method introduces the domain discriminator to reduce individual variability. However, these methods rely solely on single-modality data, which may not provide sufficient discriminative features for complex affective computing tasks. Multimodal emotion analysis can integrate complementary multi-dimensional information (*i.e.*, EEG and facial expressions) to enhance model adaptability in complex scenarios. Therefore, multimodal-based emotion recognition methods are proposed to improve emotion recognition performance [15,25]. For instance, Sun *et al.* [15] proposed a mutual information-based disentangled multimodal representation learning framework to balance and integrate the contributions of these diverse types of information. Meanwhile, Yin *et al.* [25] proposed a multimodal and channel attention fusion transformer to model inter-channel correlations and enhance emotion recognition accuracy by emphasizing global temporal dependencies. However, existing multimodal emotion recognition methods often yield suboptimal classification performance due to the lack of prior knowledge integration [27].

Brain networks based on EEG can reveal the topological associations of activation between different brain regions, which not only enhance feature discriminability for emotion recognition but also provide spatiotemporal prior information for understanding emotional fluctuations [22]. For example, Zheng *et al.* [27] proposed a prior-driven dynamic functional connectivity network to extract complex spatial-temporal features to aid emotion recognition. In addition, Chen *et al.* [3] proposed a comprehensive multi-source learning network to incorporate the information from multi-source data. However, existing methods still ignore both the time-frequency characteristics of raw EEG signals and the detailed features of facial expressions. In addition, the static weighting mechanisms in multimodal fusion algorithms may lead to misclassification in complex emotion recognition scenarios due to their limited adaptability.

To solve the above challenges, we propose a novel multi-expert collaboration and knowledge enhancement network for multimodal emotion recognition. Specifically, we propose a cross-modal fusion module (CMF) to dynamically aggregate complementary features from EEG and facial expressions through attention-guided interaction. Furthermore, we calculate the feature prototype alignment (FPA) loss to align the multimodal feature. Then, we design a prior knowledge enhancement module (PKE) that injects original dynamic brain networks into feature learning to enhance the feature representation. Finally, we propose a multi-expert collaborative decision module (MCD) to obtain robustness classification results.

2 Method

As shown in Fig.1, the proposed framework architecture consists of four interconnected core components that collectively address the challenges of multimodal emotion recognition. The cross-modal fusion module establishes deep feature interactions between EEG signals and facial expressions. The prior knowledge enhancement module incorporates domain-specific knowledge to enrich feature representations. The multi-expert collaborative decision module integrates complementary expertise, while the feature prototype alignment loss minimizes distribution discrepancies between multimodal feature representations, preserving their distinctive characteristics.

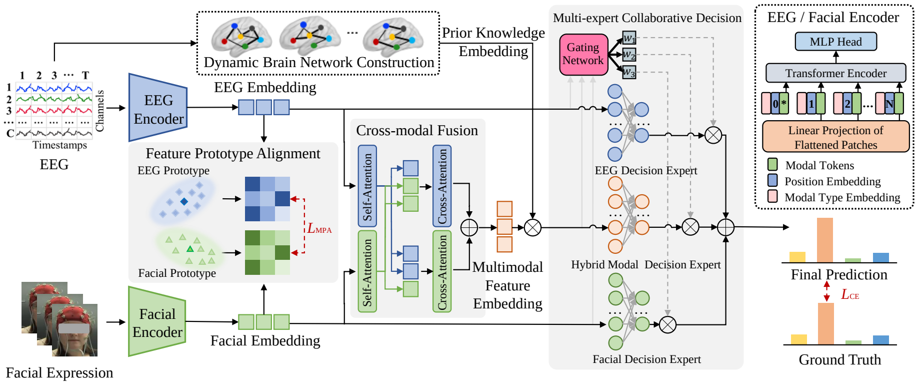

Fig. 1. Illustrate our proposed multimodal emotion recognition method.

Cross-modal Fusion Module. To effectively integrate multimodal features, we propose the cross-modal fusion module. Specifically, the EEG signals $\mathbf{X}_{\text{EEG}} \in \mathbb{R}^{T \times C_E \times D}$ and facial expression images $X_{Facial} \in \mathbb{R}^{T \times C_F \times D}$ are fed into the EEG encoder and facial encoder to obtain the EEG embedding emb_{EEG} and facial expression embedding emb_{facial}, respectively, where C_E and C_F denote the number of channels, T indicates the number of time windows, and D represents the feature dimension. It is worth noting that the EEG encoder and facial encoder use the EEGViT [24] and ViT [4] to extract the EEG feature and facial feature, respectively. Then, these modality-specific embeddings are then fused through our cross-modal fusion module, which comprises two self-attention blocks for intra-modal refinement and two cross-attention blocks for inter-modal interaction, ultimately yielding a unified multimodal feature representation emb_{MF}. Therefore, the cross-modal fusion module is defined as,

$$\begin{aligned} Q_{EEG}, K_{EEG}, V_{EEG} &= ATT_{self}(emb_{EEG}) \\ Q_{Facial}, K_{Facial}, V_{Facial} &= ATT_{self}(emb_{facial}) \\ emb_{MF} &= ATT_{cross}(Q_{Facial}, K_{EEG}, V_{EEG}) \\ &\quad \oplus ATT_{cross}(Q_{EEG}, K_{Facial}, V_{Facial}) \end{aligned} \tag{1}$$

where ATT_{self} and ATT_{cross} represent the self-attention block and the cross-attention block, respectively. $\oplus$ denotes the concatenate operation.

Prior Knowledge Enhancement Module. In EEG recordings, each electrode not only captures electrophysiological activity within its corresponding brain region, but also contributes to emotional state recognition [18,23]. Therefore, we propose to construct a dynamic brain network to establish prior knowledge for improving feature discriminability. In our paper, EEG electrode channels are simplified as nodes in constructing an EEG network to characterize information in brain regions, and signal linkages between them form edges connecting these nodes. The connection strength is quantified as edge weights in the network. Specifically, we obtain each subject's EEG data $X_{EEG} \in \mathrm{R}^{T \times C_E \times D}$. To establish the functional connectivity network, we compute the Pearson correlation coefficient between EEG from a pair of channels within the t-th time window,

$$PCC_{ij}(t) = \frac{Cov(x_i(t), x_j(t))}{\sigma_i(t)\sigma_j(t)} \tag{2}$$

where Cov denotes the covariance between two vectors, σ represents the standard deviation operation, $x_i(t)$ and $x_j(t)$ denotes the EEG of a pair of channels i and j within the t-th time window, respectively. Thus, for each subject, the dynamic brain network $DFCN_{ori} \in \mathrm{R}^{T \times C \times C}$ consists of T transient functional connectivity network $[C(1), C(2), \cdots, C(t)]^T$, where each matrix captures time-varying functional interactions between brain regions. Finally, we use the dynamic brain network as prior knowledge to guide the fusion process of multimodal features and help the model better capture key emotion-relevant features. The fusion process is defined as,

$$F_{prior} = emb_{MF} \otimes DFCN_{ori} \tag{3}$$

where $\otimes$ represents the element-wise multiplication operation.

Multi-expert Collaborative Decision Module. To reduce decision error in complex emotion recognition tasks, we proposed a multi-expert collaborative decision module to integrate cross-domain expert knowledge to improve the decision-making accuracy of the model. Specifically, the multi-expert collaborative decision module consists of three same-expert decision blocks (*i.e.*, EEG decision expert, hybrid modal decision expert, and facial decision expert) and a gating network. Furthermore, the different embedding feature is fed into the different expert decision blocks to obtain classification results. Therefore, each expert decision block is defined as,

$$\text{expert}_m = \mathrm{W}_{o,m} \cdot GELU(\mathrm{W}_{h,m} \cdot emb_m + \mathrm{b}_{h,m}) + \mathrm{b}_{o,m} \tag{4}$$

where $m \in M = [EEG, facial, MF]$ represent different modal sets. $\mathrm{W}_{o,m}$, $\mathrm{b}_{o,m}$, $\mathrm{W}_{h,m}$, $\mathrm{b}_{h,m}$ are the learnable weights and biases of the m-th expert network. The concatenated embedding features are subsequently fed into a gating network to dynamically determine weighting coefficients for each expert's output. This gating network employs an average pooling layer followed by the SoftMax function, enabling adaptive weight allocation according to diverse feature representations. Thus, the gating weights are defined as,

$$\text{weight}_m = Softmax(Avgpool(\mathrm{W}_m \cdot emb_m + \mathrm{b}_m)) \tag{5}$$

The outputs of all experts are computed and stacked together to form a tensor $\text{expert}_{EEG}(emb_{EEG})$, $\text{expert}_{facial}(emb_{facial})$, $\text{expert}_{MF}(emb_{MF})$. It follows that the output of the multi-expert collaborative decision module is the weighted sum of all expert outputs,

$$outputs = \sum_{m \in M} \text{weight}_m \cdot \text{expert}_m(emb_m) \tag{6}$$

where W_m is the m-th element of the weights vector. $\text{expert}_m(emb_m)$ represents the m-th expert's output. Finally, we obtain the final emotion recognition classification result.

Feature Prototype Alignment. Different modalities (*i.e.*, EEG signals and facial expressions) exhibit inherent structural and representational disparities that challenge their effective alignment within a shared embedding space. To address this issue, we propose a feature prototype alignment mechanism to leverage the complementary features of multimodal emotions, which can preserve their distinctive characteristics. We employed K-means clustering to capture global patterns in emotion distribution, initializing prototypes as modality-specific emotion category centroids during the feature embedding process. Thus, the prototype vector of the modal $u \in [EEG, faical]$ is defined as,

$$P_u = \frac{1}{N_u} \sum_{i=1}^{N_u} emb_u \tag{7}$$

where N_u is the number of samples for modal u, and emb_u is the feature representation of the modal u. Minimizing the distance between projected features and their respective modal prototypes ensures cross-modal alignment.

$$\mathcal{L}_P = \frac{1}{N} \sum_{i=1}^{N} \left(\left\| emb^i_{EEG} - P_{EEG} \right\|_2^2 + \left\| emb^i_{facial} - P_{facial} \right\|_2^2 \right) \tag{8}$$

where $|| \cdot ||$ represents the L2_norm.

3 Experiment Results

Multimodal Emotion Databases. The DEAP dataset is a multimodal dataset designed for the analysis of human emotional states [7]. The dataset comprises a 32-channel electroencephalogram (EEG) and peripheral physiological signals collected from 32 participants, with supplementary facial expression recordings obtained from the first 22 subjects. During the experiment, each participant viewed 40 standardized one-minute video stimuli, followed by comprehensive affective evaluations using a validated assessment scale. Post-stimulus ratings encompassed five discrete dimensions: arousal, valence, dominance, liking, and familiarity.

Data Processing. A comprehensive data preprocessing pipeline was implemented for the DEAP dataset, wherein the 32-channel EEG signals were downsampled to 128Hz. Power spectral density (PSD) features were subsequently extracted using the Welch method with a 3-second non-overlapping window segmentation. Following established methodologies in the field [26], we formulated the emotion recognition task as a binary classification problem by implementing a threshold of 5 to dichotomize the continuous rating scales. For facial expression analysis, we employed OpenFace [1] to extract facial features from video frames, ensuring temporal synchronization with corresponding physiological data.

Experiment settings. In our experiment, we employed the leave-one-subject-out (LOSO) cross-validation strategy, which is particularly suitable for subject-independent emotion recognition tasks. In each validation iteration, the samples from a single subject were held out as the testing set, while the remaining subjects' samples from the dataset constituted the training set. The final performance metrics were computed as the average values across all iterations, thereby providing a comprehensive assessment of the model's generalization capability across different individuals. The identification performance was evaluated using accuracy (ACC) and F1-score. The proposed model was implemented in PyTorch and trained on an NVIDIA GeForce RTX3090 GPU. The Adam optimizer was used to optimize our method, and the learning rate and batch size were set as 0.001 and 40, respectively. We employed the cross-entropy loss to supervise the multimodal learning process. The source code of this work is available at https://github.com/EEGBrainNet/Emotion-Recognition.

Results and Discussions. To evaluate the effectiveness of our proposed method, we compared our method with EEG-based methods (*i.e.*, DGCNN [14] and EEGNet [8]), facial-based methods (*i.e.*, ViT [4] and GAT [17]), and multimodal-based methods (*i.e.*, MKL [2], LSTM [6], DCCA [10], MoGE [11], DFCNs [27], and Milmer [20]) in DEAP dataset. The experimental results are reported in Table 1. As shown in Table 1, our proposed method achieves the best emotion recognition performance. For example, on valence and arousal, the average ACC and F1 of our method achieved 70.01%, 74.82%, 71.67%, and 76.15%, respectively. The main reason for the superiority of our method is that our method

Table 1. Experimental results of our method with comparison methods (Mean±Std%).

Modality	Method	Valence		Arousal	
		ACC	F1	ACC	F1
EEG	DGCNN [14]	62.54/09.32	61.54/11.25	62.56/09.64	68.09/11.67
	EEGNet [8]	61.11/06.19	60.41/20.47	61.52/09.59	66.07/12.37
Facial	ViT [4]	62.51/10.17	65.11/10.66	62.52/08.23	69.39/13.94
	GAT [17]	61.53/06.98	63.92/18.62	66.38/09.47	74.69/13.82
EEG+Facial	MKL [2]	56.94/08.38	64.31/13.90	55.14/09.25	65.39/10.03
	LSTM [6]	64.86/09.52	62.95/14.95	64.30/11.78	76.28/09.96
	DCCA [10]	57.08/08.79	68.42/08.04	53.75/08.59	65.04/10.26
	MoGE [11]	65.01/08.93	66.67/10.74	65.08/09.35	69.57/16.24
	DFCNs [27]	67.36/05.58	69.17/09.01	68.47/08.65	74.68/13.36
	Milmer [20]	67.77/09.49	68.06/15.43	67.78/08.49	75.02/11.52
	Ours	**70.01/09.42**	**74.82/16.59**	**71.67/07.50**	**76.15/13.35**

Table 2. Ablation studies of our proposed method (Mean±Std%).

Method	Valence		Arousal	
	ACC	F1	ACC	F1
Baseline	61.66/09.50	64.89/16.29	63.19/09.68	65.58/13.25
w/o CMF	66.51/08.25	67.28/12.45	67.44/08.44	67.51/12.13
w/o PKE	68.19/09.08	68.89/15.51	69.16/09.75	70.17/10.86
w/o MCD	67.50/08.45	69.76/12.94	68.33/08.53	68.29/14.50
w/o FPA	69.02/09.34	70.50/16.25	70.27/08.89	74.74/11.66
Ours	**70.01/09.42**	**74.82/16.59**	**71.67/07.50**	**76.15/13.35**

not only uses the dynamic brain network as prior knowledge to enhance the feature representation, but also introduces multi-expert collaborative decision to improve the classification performance of emotion recognition.

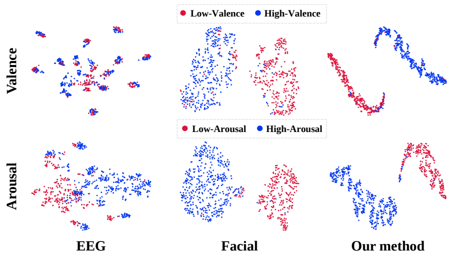

Fig. 2. The visualization of the distribution of features extracted by each modal and multimodal of the proposed model.

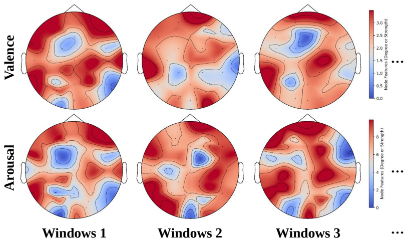

Fig. 3. Dynamic brain network topography across time windows.

To evaluate the effectiveness of our proposed module, we conduct several ablation studies on the DEAP dataset. Besides, we establish a baseline that simply fuses two modal features through concatenate operations. The experiment results are reported in Table 2. As shown in Table 2, our proposed module achieved effective improvement with the baseline method, with an increase of 8.35%, 9.93%, 8.48%, and 10.57% for ACC and F1 on valence and arousal, respectively. Specifically, we proposed the cross-modal fusion module and the prior knowledge enhancement module to obtain better improvement with the baseline. The possible reason may be that the two modules not only can enhance emotion-relevant feature representation, but also help the model focus on key emotion-relevant brain regions, reducing redundant feature learning. Furthermore, the removal of the MCD module leads to performance degradation compared to our method, resulting in accuracy reductions of 2.51% and 3.34% for valence and arousal classification, respectively. The reason lies in that MCD effectively reduces decision bias and further improves classification performance. Moreover, when the FPA module is removed from our proposed method, performance degradation occurs. Experimental results demonstrate that the FPA module can effectively reduce inter-modal heterogeneity and enhance cross-modal semantic consistency.

To analyze the performance differences among emotion recognition methods, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize high-dimensional feature distributions. As shown in Fig. 2, our method demonstrates clearer cluster separation in the embedded space, indicating enhanced discriminative power for affective states. Experimental results confirm the framework's superiority in multimodal emotion recognition, particularly in valence-arousal classification accuracy. In addition, we also display the spatiotemporal characteristics of brain activity patterns through topographic mapping in Fig. 3. It is worth noting that the color intensity of brain regions reflects EEG-facial expression correlation levels, with red hues indicating higher correlations and blue hues denoting lower ones. These findings validate that our method effectively identifies emotion-relevant electrodes and establishes corresponding brain networks as prior structural information, thereby enhancing model performance in emotion recognition tasks.

4 Conclusion

In this paper, we propose a novel multi-expert collaboration and knowledge enhancement network for multimodal emotion recognition. Specifically, we develop a cross-modal fusion module to integrate EEG and facial expression features. Second, feature prototype alignment is employed to enhance cross-modal consistency. Then, we also proposed a prior knowledge enhancement module that incorporates dynamic brain network topology as prior knowledge to strengthen feature representation. To reduce the uncertainty in decision fusion across different modalities, we develop a multi-expert collaborative strategy that systematically integrates complementary insights from experts across various modalities, thereby enhancing the robustness of the decision fusion results. Experimental evaluations demonstrate that our framework achieves better classification performance compared with state-of-the-art emotion recognition methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants 62136004, 62276130, 62376123, 62006115, 62076129, and 62371234, in part by the National Key RD Program of China (Grant No. 2023YFF1204803), in part by Key Research and Development Plan of Jiangsu Province under Grant BE2022842, in part by the Key Research and Development Plan in Jiangsu (Social Development) under Grant BE2022677, in part by Natural Science Foundation of Jiangsu Province (Grant No. BK20231438).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baltrušaitis, T., Robinson, P., Morency, L.P.: Openface: an open source facial behavior analysis toolkit. In: 2016 IEEE winter conference on applications of computer vision (WACV). pp. 1–10. IEEE (2016)
2. Cai, Q., Cui, G.C., Wang, H.X.: Eeg-based emotion recognition using multiple kernel learning. Machine Intelligence Research **19**(5), 472–484 (2022)
3. Chen, C., Li, Z., Kou, K.I., Du, J., Li, C., Wang, H., Vong, C.M.: Comprehensive multisource learning network for cross-subject multimodal emotion recognition. IEEE Transactions on Emerging Topics in Computational Intelligence **9**(1), 2471–285X (2025)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
5. Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., Hussain, A.: Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Information Fusion **91**, 424–444 (2023)
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
7. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing **3**(1), 18–31 (2011)
8. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. Journal of Neural Engineering **15**(5), 056013 (2018)
9. Li, C., Bian, N., Zhao, Z., Wang, H., Schuller, B.W.: Multi-view domain-adaptive representation learning for eeg-based emotion recognition. Information Fusion **104**, 102156 (2024)
10. Liu, W., Qiu, J.L., Zheng, W.L., Lu, B.L.: Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. IEEE Transactions on Cognitive and Developmental Systems **14**(2), 715–729 (2021)
11. Liu, X.H., Jiang, W.B., Zheng, W.L., Lu, B.L.: Moge: Mixture of graph experts for cross-subject emotion recognition via decomposing eeg. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3515–3520. IEEE (2024)

12. Pepa, L., Spalazzi, L., Ceravolo, M.G., Capecci, M.: Supervised learning for automatic emotion recognition in parkinson's disease through smartwatch signals. Expert Systems with Applications **249**, 123474 (2024)
13. Saganowski, S., Perz, B., Polak, A.G., Kazienko, P.: Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. IEEE Transactions on Affective Computing **14**(3), 1876–1897 (2022)
14. Song, T., Zheng, W., Song, P., Cui, Z.: Eeg emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing **11**(3), 532–541 (2018)
15. Sun, H., Niu, Z., Wang, H., Yu, X., Liu, J., Chen, Y.W., Lin, L.: Multimodal sentiment analysis with mutual information-based disentangled representation learning. IEEE Transactions on Affective Computing pp. 1–12 (2025)
16. Tan, H., Zeng, X., Ni, J., Liang, K., Xu, C., Zhang, Y., Wang, J., Li, Z., Yang, J., Han, C., et al.: Intracranial eeg signals disentangle multi-areal neural dynamics of vicarious pain perception. Nature Communications **15**(1), 5203 (2024)
17. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. Stat **1050**(20), 10–48550 (2017)
18. Wan, Z., Yu, Q., Dai, W., Li, S., Hong, J.: Data generation for enhancing eeg-based emotion recognition: Extracting time-invariant and subject-invariant components with contrastive learning. IEEE Transactions on Consumer Electronics pp. 1–14 (2024)
19. Wang, Y., Zhang, B., Tang, Y.: Dmmr: Cross-subject domain generalization for eeg-based emotion recognition via denoising mixed mutual reconstruction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 628–636 (2024)
20. Wang, Z., He, J., Liang, Y., Hu, X., Peng, T., Wang, K., Wang, J., Zhang, C., Zhang, W., Niu, S., et al.: Milmer: a framework for multiple instance learning based multimodal emotion recognition. arXiv preprint arXiv:2502.00547 (2025)
21. Wu, M., Chen, C.P., Chen, B., Zhang, T.: Grop: Graph orthogonal purification network for eeg emotion recognition. IEEE Transactions on Affective Computing pp. 1–14 (2024)
22. Wu, Y., Meng, T., Li, Q., Xi, Y., Zhang, H.: Study on multidimensional emotion recognition fusing dynamic brain network features in eeg signals. Biomedical Signal Processing and Control **100**, 107054 (2025)
23. Xu, J., Qian, W., Hu, L., Liao, G., Tian, Y.: Eeg decoding for musical emotion with functional connectivity features. Biomedical Signal Processing and Control **89**, 105744 (2024)
24. Yang, R., Modesitt, E.: Vit2eeg: leveraging hybrid pretrained vision transformers for eeg data. arXiv preprint arXiv:2308.00454 (2023)
25. Yin, J., Wu, M., Yang, Y., Li, P., Li, F., Wen, L., Lv, Z.: Research on multimodal emotion recognition based on fusion of electroencephalogram and electrooculography. IEEE Transactions on Instrumentation and Measurement **73**, 6502012 (2024)
26. Zhang, Z., Liu, Y., Zhong, S.h.: Ganser: A self-supervised data augmentation framework for eeg-based emotion recognition. IEEE Transactions on Affective Computing **14**(3), 2048–2063 (2022)
27. Zheng, C., Shao, W., Zhang, D., Zhu, Q.: Prior-driven dynamic brain networks for multi-modal emotion recognition. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 389–398. Springer (2023)