Cross-Modal Contrastive Learning for Emotion Recognition: Aligning ECG with EEG-Derived Features

Yi Wu[1*], Yuhang Chen[1*], Jiahao Cui[1], Jiaji Liu[2], Lin Liang[2], and Shuai Li[1]

[1] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Bejing, China
lishuai@buaa.edu.cn

[2] Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing Institute of Heart, Lung and Blood Vessel Diseases
liujiajiljj@163.com

Abstract. Emotion recognition plays a vital role in affective computing and mental health monitoring within intelligent healthcare systems. While EEG captures rich emotional patterns, its clinical applicability is limited by cumbersome acquisition and susceptibility to motion artifacts. In contrast, electrocardiogram (ECG) signals are more accessible and less prone to artifacts, but lack direct semantic representation of emotions categories. To address this challenge, we introduce a cross-modal alignment approach using contrastive learning. First, we extract emotional features from EEG signals using a pre-trained encoder. Then, we align the ECG encoder to these EEG-derived features through a contrastive learning framework, using sequence and patch level semantic alignment based on a temporal patch shuffle strategy. This method effectively combines the strengths of both modalities. Experiments on the DREAMER and AMIGOS datasets show that our method outperforms other baseline methods in emotion recognition tasks. Additional ablation studies and visualizations further reveal the contribution of core components. From a practical application perspective, our approach facilitates accurate emotion recognition in scenarios where EEG acquisition is impractical, providing a more accessible alternative for real-world affective computing applications. The code is available at https://github.com/pokking/ECG_EEG_alignment.

Keywords: Emotion Recognition · Contrastive Learning · Multi-Modal Alignment · ECG-EEG Integration.

1 Introduction

Emotion recognition has become a transformative tool in modern healthcare, enabling automated assessment of psychological states for mental health monitoring[10, 26]. Beyond clinical diagnostics for neurological disorders such as chronic

[*] These authors contributed equally.

sleep disturbances[3], these systems support personalized treatment strategies and continuous tracking of affective states. They are particularly valuable in monitoring conditions like exertion-induced fatigue[21] and acute pain[15].

Current emotion recognition systems primarily rely on two physiological signals: electroencephalography (EEG), which captures direct neural correlates of emotion processing, and electrocardiography (ECG), which reflects autonomic nervous system (ANS) activation. EEG offers high temporal resolution and well-established spectral biomarkers, such as θ oscillations linked to positive emotions[18] and β /γ activity associated with emotional variations [1]. EEG feature extraction typically focuses on power spectral density (PSD) [25, 9] and differential entropy (DE) [8], with classification performed using machine learning techniques. Advances in deep learning, including convolutional neural networks (CNNs)[13, 12] and attention mechanisms[22, 7], have significantly enhanced EEG-based emotion recognition. Meanwhile, ECG provides insights into cardiac dynamics associated with emotional states and has gained traction in deep learning applications for affective computing[19].

Despite EEG's precision in decoding emotions, its practical implementation faces significant challenges. Complex electrode setups introduce motion artifacts [5], and signal instability limits reliability in real-world applications such as telemedicine and virtual reality. In contrast, ECG is easier to acquire and more resilient to noise but lacks the emotion-specific cortical information that EEG provides. Recent research highlights the synchronization between the EEG and ECG during emotional experiences [2], suggesting the potential for ECG to infer emotional states more effectively. While previous studies have combined ECG with imaging techniques [4] or EEG with other signals [23] to apply disease diagnosis, few have explored direct temporal alignment between these complementary biosignals in emotion classification tasks.

In this paper, we propose a contrastive learning framework that enhances ECG-based emotion recognition by transferring EEG's semantic richness while maintaining ECG's practical advantages. Our key contributions comprise:

1. We introduce a contrastive learning-based framework for cross-modal semantic alignment, enabling ECG to carry the semantic information from EEG.
2. Leveraging the synchronized temporal characteristics of ECG and EEG data, we propose a patch-shuffling data augmentation technique and a two-level contrastive alignment strategy for multi-dimensional alignment.
3. Our method achieves superior performance on both the DREAMER and AMIGOS datasets, with ablation studies demonstrating its effectiveness and robustness.

2 Methods

2.1 Data Acquisition and Pre-process

We used the DREAMER dataset [11] from the University of the West of Scotland and the AMIGOS dataset [16] from Queen Mary University of London. The

DREAMER dataset consists of 23 subjects, each exposed to 18 video clips, with corresponding ECG and EEG data recorded. The ECG data comprises two channels, sampled at 256 Hz, while the EEG data includes 14 channels, also sampled at 256 Hz. Each subject assigns integer labels (1-5) to each data segment based on arousal, valence, and dominance. The AMIGOS dataset includes 40 subjects, who watched 16 short videos and 4 long videos, during which ECG and EEG signals were recorded. Similarly, the ECG consists of two channels sampled at 256 Hz, and the EEG has 14 channels sampled at 256 Hz. In this dataset, subjects assign floating-point labels (1-9) based on the arousal and valence rating scales.

For data preprocessing, we followed these steps: First, we removed unusable data, such as missing values, from both the DREAMER and AMIGOS datasets. The processed data was then segmented into non-overlapping 10-second intervals. Next, we applied Z-score normalization to the data from each channel. Specifically, the data from each subject's channels were concatenated, and normalization was performed individually on each channel. This approach helps to better capture common features across channels in relation to emotional changes. Finally, to ensure consistency across the datasets, we adopted a binary classification approach, categorizing the data into two classes for each emotional criterion.

2.2 Architecture

EEG Encoder Pre-training The EEG encoder architecture we use is iTransformer [14], which offers great flexibility and can adapt to varying numbers of EEG channels. To enhance the EEG encoder's ability to extract emotion-related features from EEG signals, we designed a pretraining scheme. During the pretraining phase, we first conduct reconstruction training on the EEG data, allowing the encoder to fully learn the relevant features of the EEG signals. Following this, we fine-tune the pretrained EEG encoder specifically for emotion classification tasks. This fine-tuning enables the encoder to learn how to map EEG features to distinct emotional categories, thereby improving its ability to accurately extract emotion-related features.

Contrastive learning The overall alignment framework is illustrated in the Fig. 1, which consists of three main components: Data Preparation, Embedding to Latent Space, and Semantic Alignment.

Data Preparation In the Data Preparation stage, we apply data augmentation to the preprocessed data. Given that ECG and EEG are collected simultaneously, they should exhibit similar semantic information along the temporal dimension. Therefore, we adopt shuffling as the data augmentation strategy. Specifically, both EEG and ECG sequences are segmented into non-overlapping patches of equal time steps, and these patches are then shuffled in the same order, resulting in newly arranged EEG and ECG sequences.

Formally, for the i-th EEG sequence $x_i \in R^{C\times L}$, we split it into $N_p = \lfloor \frac{L}{M} \rfloor$ patches, where each patch $p_{i,j} \in R^{C\times M}$. The same process is applied to the ECG sequence, ensuring consistency. This patch-based method aids in extracting local features and improves the model's ability to capture temporal dependencies.

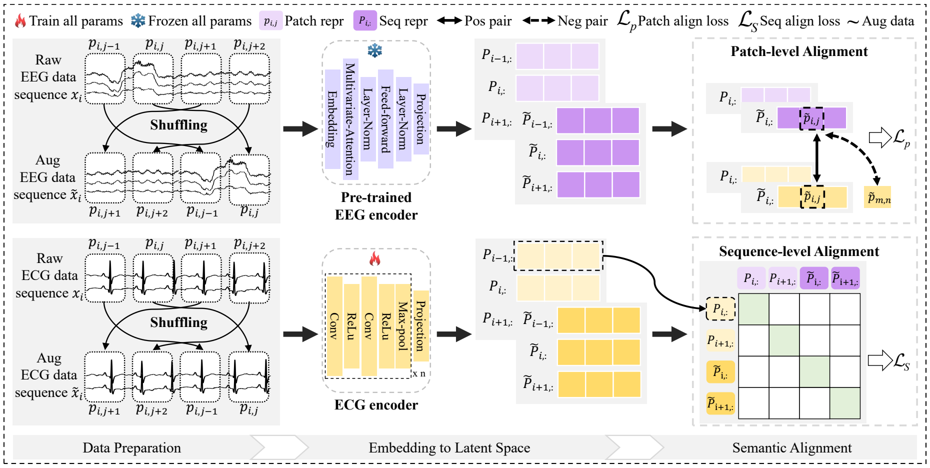


Fig. 1. Architecture of ECG and EEG Alignment: All parameters of the EEG encoder will be frozen, while the parameters of the ECG encoder are trainable, allowing the ECG features to align with the EEG features. Specially, only ECG signals are used as input during testing. Different color schemes indicate data from different modalities.

Embedding to Latent Space In the Embedding to Latent Space stage, both augmented and original EEG and ECG signals are input into the pretrained EEG encoder and non-pretrained ECG encoder to extract feature representations. Specifically, for the i-th EEG sequence x_i and its shuffled counterpart $\tilde{x}_i$, we segment the sequence into patches. The encoded feature representation of the original sequence is denoted as $P_{i,:}$, and that of the shuffled sequence as $\tilde{P}_{i,:}$, with i representing the sequence index. The same process is applied to the ECG sequences.

Semantic Alignment In the Semantic Alignment stage, positive and negative pairs are constructed based on time synchronization. Each pair consists of one EEG and one ECG feature. Only strictly synchronized EEG-ECG pairs are positive, while all others, including original and rearranged data from the same sequence, are negative. The alignment is optimized using the InfoNCE loss [17] function, defined as:

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp(\text{sim}(q, k^+)/\tau)}{\sum_{j=0} \exp(\text{sim}(q, k)/\tau)} \tag{1}$$

Where, q is the query sample, k^+ is the positive sample (semantically related to q), and k represents all contrastive samples (including k^+ and $N-1$ negative samples k^-). $sim(a, b)$ is the similarity measure between samples a and b, and τ is the temperature coefficient controlling the distribution's smoothness.

Based on the principles for constructing positive and negative pairs outlined above, we construct positive and negative pairs in both the patch and sequence dimensions and perform alignment. In the patch dimension, we divide the encoded features into non-overlapping patches, with only synchronously aligned ECG and EEG patches in the time dimension considered positive pairs. Since our goal is to align ECG and EEG such that ECG carries the semantic information of EEG, we treat $p_{i,j}^{ECG}$ as the query sample and the corresponding $p_{i,j}^{EEG}$ as the positive sample, forming a positive pair. It is important to note that, considering potential correlations between patches within the same sequence, when constructing negative pairs, patches from the same sequence (e.g., $\{p_{i,n}^{EEG}\}_{n \neq j}$) are not considered negative pairs. In contrast, EEG patches from different sequences (e.g., $\{p_{m,n}^{EEG} | m \neq i, n = 0, \ldots, N_p - 1\}$) are considered negative pairs. Therefore, the set of all contrastive samples for $p_{i,j}^{ECG}$ is $Z_{i,j}^p = p_{i,j}^{EEG} \cup \{p_{m,n}^{EEG} | m \neq i, n = 0, \ldots, N_p - 1\}$, and the contrastive learning loss in the patch dimension is as follows:

$$\mathcal{L}_p = \frac{1}{BN_p} \sum_i \sum_j -\log \frac{\exp(\mathrm{sim}(p_{i,j}^{ECG}, p_{i,j}^{EEG})/\tau_p)}{\sum_{p_{m,n}^{EEG} \in Z_{i,j}^p} \exp(\mathrm{sim}(p_{i,j}^{ECG}, p_{m,n}^{EEG})/\tau_p)} \tag{2}$$

where B represents the batch size, and N_p denotes the number of patches each sequence is divided into.

In the sequence dimension, the alignment is based on the features of the entire sequence. Similar to the alignment in the patch dimension, we treat $P_{i,:}^{ECG}$ as the query sample, $P_{i,:}^{EEG}$ as the positive sample, and the EEG features from other sequences, $\{P_{m,:}^{EEG}\}_{m \neq i}$, as negative samples. Therefore, the set of all contrastive samples for $P_{i,:}^{ECG}$ is $Z_i^S = P_{i,:}^{EEG} \cup \{P_{m,:}^{EEG} | m \neq i\}$, and the contrastive learning loss can be expressed by the following formula:

$$\mathcal{L}_S = \frac{1}{B} \sum_i -\log \frac{\exp(\mathrm{sim}(P_{i,:}^{ECG}, P_{i,:}^{EEG})/\tau_S)}{\sum_{P_{m,:}^{EEG} \in Z_i^S} \exp(\mathrm{sim}(P_{i,:}^{ECG}, P_{m,:}^{EEG})/\tau_S)} \tag{3}$$

The final contrastive loss function is as follows:

$$\mathcal{L} = \alpha \mathcal{L}_p + (1 - \alpha)\mathcal{L}_S \tag{4}$$

where $\alpha \in (0, 1)$ is a fixed scalar hyperparameter that determines the relative weights of each loss term.

Since our objective is to align ECG features as closely as possible to EEG features, we freeze the parameters of the EEG encoder during the alignment process and optimize the ECG encoder by updating its parameters using the contrastive loss.

Classification Although the pre-trained ECG encoder learns partial EEG semantics through cross-modal alignment, it lacks full modeling of emotion-related temporal-spectral features in ECG. Moreover, it cannot directly perform emotion recognition without a classifier. To address this, we attach a fully connected classifier and fine-tune both on the target dataset. By minimizing cross-entropy loss, the model extracts emotion-specific features while retaining cross-modal knowledge. The final fine-tuned model enables end-to-end emotion recognition. And it should be noted that only ECG signals are used as input at this stage.

3 Experiments and Results

3.1 Experimental Setup

We evaluate our method by randomly dividing the preprocessed DREAMER and AMIGOS datasets into five folds for training and validation. The hardware used includes an NVIDIA GeForce RTX 4090D GPU (driver version 550.67, CUDA 12.4) and an Intel(R) Xeon(R) Platinum 8474C CPU. We use PyTorch (version 1.10.1) as the deep learning framework.

3.2 Experimental Results

Our baseline selection addresses multiple aspects to ensure a thorough evaluation. First, since we adopt a temporal contrastive learning approach, we selected two relevant methods from this domain: TS-TCC [6] and TFC [24], to highlight the advantages of our method over other similar approaches. Second, to demonstrate the effectiveness of our method in emotion recognition, we chose two existing methods that focus on different modalities: SSL-ECG [19], which specializes in ECG, and EEG-Conformer [20], which focuses on EEG. The detailed experimental results are presented in Table 1 and Table 2.

Table 1. Average performance on the DREAMER dataset.

Methods	Arousal			Valence			Dominance		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
TS-TCC	0.6099	0.6037	0.6479	0.5633	0.5591	0.5902	0.6029	0.6025	0.6314
TFC	0.5407	0.5336	0.5432	0.6006	0.3990	0.5145	0.5272	0.4732	0.5264
SSL-ECG	0.8546	0.6743	0.7423	0.8256	0.6649	0.7657	0.8205	0.7194	0.7714
EEG-Conformer	0.7131	0.7545	0.7073	0.6766	0.6422	0.6233	0.7072	0.7076	0.6806
w/o align	0.8140	0.8136	0.8969	0.7924	0.7849	0.8711	0.8039	0.8022	0.8858
w/o shuffle	0.8237	0.8237	0.9035	0.8431	0.8371	0.9188	0.8786	0.9489	0.8777
w/o patch-align	0.8497	0.8494	0.9295	0.8128	0.80674	0.8928	0.8349	0.8338	0.9140
w/o seq-align	0.8859	0.8857	0.9579	0.8616	0.8578	0.9373	0.8823	0.8814	0.9529
Ours	**0.8923**	**0.8921**	**0.9611**	**0.8787**	**0.8758**	**0.9482**	**0.8866**	**0.8858**	**0.9558**

The results show that our method surpasses baseline models in accuracy, F1, and AUC on both datasets. While SSL-ECG is optimized for ECG-based emotion recognition, our approach outperforms it by leveraging EEG alignment, demonstrating that integrating EEG's semantic information enhances emotion recognition.

3.3 Ablation Study

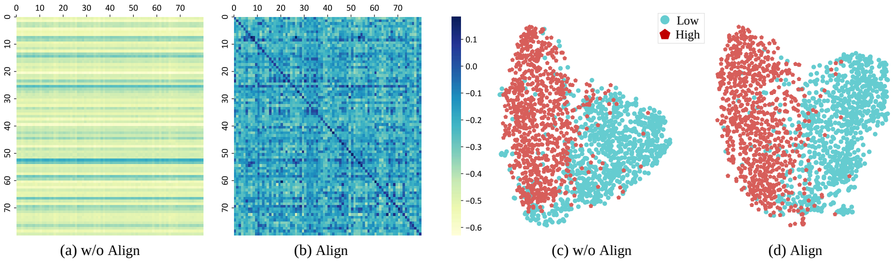
To assess the impact of each component, we conducted ablation experiments by removing alignment, shuffle, patch-level alignment, and sequence-level alignment, denoted as w/o align, w/o shuffle, w/o patch-align, and w/o seq-align. The results are presented in Table 1 and Table 2.

The results demonstrate that our method outperforms all variants, with the absence of EEG alignment having the most significant impact, thereby underscoring the crucial role of EEG semantic information in emotion recognition. Furthermore, the results of the "w/o shuffle" condition confirm that time synchronization serves as a reliable criterion for defining positive and negative pairs. Additionally, alignment at both the patch and sequence levels contributes to enhanced performance, highlighting the effectiveness of multi-scale alignment between ECG and EEG.

Table 2. Average performance on the AMIGOS dataset.

Methods	Arousal			Valence		
	ACC	F1	AUC	ACC	F1	AUC
TS-TCC	0.8110	0.5044	0.5174	0.7374	0.5624	0.6308
TFC	0.8101	0.4475	0.5377	0.7254	0.6128	0.5281
SSL-ECG	0.8472	0.7300	0.7764	0.8239	0.7669	0.8213
EEG-Conformer	0.8187	0.5481	0.6423	0.7475	0.5374	0.6072
w/o align	0.8188	0.6984	0.7197	0.7813	0.7508	0.7791
w/o shuffle	0.8278	0.6982	0.8067	0.8095	0.7526	0.8497
w/o patch-align	0.8277	0.6978	0.8124	0.8071	0.7525	0.8530
w/o seq-align	0.8327	0.7132	0.8032	0.8280	0.7767	0.8497
Ours	**0.8489**	**0.7511**	**0.8124**	**0.8341**	**0.8271**	**0.8735**

To provide a more intuitive demonstration of the architecture's effectiveness, we conducted two visualization experiments. The first visualizes the similarity matrix of ECG and EEG features before and after alignment, while the second uses t-SNE for dimensionality reduction to analyze feature distributions. As shown in parts (a), (b) of Fig. 2 , by comparing the cosine similarity of features encoded by aligned and non-aligned ECG and EEG encoders, we observe a marked improvement following alignment. For the aligned case, the cosine similarity range is [-0.4661, 0.2892], with an average of -0.1802, whereas for the non-aligned case, the range is [-0.6261, -0.2758], with an average of -0.4963. This alignment method improves the average cosine similarity by approximately 31.7%, demonstrating its efficacy in enhancing the similarity between ECG and EEG features.



(a) w/o Align (b) Align (c) w/o Align (d) Align

Fig. 2. Results of the visualization experiments.Parts (a), (b) are the similarity matrixs of w/o align and align. Parts (c), (d) are the feature distributions of w/o align and align.

As shown in parts (c), (d) of Fig. 2, the aligned encoder exhibits clearer class boundaries and more compact intra-class distributions compared to the unaligned encoder. Without alignment, the feature distributions of different classes are more entangled, class boundaries are ambiguous, and intra-class features are scattered, resulting in weaker discriminative capability. In contrast, after alignment, the clustering effect is significantly improved, inter-class separability is enhanced, and intra-class consistency is strengthened, indicating that the encoder can extract more discriminative features. This demonstrates that aligning ECG and EEG signals effectively enhances the model's classification performance and semantic coherence.

4 Conclusion

In this work, we propose a contrastive learning framework for cross-modal physiological signal alignment, enabling ECG signals to acquire EEG-like semantic representations. Our approach leverages a two-level alignment strategy (patch-level and sequence-level) and data shuffling to effectively align the semantics between EEG and ECG signals. Experiments on the DREAMER and AMIGOS datasets achieve the best performance compared to other baselines, confirming the framework's ability to align EEG semantics with ECG features. Looking ahead, we plan to extend this framework to other physiological signals and broaden its application to various tasks.

Acknowledgments. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFF1203803, and Beijing Science and Technology Plan Project Z231100005923039, and in part by the National Natural Science Foundation of China under Grant 62102036, Beijing Natural Science Foundation (L242141).

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

References

1. Aftanas, L., Reva, N., Savotina, L., Makhnev, V.: Neurophysiological correlates of induced discrete emotions in humans: an individually oriented analysis. Neuroscience and Behavioral Physiology **36**, 119–130 (2006)
2. Billones, R.K.C., Bedruz, R.A.R., Caguicla, S.M.D., Ilagan, K.M.S., Monsale, K.R.C., Santos, A.G.G., Valenzuela, I.C., Villanueva, J.P., Dadios, E.P.: Cardiac and brain activity correlation analysis using electrocardiogram and electroencephalogram signals. In: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). pp. 1–6. IEEE (2018)
3. Dahl, R.E., Harvey, A.G.: Sleep in children and adolescents with behavioral and emotional disorders. Sleep medicine clinics **2**(3), 501–511 (2007)
4. Ding, Z., Hu, Y., Li, Z., Zhang, H., Wu, F., Xiang, Y., Li, T., Liu, Z., Chu, X., Huang, Z.: Cross-modality cardiac insight transfer: A contrastive learning approach to enrich ecg with cmr features. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15003, pp. 109–119. Springer, Marrakesh (2024). https://doi.org/10.1007/978-3-031-72384-1_11
5. Egger, M., Ley, M., Hanke, S.: Emotion recognition from physiological signal analysis: A review. Electronic Notes in Theoretical Computer Science **343**, 35–55 (2019)
6. Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X., Guan, C.: Time-series representation learning via temporal and contextual contrasting. arXiv preprint arXiv:2106.14112 (2021)
7. Fan, X., Xu, P., Zhao, Q., Hao, C., Zhao, Z., Wang, Z.: A domain adaption approach for eeg-based automated seizure classification with temporal-spatial-spectral attention. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15005, pp. 14–24. Springer, Marrakesh (2024). https://doi.org/10.1007/978-3-031-72086-4_2
8. García-Martínez, B., Martinez-Rodrigo, A., Alcaraz, R., Fernández-Caballero, A.: A review on nonlinear methods using electroencephalographic recordings for emotion recognition. IEEE Transactions on Affective Computing **12**(3), 801–820 (2019)
9. Jin, M., Du, C., He, H., Cai, T., Li, J.: Pgcn: Pyramidal graph convolutional network for eeg emotion recognition. IEEE Transactions on Multimedia (2024)
10. Kamble, K., Sengupta, J.: A comprehensive survey on emotion recognition based on electroencephalograph (eeg) signals. Multimedia Tools and Applications **82**(18), 27269–27304 (2023)
11. Katsigiannis, S., Ramzan, N.: Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. IEEE journal of biomedical and health informatics **22**(1), 98–107 (2017)
12. Kumar, V., Reddy, L., Kumar Sharma, S., Dadi, K., Yarra, C., Bapi, R.S., Rajendran, S.: muleeg: a multi-view representation learning on eeg signals. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 398–407. Springer, Singapore (2022). https://doi.org/10.1007/978-3-031-16437-8_38
13. 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)
14. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625 (2023)

15. Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., Prkachin, K.M.: Automatically detecting pain in video through facial action units. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **41**(3), 664–674 (2010)
16. Miranda-Correa, J.A., Abadi, M.K., Sebe, N., Patras, I.: Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE transactions on affective computing **12**(2), 479–493 (2018)
17. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
18. Sammler, D., Grigutsch, M., Fritz, T., Koelsch, S.: Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. Psychophysiology **44**(2), 293–304 (2007)
19. Sarkar, P., Etemad, A.: Self-supervised ecg representation learning for emotion recognition. IEEE Transactions on Affective Computing **13**(3), 1541–1554 (2020)
20. Song, Y., Zheng, Q., Liu, B., Gao, X.: Eeg conformer: Convolutional transformer for eeg decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 710–719 (2022)
21. Sun, J., Han, J., Wang, Y., Liu, P.: Memristor-based neural network circuit of emotion congruent memory with mental fatigue and emotion inhibition. IEEE Transactions on Biomedical Circuits and Systems **15**(3), 606–616 (2021)
22. Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., Chen, X.: Eeg-based emotion recognition via channel-wise attention and self attention. IEEE Transactions on Affective Computing **14**(1), 382–393 (2020)
23. Zhang, D., Yuan, Z., Chen, J., Chen, K., Yang, Y.: Brant-x: A unified physiological signal alignment framework. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 4155–4166 (2024)
24. Zhang, X., Zhao, Z., Tsiligkaridis, T., Zitnik, M.: Self-supervised contrastive pretraining for time series via time-frequency consistency. Advances in Neural Information Processing Systems **35**, 3988–4003 (2022)
25. Zhao, Y., Gu, J.: Feature fusion based on mutual-cross-attention mechanism for eeg emotion recognition. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15011, pp. 276–285. Springer, Marrakesh (2024). https://doi.org/10.1007/978-3-031-72120-5_26
26. Zheng, C., Shao, W., Zhang, D., Zhu, Q.: Prior-driven dynamic brain networks for multi-modal emotion recognition. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S.E., Duncan, J., Syeda-Mahmood, T.F., Taylor, R.H. (eds.) MICCAI 2023. LNCS, vol. 14227, pp. 389–398. Springer, Vancouver, BC (2023). https://doi.org/10.1007/978-3-031-43993-3_38