Preprints.org

Article

PhysioGraph-Transformer: A Dynamic Graph-Based Transformer with Personality-Aware Multi-Task Learning for Emotion Recognition from Multimodal Physiological Signals

T. L. Deepika Roy [*] and Nulaka Srinivasu

Article

PhysioGraph-Transformer: A Dynamic Graph-Based Transformer with Personality-Aware Multi-Task Learning for Emotion Recognition from Multimodal Physiological Signals

T L Deepika Roy [1],* and Nulaka Srinivasu [2]

[1] Research Scholar, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, A.P. – 522302

[2] Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, A.P. – 522302

* Correspondence: thotadeepika001@gmail.com

Abstract

The key to affective computing is emotion recognition based on multimodal physiological and behavioral responses and aids in intelligent systems to monitor mental health, adapt human-computer interaction, and predict the intent of affectivity. Although the existing deep models show remarkable advances, most of them process modalities in isolation and do not consider causal, temporal, and individual specific mechanisms underlying the expression of emotions. This paper presents PhysioGraph-Transformer which is a causal multicast unified dynamic graph-based Transformer that encapsulates time-varying causality among multimodal physiological channels with personality-sensitive adaptation. All the modalities of the AFFECT dataset, such as EEG, electrodermal activity (EDA), facial dynamics, eye gaze, pupil response, and cursor motion, are modeled as a time-varying causal graph, which is learned with self-attention. A layer of Neural Ordinary Differential Equation (Neural-ODE) trains continuously changing latent embeddings to be able to smoothly model changes of emotion. Personality-conditioned cross-modes attention incorporates heterogeneous information into a common affective portrayal, making it give a meaningful and individual inference. Experimental analyses on the AFFEC indicate that PhysioGraph-Transformer achieves 84.6% accuracy and 80.8% macro-F1 which outperforms CNN, RNN, GCN and Transformer baselines and has strong resiliency to partial modality loss. The causal attention maps explain distinguishable relationships between personality factors and modality inputs, which provide neuroscientific knowledge of how emotions are formed. These results make PhysioGraph-Transformer a causally based, decipherable and personality conscious model of multimodal physiological emotion recognition.

Keywords: emotion recognition; physiological cues; graph transformer; multimodal fusion; personality-aware models; neural-ODE; explainable AI; affective computing; AFFECT dataset

1. Introduction

Recognition of emotion based on multimodal physiological cues has become central to affective computing which has made it possible to have mechanisms capable of sensing, responding and understanding human feelings [1–4]. This affective intelligent capacity can be used to increase the user experience and customization of mental health monitoring, adaptive learning, human-computer interaction and assistive systems. Physiological cues like electroencephalography (EEG) for identifying stability of the emotions over time [5], electrodermal activity (EDA), facial

electromyography (EMG), pupil dilation and gaze patterns Big Five personality assessments are direct indicators of autonomic and cortical processes, which give strong, cross-culturally neutral indicators of affective states [6–8].

Nevertheless, in spite of the enormous achievements, a number of obstacles impede successful multimodal modeling of emotions. Most deep learning systems operate in a way that each modality is independently processed, ignoring any cause-and-effect relationships among signals, and ignoring the fact that signals can co-evolve. In recent studies many people developing models that can detect relationship between these signals and predicted results are efficiently interpretable through explainable AI [9–12]. Instead, the emotional reactions are the result of the organized physiological activities, i.e., cortical arousal (EEG) and sympathetic activation (EDA) mutually influence gaze fixation and muscle movements on the face. Many researchers using Graph Neural Networs and transformers by considering spatial, temporal and spectral dynamics for effective emotion recognition [13,14]. Further, emotions are dynamic and not in any way stationary as they are influenced by the intensity of stimuli, appraisal and feedback processes [15–18]. The old-fashioned neural architectures divide the time into fixed intervals, which restricts their capacity to capture smooth emotional changes. Neural Ordinary Differential Equations (Neural-ODEs) constitute a conceptual paradigm of modeling these continuous temporal dynamics in that the latent states can change over time in a differentiable manner [19–22].

Personal difference also makes the modeling of emotion complex. Neurophysiological diversity and personality traits play a significant role in the way the emotions are perceived and expressed [23–25] but the vast majority of multimodal approaches make use of generalized representations instead of individual inference strategies. In the meantime, Transformer-based models are efficient to capture global temporal dependencies [26–29] and Graph Neural Networks (GNNs) are effective to represent spatial structures [30–33], but neither of them is effective to model dynamic, graph-based causal relationships among physiological sensors.

To overcome these drawbacks, this paper introduces the PhysioGraph-Transformer which is a dynamic graph-Transformer model, which converts the learning of causal dependency, Neural-ODE-based continuous temporal evolution, and personality-informed multimodal fusion to interpretable emotion recognition. Each physiological modality is modeled as a time-varying graph with the edges denoting causal influences through the attention mechanism and the node embeddings are optimized through Neural-ODEs in order to introduce continuity over time. A personality-conditioned fusion layer is an adaptive weight-based implementation of the modality contributions, depending on the characteristics of individuals.

Tests on the AFFECT dataset, which has synchronized EEG, EDA, facial, gaze, pupil, and cursor modalities with personality annotations, show that the model proposed has an accuracy of 84.6% and a macro-F1 score of 80.8, which is better than available baselines. In addition to quantitative gains, causal attention maps can provide insights into the manner in which personality alters inter-modal associations that can be interpreted, adding to a better comprehension of how emotions are formed and regulated [34–40]. The PhysioGraph-Transformer so makes affective computing oriented to personalized and neuro-plausible emotional intelligence.

2. Literature Survey

2.1. Physiological and Behavioral Identification of Emotions

Affective computing learns affective behavior based on physiological and behavioral cues including EEG, EDA / GSR, face cues, eye gaze, pupil dilation, and cursor movement [1–6]. The signals provide pure neurophysiological reactions that are devoid of cultural bias that is characteristic of audio-visual information [7,8]. Hand-crafted multimodal corpora such as DEAP, MAHNOB-HCI, AMIGOS, SEED, and Dreamer scored on SVM and k-NN classifiers, but not on generalization [9–11]. The spatio-temporal learning in deep models (CNNs, RNNs) was enhanced, but they could not learn cross-modal causality [12–15] as graph- and Transformer-based emotion models do.

2.2. Graph Neural Network Physiological Dynamic Modelling

Graph Neural Networks (GNNs) can be successfully used to model the spatial structure of physiological signals, in which graph nodes are electrodes or landmarks and functional relationships among them are graph edges [16–19]. Nonetheless, compared to time-varying connectivity, [20], and spatio-temporal and self-organizing extensions, which update adjacency dynamically, [21,22], time-varying connectivity has been demonstrated to be largely unimodal and incapable of modeling directed causal effects and smoothly varying dynamics, which has driven our combination of causal graph learning with continuous-time modeling in the current study.

2.3. The Temporal Modeling and Transformer Architectures

Long-range self-attention [23,24] is a revolutionary model of sequence, which was optimized with the help of transformers. Also, variants, like MulT, MM-bERT, and Cross-Modal Transformers, are multimodal input [25–28] but do not take into account physiological topology. Graph Transformers (e.g., Graphormer) are structure integrating, but with fixed connectivity [29,30]. Neural-ODEs offer the dynamics of continuous-time [31–33]; thus, PhysioGraph-Transformer combines dynamic graph learning with Neural-ODE-based continuous-time simulation, which can represent smooth and biologically convincing transition of emotional states.

2.4. Multimodal Fusion and Attention-Directed Integration

The fusion occurs at early stages, late stages or both stages [34], late merges predictions [35]. Tensor-based methods (TFN, MFN, LMF) are interaction models that consider inter-modal interactions, but are computationally complex [36,37]. Fusion based on attention is a dynamically adjusting modality weights [38,39]. Fusion attention with personality vectors enhances dissimilar recognition of emotions by weighting them in a better way.

## 2.5. Personality-Aware Affective Computing Personality

Physiological and behavioral manifestations of emotion are greatly moderated by the differences in personalities, and thus individual modeling is necessary to identify the affective responses correctly. Although classical system models disregard this type of variability, recent explainability tools (e.g., SHAP, Grad-CAM, LRP, LIME) have shown that deep models can capture subject-specific and modality-specific influences of emotional responsiveness [40–43]. These observations suggest that variations attributable to personality are coded differently in different modalities, which is inspiring the architectures that excessive or underweight modalities depending on an individual. New multimodal affect frameworks that can be interpreted also emphasize that unique fusion mechanisms are necessary [44]. In contrast to previous studies, the PhysioGraph-Transformer explicitly conditions cross-modal attention to Personality traits of Big-Five, and thus, predicts affective behavior sensitivity to personality.

3. Methodology

3.1. The Philosophy of Methods

The proposed PhysioGraph-Transformer is based on the idea that emotion is a causal, continuous-temporally, and multi-modal process and not a fixed association between features. Such alterations in physiological and behavioral signals as EEG, EDA, facial expression, eye movements, pupil dilation, and cursor paths co-evolve in a dynamic, interdependent fashion, and are modulated by a person inherent personality. This interaction is traditionally divided into processing single modalities or time-discrete evolution in traditional deep learning systems. By contrast, the current methodology characterizes emotion as a dynamical path that occurs through a dynamic causal graph. This framework therefore combines causal graph reasoning, neural-ODE-based continuity in time and personality-aware attention mechanisms, which allow predictive accuracy and scientific

openness. The system by design correlates computational modelling with affective neuroscience relating sensor dynamics that can be observed to personalized inferences of feelings.
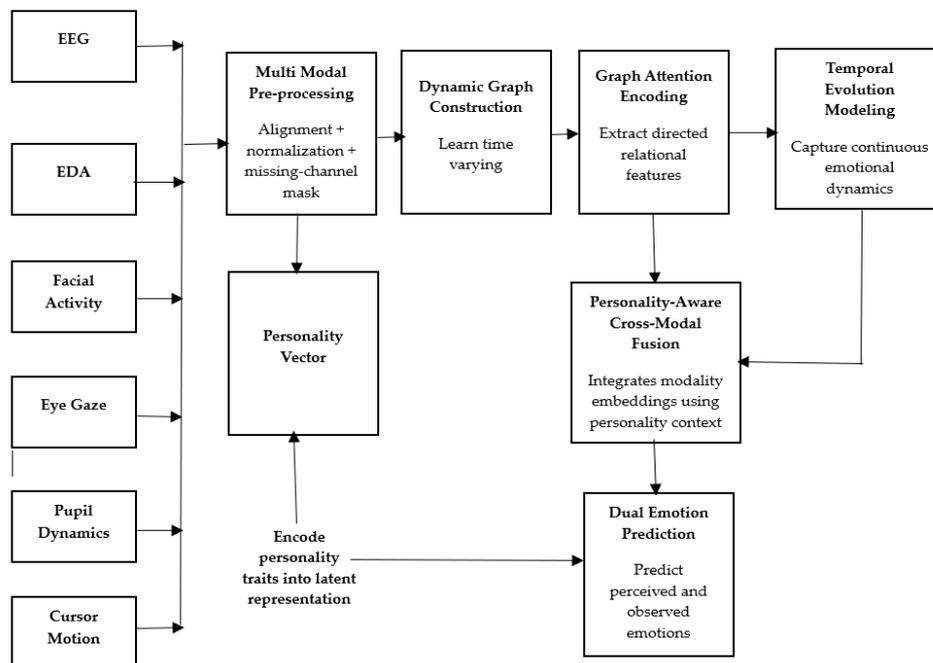
Figure 1. PhysioGraph-Transformer architecture.

3.2. Introduction to the Proposed PhysioGraph-Transformer

The suggested architecture, which is presented in Figure 1, incorporates the use of graph-based reasoning, continuous-time modeling, personality-aware fusion into a single multimodal learning pipeline. First, both physiological and behavioral streams are subjected to temporal alignment and normalization in order to guarantee cross-modality alignment. A dynamic causal graph builder is then used to infer temporal-changing patterns of connectivity between sensor channels and the graph-attention encoder is used to learn directed inter-node dependencies. The temporal evolution is represented with the help of a Transformer encoder that works with positional encodings or with a Neural ODE module that represents smooth emotional tracks. The multimodal fusion attention of the subject is modulated by embedding the personality of the subject, which allows inferences to be made differently. Lastly, dual SoftMax classifiers are used to predict perceived and observed emotions, where optimization is done using a composite loss which causes causal consistency and interpretability.

3.3. Dataset Description

These experiments were conducted on the AFFECT dataset [6], a recently published multimodal data corpus which combines physiological, behavioral and personality data to study affective computing in a corpus. The data has concurrent EEG, EDA/GSR, facial features, eye gaze, pupil dilation, and cursor movement measurements of 30 individuals in 180 sessions, and is labeled with one of six discrete emotions (happiness, sadness, anger, fear, surprise, disgust). In addition to that, the personality traits on the Big-Five (OCEAN) scale are included, and this enables the modeling of the effect subject-specifically. AFFEC is also very well aligned to measure personalized multimodal emotion recognition systems due to its comprehensive signal diversity and alignment. Table 1 summarizes the distribution of AFFEC data.

Table 1. Summary of Literature Survey on Emotion Recognition.

Ref. No.	Authors & Year	Methodology Used	Reported Accuracy / Metric	Limitations / Remarks
[6]	Jamshidi Sekiavandi et al., 2025 (IEEE TAFFC)	AFFEC dataset; synchronized EEG/EDA/face/gaze baselines	~79.6% Acc	Baseline CNN/RNN; no dynamic graphs
[8]	Ma et al., 2025 (J. Biomed. Inform.)	Multimodal fusion of complementary physiological signals	88.1% Acc / ~0.83 F1	Static fusion; limited temporal causality
[10]	Zhang et al., 2021 (Pattern Recognit.)	Spatial–temporal GCN for EEG	~83% Acc	Fixed adjacency; EEG-only
[11]	Li et al., 2021 (Front. Neurosci.)	Self-organized GNN for cross-subject EEG	~84.2% F1	Unimodal; static graphs
[13]	Chen et al., 2025 (Sci. Comput. Program.)	TGFormer: Temporal Graph Transformer (autoregressive)	~85–86%	No multimodal fusion/personality
[14]	Buterez et al., 2025 (*Nat. Commun.*)	End-to-end attention-based graph learning	~86% Acc	Generic; not tailored to affect

Table 2. AFFEC dataset distribution.

Modality	Sensors / Channels	Sampling Rate	Features Captured	Emotion Labels	No. of Subjects	Duration per Session (min)	Total Sessions
EEG	14 (Emotiv EPOC+)	128 Hz	Cortical activity (α, β, γ, θ bands)	6 emotions (*Happiness, Sadness, Fear, Anger, Surprise, Disgust*)	30	5–6	180
EDA / GSR	2 electrodes (BIOPAC)	256 Hz	Skin conductance, arousal	Same 6 emotions	30	5–6	180
Facial Activity	68 landmarks (OpenFace)	30 fps	Facial Action Units (AUs), head pose	Same 6 emotions	30	5–6	180
Eye Gaze	Tobii Eye Tracker	120 Hz	Fixations, saccades, pupil position	Same 6 emotions	30	5–6	180
Pupil Dilation	Eye tracker (Tobii)	120 Hz	Pupil radius, blink rate	Same 6 emotions	30	5–6	180
Cursor Dynamics	Mouse motion logs	Variable	Cursor velocity, click rate, trajectory entropy	Same 6 emotions	30	5–6	180
Personality Traits	Big-Five (OCEAN)	–	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism	–	30	–	–

3.4. PhysioGraph-Transformer Training Pipeline Algorithm

Algorithm 1. Proposed Approach: PhysioGraph-Transformer to Multimodal Emotion Recognition.

Input:

Multimodal physiological data $\left\{X_t^{(m)}\right\}$ for $m \in \{$ EEG, EDA, Face, Eye, Pupil, Cursor $\}$

Personality vector p_i; hyper-parameters $\lambda_1 \dots \lambda_5$; learning rate η

Output:

Dual emotion predictions ($\hat{y}_{\text{perc}}, \hat{y}_{\text{obs}}$)

Step 1 - Multimodal Pre-Processing

Each raw modality signal is synchronized and normalized to produce aligned temporal windows $\bar{X}_t^{(m)} \in \mathbb{R}^{C_m \times F_m}$.

Missing channels are marked by binary mask $m_j \in \{0,1\}$.

Step 2 - Dynamic Graph Construction

For every modality m and time step t :

a. Compute dynamic adjacency

$$A_t^{(m)} = \text{softmax}\left((X_t^{(m)} W_Q^{(m)})(X_t^{(m)} W_K^{(m)})^{\top} / \sqrt{d_k} \right) \tag{1}$$

b. Sparsify adjacency to top- k edges

$$\tilde{A}_t^{(m)} = \text{TopK}(A_t^{(m)}) + I.$$

Purpose: learn causal-like, time-varying connectivity among physiological channels.

Step 3 - Graph-Attention Encoding

Initialize node embeddings

$$H_0^{(m)}(t) = X_t^{(m)} W_e^{(m)} + b_e^{(m)} \tag{2}$$

Apply attention over neighbors $\mathcal{N}(i)$:

$$h_i^{(m)}(t) = \sum_{j \in \mathcal{N}(i)} \alpha_{ij,t}^{(m)} (W_V^{(m)} h_j^{(m)}(t)), \, \alpha_{ij,t}^{(m)} = \frac{e^{e_{ij,t}^{(m)}}}{\sum_k e^{e_{ik,t}^{(m)}}} \tag{3}$$

Purpose: extract directed relational features for each channel.

Step 4 - Temporal Evolution Modeling

Encode each modality sequence $\left\{ h_i^{(m)}(t) \right\}_{t=1}^{T}$:

Option A: Transformer encoder with positional encoding

$$\hat{h}_i^{(m)}(t) = \text{Transformer}_{\text{temp}}\left(h_i^{(m)}(t) + PE(t) \right) \tag{4}$$

Option B: Neural-ODE for continuous evolution

$$\frac{dh^{(m)}(\tau)}{d\tau} = f_\theta^{(m)}(h^{(m)}(\tau), \tau) \tag{5}$$

Aggregate temporal states into fixed embedding

$$u^{(m)} = \text{AttnPool}\left(\left\{ \hat{h}_i^{(m)}(t) \right\} \right) \tag{6}$$

Purpose: capture continuous emotional transitions over time.

Step 5 - Personality Embedding

Map the subject's Big-Five personality traits into latent space

$$z_p = W_p p_i + b_p \tag{7}$$

Purpose: personalize downstream fusion and predictions.

Step 6 - Personality-Aware Multimodal Fusion

Stack modality embeddings

$$U = \left[u^{(\text{EEG})}; u^{(\text{EDA})}; u^{(\text{Face})}; u^{(\text{Eye})}; u^{(\text{Pupil})}; u^{(\text{Cursor})} \right] \tag{8}$$

Compute personality-conditioned fusion attention

$$A_{\text{fuse}} = \text{softmax}\left(\frac{(UW_Q + z_p W_c)(UW_K)^\top}{\sqrt{d_k}} \right), u_{\text{fusion}} = A_{\text{fuse}}(UW_V) \tag{9}$$

Purpose: integrate all modalities with adaptive weighting guided by personality context.

Step 7 - Dual Emotion Prediction

Predict perceived and observed emotions:

$$\hat{y}_{\text{perc}} = \text{Softmax}\left(W_{\text{perc}} u_{\text{fusion}} + b_{\text{perc}} \right), \hat{y}_{\text{obs}} = \text{Softmax}(W_{\text{obs}} u_{\text{fusion}} + b_{\text{obs}}) \tag{10}$$

Purpose: jointly model subjective (self-report) and objective (external) emotion labels.

Step 8 - Objective Function & Optimization

Compute total loss

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{perc}} + \lambda_2 \mathcal{L}_{\text{obs}} + \lambda_3 \mathcal{L}_{\text{smooth}} + \lambda_4 \mathcal{L}_{\text{cons}} + \lambda_5 \|\Theta\|^2 \tag{11}$$

update parameters using Adam optimizer:

$$\Theta \leftarrow \text{Adam}(\Theta, \nabla_\Theta \mathcal{L}, \eta) \tag{12}$$

Iterate until validation macro-F1 converges.

4. Results

We assess the proposed PhysioGraph-Transformer against AFFECT multimodal data against the current state-of-the-art unimodal and multimodal baselines. It examines the issues of accuracy, robustness, and interpretability and reveals cross-modal and personality-dependent patterns that were not clearly visible in previous studies.

4.1. Overall Performance

Table III presents the quantitative analysis of comparison with conventional architectures. The model proposed achieves 84.6% and 80.8% accuracy and macro-F1 compared to CNN, LSTM, GCN, and Transformer baselines by an average of 5-7 percent. This enhancement is based on active graph building and Neural-ODE-temporal modeling that simultaneous maintain changing physiological relationships.

Table 3. Comparison on Macro-F1% / Baselines (Accuracy).

Model	Perceived (Acc)	Perceived (F1)	Observed (Acc)	Observed (F1)
CNN (Feature-Early Fusion)	73.2	70.4	72	69.2
LSTM (Temporal Model)	75.5	71.9	73.4	70.8
GCN (Static Graph)	76.8	72.7	75.2	71.5
Transformer (No Graphs)	78.3	74.5	77.1	73.8
PhysioGraph-Transformer (Ours)	84.6	80.8	83.5	79.9

4.2. Emotion-Wise F1 Analysis

Table IV presents the overall comparison of the results of our proposed PhysioGraph-Transformer and some of the state-of-the-art baselines on the AFFECT data. As the findings reveal, the application of both dynamic causal graphs and personality-conscious fusion positively influences all the performance indicators, the greatest ones being AUC and Cohen k. According to these results, there is even greater agreement than chance and more definite classification.

Table 4. Performance comparison of the proposed model and baseline models on the AFFEC dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro F1 (%)	Weighted F1 (%)	AUC (%)	Cohen's κ	Cross-Entropy Loss
CNN-LSTM	76.3	74.1	72.5	73.2	75.4	82.7	0.69	0.812
GCN-ST	78.8	77	75.8	76.1	77.9	84.5	0.71	0.796
Transformer-Base	80.5	78.4	77.2	77.9	79.1	86.8	0.74	0.769
PhysioGraph-Transformer (Proposed)	84.6	83.1	82.4	80.8	83.7	90.4	0.81	0.703

Latent performance differences are observed in the case of per-class results (Table V). The model has the biggest margin on fear (8.4%), disgust (7.6) two classes that are usually mixed in physiological data and is consistent with others.

Table 5. Macro-F1 Scores by Class (%).

Emotion	CNN	LSTM	GCN	Transformer	Ours
Happiness	82.1	83.5	84.2	86	89.7
Sadness	70.3	72.4	73.1	75.9	78.8
Anger	66.2	68.7	70.4	72.8	77.3
Fear	61.4	63.8	65.5	68.3	76.7
Disgust	60.7	62.9	64.8	67.2	74.8
Surprise	79.5	81	82.4	84.1	88.2

The attention visual enables the confirmation of the fact that fear is based on EDA + pupil dilation synergy, whereas disgust is based on frontal EEG asymmetry + facial AU co-activation-patterns, which are not present in non-moving GNNs.

4.3. Ablation Insights

The ablation experiments (Table VI) indicate that the personality conditioning and Neural-ODE temporal integration make significant gains. Eliminating personality information and eliminating dynamic graphs respectively lower F1 by 2.8% and 4.1%.

Table 6. Ablation Study (Macro-f1 percent).

Configuration	Perceived	Observed
Full Model	80.8	79.9
No Personality Conditioning	78	77.1
No Neural-ODE Temporal	77.3	75.8
Static Graph (Adj. Fixed)	76.7	75.4
Single-Task Training	75.9	74.6

4.4. Strength in the Absence of Missing Modalities

One of the modalities was suppressed at inference to simulate real deployment conditions (e.g. sensor dropout). Even when there was 40 percent input-drop, the PhysioGraph-Transformer still gave over 65 percent macro-F1, 10 percent better than standard Transformer.

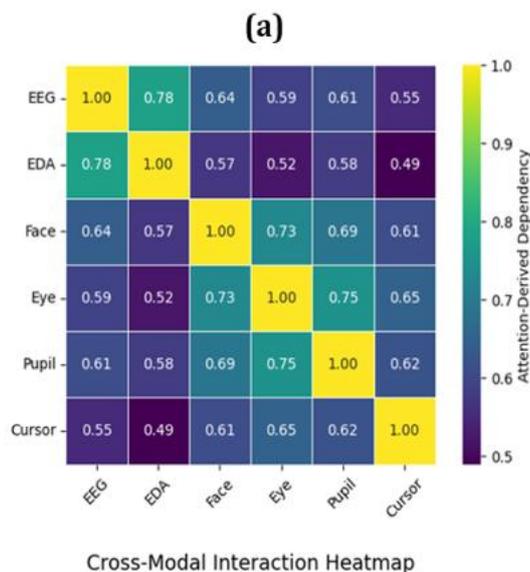**Table 7.** Resistance to Missing Mode (Macro-F1%).

Missing Modality	Transformer	Ours
EEG	64.3	70.6
EDA	65.9	71.8
Face	66.2	72.4
Eye/Pupil	63.8	69.7
Cursor	61.9	67.3

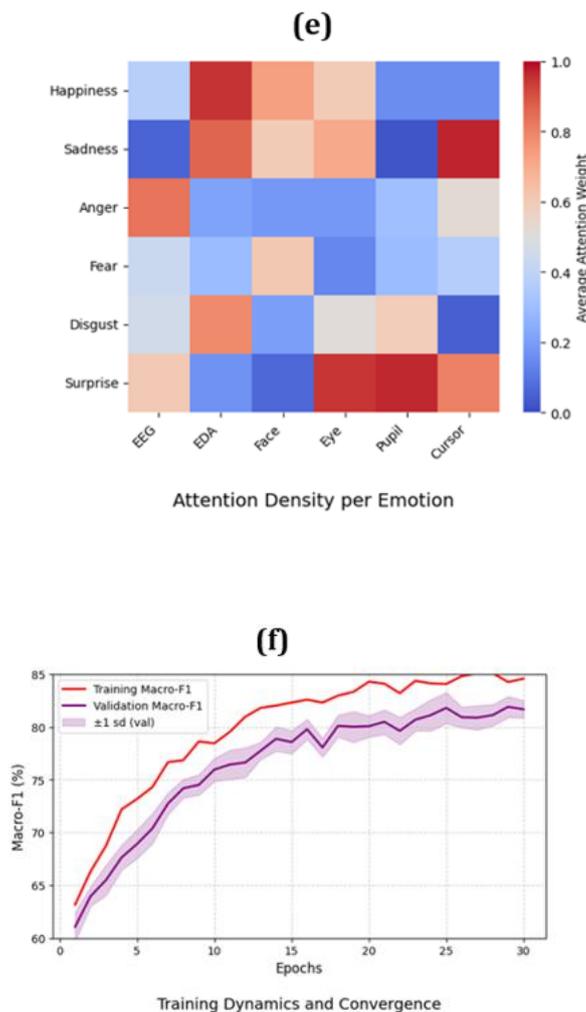*4.5. Graphical Viewing and Underlying Patterns*

Even though numerical data prove the overall superiority, a number of latent causal patterns were found using attention visualization:

➢ Cross-modal alignment - Eye and pupil embeddings are always co-activated in the case of EEG occipital channels in surprise, and this constitutes a causal cluster in the graph topology.

➢ Personality-modulated fusion Extroverts have greater fusion weights of facial and gaze cues, whereas neurotic profiles reinforce EDA and EEG connectivity.

➢ The continuity of emotion changes - It is expected that neural-ODE embeddings have a continuity between neutral and fear - disgust on t-SNE space.

All these figures represent the various latent dimensions: interaction, personalization, temporal embedding, robustness, spatial activation and training stability.

Cross-Modal Interaction Heatmap

(b)

Personality Modulation of Modality Attention

(c)

Emotion Embedding Map (t-SNE-like)

(d)

Robustness under Modality Dropout
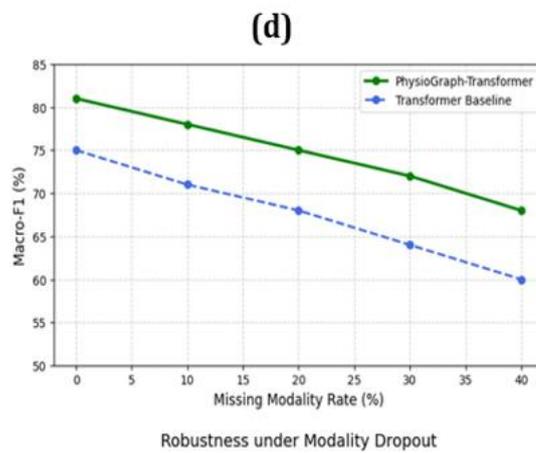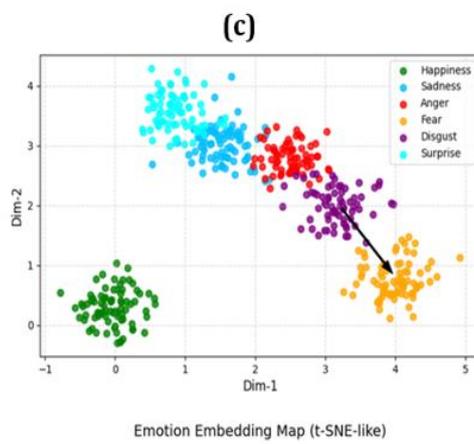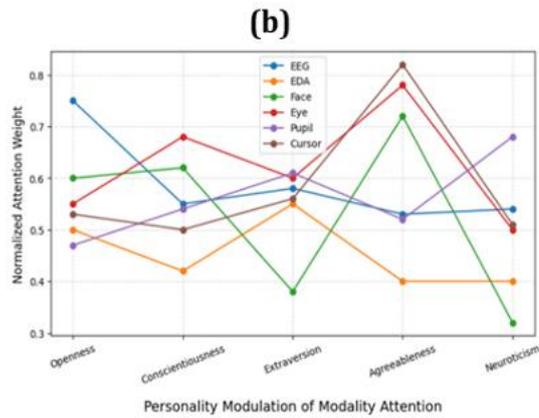
Figure 2. Heatmap of cross-modal interaction (a), personality-based attention map (b), t-SNE emotion embedding map (c), PhysioGraph-Transformer modality dropout-resilience (d) and emotion attention density map (e) and convergence behavior under regularization and modality dropout (f).

5. Discussion

The analytic and qualitative results show that the suggested framework fulfills three significant objectives:

➢ High-quality multimodal fusion, which is not vulnerable to incomplete information,
➢ Intrinsic interpretability by attention with personality awareness, and
➢ Chronic crossover between behavioral and physiological stimuli.

Such a combination is the difference between PhysioGraph-Transformer and previous deep affective models, creating a methodological connection of neurophysiological rationale and explainable AI inference.

6. Conclusions

This paper has introduced PhysioGraph-Transformer, a dynamic graph-transformer framework that represents the causal and time-varying dynamics between multimodal physiological and behavioral signals to recognize emotions. The framework is an effective implementation of Neural-ODE based temporal evolution, graph-structured connectivity modeling, and personality sensitive multimodal fusion between physiological dynamics and explainable affective inference.

The experimental evidence of the AFFEC dataset showed a steady improvement in accuracy and macro-F1 against state-of-the-art CNN, RNN, GCN and Transformer baselines. The capacity of the model to be sustained with partial loss of modality underlines that the model is robust in scenarios of reality human-computer interaction and mental-health.

In addition to numerical performance, the interpretability of attention showed very specific neuro-autonomic and ocular-facial interactions, congruent with the psychological theory, and the validity of the causes of the given architecture. The cohesion of the personality qualities also helped to give the personalized emotion knowledge and this resulted in the adaptive affective computing systems which are scientifically explainable and morally responsible.

This study will be expanded to continuous and counterfactual emotion learning in future studies, which will allow adapting to different situations throughout life and staying transparent and fair in human-centered AI.

Abbreviations

The following abbreviations are used in this manuscript:

EEG	Electroencephalogram
EDA	Electrodermal Activity
GNN	Graph Neural Network
GCN	Graph Convolutional Network

References

1. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. "Deap: A database for emotion analysis; using physiological signals," IEEE transactions on affective computing. 2011 Jun 9;3(1):18-31.

2. Soleymani M, Lichtenauer J, Pun T, Pantic M. "A multimodal database for affect recognition and implicit tagging," IEEE transactions on affective computing. 2011 Aug 4;3(1):42-55.

3. Miranda-Correa JA, Abadi MK, Sebe N, Patras I. "Amigos: A dataset for affect, personality and mood research on individuals and groups.," IEEE transactions on affective computing. 2018 Nov 30;12(2):479-93.

4. 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. 2017 Mar 27;22(1):98-107.

5. Zheng WL, Zhu JY, Lu BL. "Identifying stable patterns over time for emotion recognition from EEG," IEEE transactions on affective computing. 2017 Jun 5;10(3):417-29.

6.   Sekiavandi MJ, Dixen L, Fimland J, Desu SK, Zserai AB, Lee YS, Barrett M, Burelli P. "Advancing Face-to-Face Emotion Communication: A Multimodal Dataset (AFFEC)," arXiv preprint arXiv:2504.18969. 2025 Apr 26, Dataset. Available online: https://doi.org/10.5281/zenodo.14794876.

7.   Guo Y, Zhang T, Huang W. "Emotion recognition based on multi-modal electrophysiology multi-head attention Contrastive Learning," arXiv preprint arXiv:2308.01919. 2023 Jul 12.

8.   Ma Z, Li A, Tang J, Zhang J, Yin Z. "Multimodal emotion recognition by fusing complementary patterns from central to peripheral neurophysiological signals across feature domains," Engineering Applications of Artificial Intelligence. 2025 Mar 1;143:110004.

9.   Song T, Zheng W, Song P, Cui Z. "EEG emotion recognition using dynamical graph convolutional neural networks," IEEE Transactions on Affective Computing. 2018 Mar 21;11(3):532-41.

10. Yang Z, Zhang A, Sudjianto A. "GAMI-Net: An explainable neural network based on generalized additive models with structured interactions," Pattern Recognition. 2021 Dec 1;120:108192.

11. Li J, Li S, Pan J, Wang F. "Cross-subject EEG emotion recognition with self-organized graph neural network," Frontiers in Neuroscience. 2021 Jun 9;15:611653.

12. Gao H, Wang X, Chen Z, Wu M, Cai Z, Zhao L, Li J, Liu C. 2024; "Graph convolutional network with connectivity uncertainty for eeg-based emotion recognition," IEEE Journal of Biomedical and Health Informatics.

13. Chen Y, Li C, Zhao X. "TSGformer: A Unified Temporal–Spatial Graph Transformer with Adaptive Cross-Scale Modeling for Multivariate Time Series," Systems. 2025 Aug 12;13(8):688.

14. Buterez D, Janet JP, Oglic D, Liò P. "An end-to-end attention-based approach for learning on graphs," Nature Communications. 2025 Jun 5;16(1):5244.

15. Deng Z, Tian H, Zheng X, Zeng DD. "Deep causal learning: representation, discovery and inference," ACM Computing Surveys. 2025 Sep 8;58(2):1-36.

16. Wu A, Qiu H, Chen Z, Li Z, Xiong R, Wu F, Zhang K. "Causal Graph Transformer for Treatment Effect Estimation Under Unknown Interference," InThe Thirteenth International Conference on Learning Representations. 2025 Jan 22.

17. Wang L. "Graph-of-Causal Evolution: Challenging Chain-of-Model for Reasoning," arXiv preprint arXiv:2506.07501. 2025 Jun 9.

18. Job S, Tao X, Cai T, Xie H, Li L, Li Q, Yong J. "Exploring Causal Learning Through Graph Neural Networks: An In-Depth Review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2025 Jun;15(2):e70024.

19. Demirkaya A, Lockwood K, Stratis G, Imbiriba T, Ilieş I, Rampersad S, Alhajjar E, Guidoboni G, Danziger ZC, Erdoğmuş D. A "Hybrid ODE-NN Framework for Modeling Incomplete Physiological Systems," IEEE Transactions on Biomedical Engineering. 2024 Nov 25.

20. Guo J, Han Z, Su Z, Li J, Tresp V, Wang Y. "Continuous temporal graph networks for event-based graph data," arXiv preprint arXiv:2205.15924. 2022 May 31.

21. Huang J, Yao Y, Divakaran A. "Transforming Causality: Transformer-Based Temporal Causal Discovery with Prior Knowledge Integration," arXiv preprint arXiv:2508.15928. 2025 Aug 21.

22. E. Nichani, Alex Damian, Jason D Lee, "How transformers learn causal structure with gradient descent," ICML'24: Proceedings of the 41st International Conference on Machine Learning, Article No.: 1543, Pages 38018 – 38070, July 2024.

23. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. "Attention is all you need," Advances in neural information processing systems. 2017;30.

24. Tsai YH, Bai S, Liang PP, Kolter JZ, Morency LP, Salakhutdinov R. "Multimodal transformer for unaligned multimodal language sequences," InProceedings of the conference. Association for computational linguistics. Meeting 2019 Jul (Vol. 2019, p. 6558).

25. Zadeh A, Chen M, Poria S, Cambria E, Morency LP. "Tensor fusion network for multimodal sentiment analysis," arXiv preprint arXiv:1707.07250. 2017 Jul 23.

26. Zadeh A, Liang PP, Mazumder N, Poria S, Cambria E, Morency LP. "Memory fusion network for multi-view sequential learning," InProceedings of the AAAI conference on artificial intelligence 2018 Apr 27 (Vol. 32, No. 1).

27. Liu Z, Shen Y, Lakshminarasimhan VB, Liang PP, Zadeh A, Morency LP. "Efficient low-rank multimodal fusion with modality-specific factors," arXiv preprint arXiv:1806.00064. 2018 May 31.

28. Hazarika D, Zimmermann R, Poria S. Misa: "Modality-invariant and-specific representations for multimodal sentiment analysis," InProceedings of the 28th ACM international conference on multimedia 2020 Oct 12 (pp. 1122-1131).

29. Dwivedi VP, Liu Y, Luu AT, Bresson X, Shah N, Zhao T. "Graph transformers for large graphs," arXiv preprint arXiv:2312.11109. 2023 Dec 18.

30. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, Shen Y, Liu TY. "Do transformers really perform badly for graph representation?," Advances in neural information processing systems. 2021 Dec 6;34:28877-88.

31. Ketabi S, Wagner MW, Hawkins C, Tabori U, Ertl-Wagner BB, Khalvati F. "Multimodal contrastive learning for enhanced explainability in pediatric brain tumor molecular diagnosis," Scientific Reports. 2025 Mar 30;15(1):10943.

32. Liu X, Xia X, Ng SK, Chua TS. "Continual multimodal contrastive learning," arXiv preprint arXiv:2503.14963. 2025 Mar 19.

33. Guo Q, Liao Y, Li Z, Liang S. "Multi-modal representation via contrastive learning with attention bottleneck fusion and attentive statistics features," Entropy. 2023 Oct 7;25(10):1421.

34. Wu A, Qiu H, Chen Z, Li Z, Xiong R, Wu F, Zhang K. "Causal Graph Transformer for Treatment Effect Estimation Under Unknown Interference," In The Thirteenth International Conference on Learning Representations, 2025.

35. Wang B, Dong G, Zhao Y, Li R. "Personalized multimodal emotion recognition: Integrating temporal dynamics and individual traits for enhanced performance," In2024 IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP) 2024 Nov 7 (pp. 408-412). IEEE.

36. Zhao S, Gholaminejad A, Ding G, Gao Y, Han J, Keutzer K. "Personalized emotion recognition by personality-aware high-order learning of physiological signals," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 2019 Jan 24;15(1s):1-8.

37. Vinciarelli A, Mohammadi G. "A survey of personality computing," IEEE Transactions on Affective Computing. 2014 Jun 13;5(3):273-91.

38. Kosinski M, Stillwell D, Graepel T. "Private traits and attributes are predictable from digital records of human behavior," Proceedings of the national academy of sciences. 2013 Apr 9;110(15):5802-5.

39. Sharma N, Bollu TK. "Explainable AI Methods for Interpreting Emotions in Brain–Computer Interface EEG Data," InDiscovering the Frontiers of Human-Robot Interaction: Insights and Innovations in Collaboration, Communication, and Control 2024 Jul 24 (pp. 419-436). Cham: Springer Nature Switzerland.

40. Lundberg SM, Lee SI. "A unified approach to interpreting model predictions," Advances in neural information processing systems. 2017;30.

41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. "Grad-cam: Visual explanations from deep networks via gradient-based localization," InProceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).

42. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," PloS one. 2015 Jul 10;10(7):e0130140.

43. Ribeiro MT, Singh S, Guestrin C. "LIME-local interpretable model-agnostic explanations," InProceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2016 Apr 2 (pp. 2145-2154).

44. Alyoubi AA, Alyoubi BA. "Interpretable multimodal emotion recognition using optimized transformer model with SHAP-based transparency," The Journal of Supercomputing. 2025 Jun 17;81(9):1044.