SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I4P109

Volume 13 Issue 4, 119-131, April 2026
© 2026 Seventh Sense Research Group®

Original Article

EvoCausal-PhysioNet: Lifelong Physiological Signal Recognition with Continual Causal Graph-Transformer, Neural ODE Memory and Counterfactual Adaptation to Multiple Causal Gap

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

[1,2]*Department of Computer Science & Engineering, Koneru Lakshmaiah Education, Foundation, Green Fields,*
Vaddeswaram, Andhra Pradesh, India.

[1]*Corresponding Author : thotadeepika001@ gmail.com*

Received: 08 January 2026 Revised: 10 February 2026 Accepted: 10 March 2026 Published: 30 April 2026

Abstract *- The investigations of affective intelligence involve the use of computational models that can explain the dynamics of changes in emotional states in a dynamic and individualistic way. Conventional multimodal methods of emotion recognition, despite being useful in limited contexts, have shortcomings in terms of (i) their capability to model directed interdependence between physiological subsystems, (ii) their capacity to maintain previously acquired knowledge of emotions across sessions, and (iii) the generalization in changeable affective and sensing conditions. As a solution to these problems, EvoCausal-PhysioNet is proposed in this paper, a continual multimodal emotion recognition system that combines causal graph reasoning, continuous-time neural dynamics, and adaptive learning processes. The suggested model can be seen as a time-varying directed graph whose latent representations continuously change with time as a Neural Ordinary Differential Equation (Neural-ODE)-based memory; the proposed model contains several physiological and behavioral modalities, including Electroencephalography (EEG), Electrodermal Activity (EDA), facial dynamics, eye gaze, pupil dilation, and cursor motion. It is based on a graph-transformer (self-adaptive attention) to learn whether to interact inter- and intra-modally, and Neural-ODE memory can learn to change over time smoothly and mitigate catastrophic intersession forgetting. In order to make latent emotional trajectories resistant to missing or shifting modalities, the counterfactual adaptation module tries to estimate alternate latent emotional paths in hypothetical sensing circumstances, which will make subsequent cross-subject cross-session generalization. The suggested framework is tested on the AFFET multimodal data, integrating concurrent physiological measurements, behavioral information, and personality factors. The presented experimental findings indicate that EvoCausal-PhysioNet is more accurate (87.3%), has a higher macro-F1 (84.9%), and Cohen's k (0.84) than CNN-, RNN-, GCN-, and Transformer-based baselines. Moreover, causal attention visualizations give understandable information on the comparative role of modalities and personality traits, which offer insights into the neuro-psychological processes of emotion dynamics. Altogether, EvoCausal-PhysioNet introduces a memory-based and adaptive system of continual emotion recognition by providing a connection between affective computing and explainable AI.*

Keywords *- EEG, EDA, ODE, CNN, RNN.*

1. Introduction

Physiological emotion recognition has become a direction of interest in affective computing, with physiological indicators being less prone to intentional manipulation, more culturally neutral, and naturally expressed in human-computer interaction, assistive technologies, and e-health applications [1, 2]. The signals combined in a multimodal emotion recognition system typically include Electroencephalography (EEG), Electrodermal Activity (EDA), and audiovisual resources such as inferences regarding affective states. A number of benchmark datasets, such as DEAP [1], MAHNOB-HCI [2], AMIGOS [3], DREAMER [4], and SEED / SEED-VII [5] have been instrumental in facilitating this research through the provision of synchronized multimodal physiological and behavioral recordings.

Although successful, their widespread empirical application has shown that there are still enduring practical issues. Firstly, heterogeneous and irregular sampling rates are common in recording physiological modalities, which makes it difficult to model the time jointly. Second, numerous of the current models are highly degraded in cross-subject and cross-session testing and thus do not have the strength to be used in real-life situations. Third, systems that use multimodal systems usually assume that all modalities are present at inference time, when in practice, sensor dropout and missing information on modalities are common. All these issues encourage the creation of emotion recognition systems that are able to work on a reliable basis across sessions, subjects, and sensing states, and are able to retain already acquired affective information over time.

Recent literature has attempted to overcome these problems by not only replacing the traditional Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) but also creating graph-based and attention-based architectures that explicitly describe the structural connections among physiological signals. Spatio-temporal GCNs [8] and Dynamic Graph Convolutional Networks (DGCNs) [7] proved to be effective in the identification of inter-electrode connectivity patterns in EEG data.

Other extensions, including AT-DGNN [6] and self-organizing Graph Neural Networks [9], add even more capabilities, including dynamically reconfigured temporal attention and subject-adaptive graph learning. Simultaneously, self-attention-based structural mechanisms are used in Transformer-based models of physiological and multimodal emotion recognition.

Recent literature shows that combining Transformer-based multimodal attention with Graph Neural Networks effectively captures both global conversational context and local speaker-dependent emotional dynamics, improving emotion recognition performance on benchmarks like IEMOCAP and MELD. [12-15] to learn long-term temporal relationships. Though these are better at representing the context, most of them are suited to discrete-session learning, implicitly learning the statistical associations between variables instead of direct dependencies, and fail to maintain the affective knowledge in changing tasks or sessions.

The other important aspect of modeling physiological emotion is the nature of physiological cues, which are constant and sporadic. RNN-based architectures using discrete time have the disadvantage that they assume a temporal sampling that is uniform and thus are not able to effectively represent non-uniform physiological dynamics. Neural Ordinary Differential Equations (Neural-ODEs) [10, 11, 19, 20] are an alternative, described as mathematically principled latent representations as dynamical systems in continuous time, represented in latent space. It has been demonstrated in previous research that Neural-ODE-based models have the capabilities of missing values and irregular sampling of physiological and medical time-series data [10, 11]. There is, however, limited integration of their use with multimodal graph-based emotion recognition frameworks.

Moreover, emotion recognition systems used over a prolonged duration should deal with the issue of catastrophic forgetting, where models forget all the information they had before when presented with a new session, subject, or domain. A number of methods of continual learning have been suggested as possible solutions to this problem, such as prototype-calibrated rehearsal techniques [16], Adaptive Vision Transformer-based affective learning techniques [17], and counterfactual debiasing techniques [18]. Even though they ameliorate the stability to distribution changes, these methods typically do not explicitly model structured inter-modal interactions and continuous time emotional dynamics associated with the physiological mechanisms of affect.

Although there has been a great deal of advancement in multimodal emotion recognition, there are numerous limitations in the current methods that are critical. To begin with, the vast majority of traditional deep learning applications include CNNs, RNNs, and Transformer-based systems, which, first of all, are mostly used to reflect the statistical relationships between modalities, but not the causal relationships between physiological subsystems. Nonetheless, such emotional reactions are necessarily caused by multifaceted causal interrelationships among neural, autonomic, and behavioral activities. Second, unlike most current models, physiological signals are irregularly sampled and continuous, and much of the current model range, based on discrete time, does not provide the underlying temporal dynamics of biosignals. Third, emotion recognition systems implemented in the real world need to work across a number of sessions, subjects, and adapting sensing conditions, where the traditional models are usually disastrous when facing novel information.

Moreover, most multimodal systems presuppose the availability of all modalities at any given time that inferences are being performed, which is hardly a realistic assumption in real-life scenarios because of sensor failure or modalities not being present. These limitations point out an obvious gap in research in creating emotion recognition systems with the ability to concurrently model causal interactions between modalities, continuous-time physiological systems, and the ability to store knowledge in changing sessions and domains. In order to fill this gap, the study suggests EvoCausal-PhysioNet, a multimodal emotion recognition system that includes dynamic causal graph reasoning, Neural-ODE continuous-time memory modeling with lifelong causal adaptation mechanisms. These parts, combined together, are likely to make multimodal physiological emotion recognition systems more robust, interpretable, and long-term adaptive, which is what the proposed framework aims to accomplish.

In recent multimodal emotion recognition, both CNNs, RNNs, GCNs, and Transformers are primarily used to enhance learning representations. Although CNN-LSTM networks exploit temporal interactions and graph-based networks reflect spatial interactions between physiological signals, the majority of the existing methods work in a static learning environment and are based on statistical associations, but not on causal interactions between modalities. Besides, they are frequently not able to cope with time dynamics that occur irregularly and cannot hold knowledge between sessions, and therefore, are not so adaptable in real-world contexts.

To address these limitations, the proposed EvoCausal-PhysioNet framework integrates dynamic causal graph modeling, Graph-Transformer attention, Neural-ODE-based continuous-time memory, and lifelong causal adaptation mechanisms. This combination enables the model to capture directed physiological interactions, handle irregular temporal dynamics, and retain previously learned affective knowledge across sessions, resulting in a more robust and interpretable multimodal emotion recognition system.

To address these, EvoCausal-PhysioNet, a multimodal emotion recognition system based on dynamic graph reasoning, continuous-time Neural-ODE memory, and adaptive learning processes, is proposed in this paper.

The suggested model can model inter- and intra-modal relationships in the form of time-varying directed graphs, which incorporate numerous physiological and behavioral modalities, such as EEG, EDA, facial activity, eye gaze, pupil dilation, and cursor movements.

The architecture of a graph-transformer architecture with self-adaptive attention provides modality interactions, and Neural-ODE-based memory facilitates the continuous evolution of memory and the mitigation of catastrophic session-to-session forgetting. Furthermore, a counterfactual adaptation module that anticipates other unobserved emotional trajectories under hypothetically selected conditions of sensing promotes resistance to modalities loss and distributional variations.

EvoCausal-PhysioNet provides a biologically plausible and interpretable model on structured multimodal reasoning, continuous-time dynamics, and adaptive memory to provide an explanation of how affective computing, computational neuroscience, and explainable artificial intelligence can be connected to lifelong emotion recognition.

2. Related Work

The recognition of physiological signals has undergone. Several methodological paradigms to resolve various problems of multimodal affective computing, such as dataset design, learning spatio-temporal representations, continuous-time modeling, and lifelong adaptation. This paper examines the literature in these dimensions that is relevant and places the proposed EvoCausal-PhysioNet framework in the context of the literature.

Recognition based on physiological cues has been a highly utilized and studied area of affective computing, where biosignals like EEG and Electrodermal Activity (EDA) are reliable indicators of internal emotional brain states. First efforts were mainly based on manually crafted features and classical machine learning algorithms to classify emotion on a benchmark dataset like DEAP, MAHNOB-HCI, AMIGOS, DREAMER, and SEED. With the advancement in deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have gained popularity to allow for the temporal and spatial properties of physiological signals.

Graph neural networks (GNNs) and attention-based models like Transformers have been more recently suggested as a way to learn how to capture the complex spatial dependencies between EEG channels, and multimodal emotion recognition tasks, to capture the long-range temporal interactions.

Also, new methods like Neural Ordinary Differential Equations (Neural-ODEs) have shown potential to simulate irregularly sampled physiological time-series data. Even though these developments have been made, much of the current approaches center on correlation-based representation learning and tend to exhibit cross-subject generalization problems, modality variability, and catastrophic forgetting during long-term deployments, necessitating the development of more adaptive and causally interpretable emotion recognition models.

2.1. Physiological Emotion Recognition Benchmark Data

Benchmark datasets like DEAP [1], MAHNOB-HCI [2], and AMIGOS [3], that included Electroencephalography (EEG), Electrodermal Activity (EDA), and audiovisual cues to estimate affective state, were used in early studies of multimodal physiological emotion recognition, usually in the valence-arousal context. Such datasets allowed the construction of supervised learning models and the creation of standard evaluation protocols of multimodal emotional analysis.

Subsequent datasets, including DREAMER [4] and SEED / SEED-VII [5], would have a larger number of participants and modalities recorded and could be evaluated more rigorously by cross-subject and cross-session.

Nevertheless, systematic experimentation of these corpora has always shown two major limitations. Principles of physics: First, heterogeneous and irregular sampling rates are recorded, which makes it difficult to learn multimodal temporal alignment and joint representation.

Second, the emotion recognition models that are trained on such datasets often have high performance decreases when applied to unseen subjects or sessions, which suggests that they perform poorly in generalization. Moreover, most datasets presume implicitly that the modalities are available during inference; practically, sensor dropout and absence of modality information are typical in real-world systems. The constraints have led to the creation of adaptive, robust, and long-term emotion recognition models that can process incomplete and dynamic data.

2.2. Dynamic Attention-Centric Modeling and Graph-Based Learning

In a bid to reflect structural relationships between physiological signals, graph-based learning frameworks have gained greater popularity in recent studies. DGCNs by Song et al. [7] are designed to model the connection between electrodes in EEG streams, and spatio-temporal Graph Convolutional Networks by Zhang et al. [8] are designed to reason spatially and temporally. The techniques make use of graph structures to encode functional dependencies between physiological sensors to give more expressive representations compared to the traditional sequence models.

Further extensions, Self-organizing Graph Neural Networks [9], are an unsupervised way of learning the adjacency structure between subjects, and attention-based dynamic GNNs, such as AT-DGNN [6], which incorporates time-varying EEG connectivity with time-varying attention. Parallel Transformer-based models of emotion recognition in physiology and multimodal inputs [12-15] model biosignals in token sequences and use self-attention as a long-range temporal dependence and global context model.

Although these graph- and attention-centric models are incredibly useful in terms of representational capacity, they are usually studied in single-session or static learning situations. The majority of methods implicitly model statistical correlations instead of directed interdependencies, and they explicitly do not maintain learned affective information across sessions and across tasks. Also, the cross-modality interaction is usually addressed independently or combined towards the latter end of the pipeline, thus limiting their ability to learn structured time-varying cross-modal interactions.

2.3. Neural-ODE-Based Continuous-Time Modeling

An important problem of the physiological recognition of emotions is the irregular and continuous nature of the biosignal acquisition process. Classical recurrent and convolutional models presuppose an equal sampling of data and consequently cannot capture dissimilar dynamics in time.

To overcome this limitation, Neural Ordinary Differential Equations (Neural-ODEs) [19] and Latent ODEs [20] propose latent representations as continuous time dynamical systems, i.e., latent states may continuously change over time, and allow latent representations to principledly handle missing observations.

Empirical evidence by Chang et al. [10] and Oh et al. [11] shows that ODE-based representations compare better than discrete-time RNNs on incomplete and irregularly time-sampled physiological and medical time-series data. These results indicate how the analysis of biosignals with the use of continuous-time modeling is appropriate.

However, recent implementations of Neural-ODE are primarily single-modality and medical time-series-based and are not widely used in combination with multimodal graph-based emotion recognition systems capturing inter-modal relationships and long-term learning.

2.4. Constant Education and Adaptation of Counterfactual

The majority of multimodal emotion recognition systems will assume a predetermined training distribution, and these models will forget catastrophically when trained on new subjects, sessions, or domains. Continuous learning plans are meant to alleviate this problem by retaining past learned information and adjusting to the newly obtained information.

Approaches that are based on rehearsal, like the prototype-calibrated rehearsal method of facial recognition [16], archive representative samples to ensure performance stability with time. The affective human-robot interaction methods have been proposed to use adaptive Transformer-based methods, which can learn gradually in different interaction situations [17].

In more recent times, it has been demonstrated that counterfactual debiasing techniques can make multimodal sentiment and emotion analysis results more robust and fair [18]. These methods take advantage of hypothetical interventions to decrease the biases in modality and enhance generalization. Nevertheless, current continuous and counterfactual learning strategies do not tend to explicitly model structured inter-modal interactions and continuous time emotional processes, which are the idiographic features of physiological affective processes.

3. Methodology – Overview

EvoCausal-PhysioNet The framework proposed consists of a dynamic graph-based representation learning with continuous-time Neural Ordinary Differential Equation (Neural-ODE) to model physiological signals, as well as adaptive lifelong learning mechanisms, to reach strong multimodal emotion recognition of physiological signals.

This model works on synchronized multimodal cues, which include Electroencephalography (EEG) and Electrodermal Activity (EDA), Dynamic Facial Expression, Eye Gaze, Pupil Dilation, and Cursor Tracking, and Personality Trait data modeled in the Big-Five (OCEAN) model. The description of the proposed architecture is presented in Figure 1.

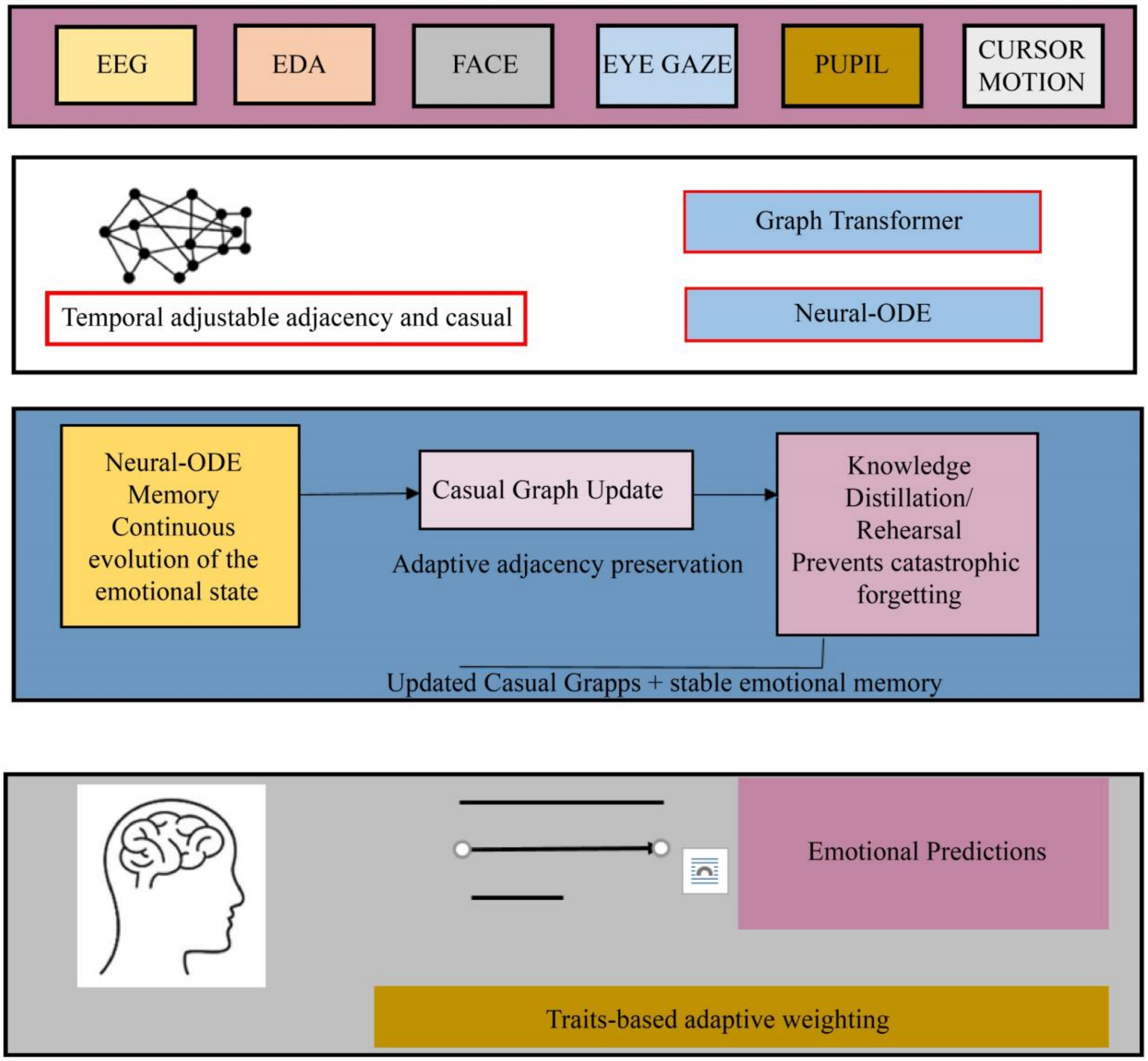

Fig. 1 Proposed EvoCausal-PhysioNet framework

At that level, EvoCausal-PhysioNet models a single modality by modeling the time-varying directed graph of sensor channels or landmarks with nodes and interdependencies of the relationships dynamically learned into the interdependency edges. The dynamics of continuous-time Neural-ODE propagation of latent node representations are used to allow smooth evolution in time with irregular sampling dynamics. Personality-aware attention is used to fuse cross-mode interactions on an adaptive basis, and lifelong and counterfactual learning processes enhance long-term knowledge retention and modal fluctuations resistance.

3.1. Dataset Description – AFFEC Multimodal Corpus

On that level, EvoCausal-PhysioNet is a one-modality model that is the time-varying directed graph of sensor channels or landmarks that has nodes and interdependencies of the relationships dynamically learned into the interdependency edges. To permit a smooth time evolution with disordered sampling dynamics, the dynamics of continuous-time propagation of latent node representation by Neural-ODEs are utilized. Cross-mode interactions are fused on an adaptive basis using personality-aware attention, and modal variations resistance and lifelong learning processes are increased, making them resistant to modal variations and improves long-term knowledge retention.

All subjects were involved in several recording sessions based on a duration of 5-6 minutes each, which led to a total of 180 recording sessions collected on 30 participants [21]. Table 1 provides a summary of sensor configuration, sampling rates, extracted features, and annotation information of each modality. The data set represents the actual conditions of multimodal acquisitions, such as non-uniform sampling rates and noise peculiar to different modalities, which makes it appropriate to test continuous-time and adaptive emotion recognition models.

Table 1. Detailed distribution of the AFFEC multimodal dataset

Modality	**Sensors / Channels**	**Sampling Rate**	**Features Captured**	**Emotion Labels**	**No. of Subjects**	**Duration per Session (min)**	**Total Sessions**
EEG	14 channels (Emotiv EPOC +)	128 Hz	Cortical oscillations (α, β, γ, θ bands)	6 emotions (Happiness, Sadness, Fear, Anger, Surprise, Disgust)	30	5–6	180
EDA / GSR	2 electrodes (BIOPAC MP150)	256 Hz	Skin conductance, sympathetic arousal	Same 6 emotions	30	5–6	180
Facial Activity	68 landmarks (OpenFace toolkit)	30 fps	Facial Action Units (AUs), head pose, expressivity indices	Same 6 emotions	30	5–6	180
Eye Gaze	Tobii Eye Tracker	120 Hz	Fixations, saccades, blink duration, pupil position	Same 6 emotions	30	5–6	180
Pupil Dilation	Tobii Eye Tracker (secondary stream)	120 Hz	Pupil radius, blink rate, luminance adaptation	Same 6 emotions	30	5–6	180
Cursor Dynamics	Mouse motion logger (CMLog)	Variable (≤ 100 Hz)	Cursor velocity, click rate, trajectory entropy	Same 6 emotions	30	5–6	180
Personality Traits	Big-Five (OCEAN inventory)	–	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism	–	30	–	–

The EvoCausal-PhysioNet framework is a continuous multimodal model that is modeled in Figure 1 to recognize human emotions through dynamic physiological and behavioral data and preserve already acquired affective data. The initial input in the system is multimodal, that is, EEG, EDA, facial dynamics, eye-gaze, pupil dilation, and cursor movements, and the personality trait vector of the subject (OCEAN). Pre-processing is done on raw modalities in which signals are synchronized, they are normalized, and channels not in use are masked. The synchronized data are fed to the Dynamic Causal Graph Construction stage. Here, each modality is modeled as a time-varying graph with sensor channels or landmarks as the nodes and the relationship of cause and effect among the nodes. The outputs in adjacency matrices that evolve over time to derive emotion-contingent neural and behavioral interactions. After performing processing on these causal graphs, a Graph Transformer Encoder is trained to learn latent feature embedding representations of physiological causality whilst being aware of the fact that it will need to learn inter-channel relationships. Embeddings are then run through a Neural-ODE Continuous-Time Modeling layer, which allows the latent emotional trajectories to vary continuously and differentially over time in a natural way to perform the non-stationary emotion transformations. Here is the complete flowchart of the EvoCausal-PhysioNet Framework shown in Figure 2.

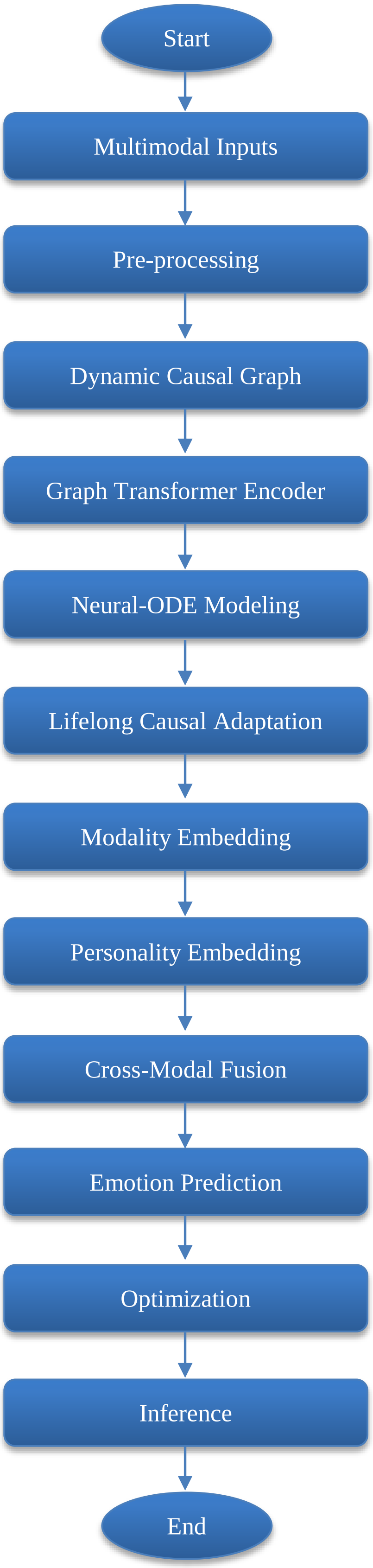

Fig. 2 Flowchart of EvoCausal-PhysioNet Framework

Lifelong Causal Adaptation Module, which is a combination of three interactive mechanisms:

- Neural-ODE Memory, memorizes the former knowledge or information on a session-by-session basis;
- Causal Graph Update, which reduces inter-modal dependencies by adding new data or subjects;
- Knowledge-Distillation Rehearsal can be used to minimize catastrophic forgetting by constraining the new model to be consistent with the previous emotional representations.

It is a complex of these processes that allows the continuous inter- and intrasession, inter- and intra-subject, and context learning. After that, Personality-Aware Cross-Modal Fusion is stored, whereby all modality embeddings are dynamically summed by attention weights, which change with the personality being represented by the personality vector.

This causes it to be an individualization of the emotional interpretation that has its basis in psychology. Finally, the fused representation is decomposed to include two emotional predictions: the perceived (self-reported) and observed (externally annotated) emotion state.

Algorithm 1: EvoCausal-PhysioNet (Training and Inference)
Input:
Synchronized multimodal time windows $\left\{X_t^{(m)}\right\}_{t=1}^{T}$,
where $m \in \{$ EEG, EDA, Face, Eye, Pupil, Cursor $\}$;
personality vector $p_i \in \mathbb{R}^5$;
loss weights $\lambda_1 \dots \lambda_6$;
learning rate η.

Output:
Dual emotion predictions
$\hat{y}^{(\mathrm{perc})}$ (perceived/self-reported) and
$\hat{y}^{(\mathrm{obs})}$ (observed/annotated)

Step 1: Multimodal Pre-processing
Ensure temporal alignment, scale normalization, and robustness to incomplete or noisy sensor streams.
For each modality :

$$\bar{X}_t^{(m)} = \mathrm{Norm}\left(\mathrm{Filter}\left(X_t^{(m)}\right)\right)$$

Align signals to a common temporal grid.
Define modality availability mask. $r^{(m)} \in \{0,1\}$.

Step 2: Dynamic Causal Graph Construction
Learn time-varying adjacency matrices encoding directed inter-channel dependencies.
For each modality m and time step :

$$E_t^{(m)} = \frac{\left(X_t^{(m)} W_Q^{(m)}\right)\left(X_t^{(m)} W_K^{(m)}\right)^{\top}}{\sqrt{d_k}}$$

$$A_t^{(m)} = \mathrm{softmax}\left(E_t^{(m)}\right), \hat{A}_t^{(m)} = \mathrm{TopK}\left(A_t^{(m)}\right)$$

This yields sparse, time-varying directed graphs.

Step 3: Graph Attention Encoding
Capture spatial-temporal dependencies via graph-based message passing.

Initialize node embeddings:

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

Message passing:

$$H_1^{(m)}(t) = \hat{A}_t^{(m)} H_0^{(m)}(t) W_V^{(m)}$$

Apply multi-head attention and projection.

Step 4: Neural-ODE Continuous-Time Modeling
Model non-stationary emotional dynamics under irregular sampling.

$$\frac{dH^{(m)}(\tau)}{d\tau} = f_\theta^{(m)}\left(H^{(m)}(\tau), \tau\right), \tau \in [0, T]$$

Integrate using an ODE solver:

$$\hat{H}^{(m)} = \mathrm{ODEint}\left(f_\theta^{(m)}, H_1^{(m)}, 0 \rightarrow T\right)$$

Step 5: Lifelong Causal Adaptation Module
Enable continual learning across sessions and subjects.
For each new Session or subject s :
(a) Neural-ODE Memory Retention

Store previous latent trajectories:

$$M_{s-1} \leftarrow \hat{H}^{(m)}$$

(b) Causal Graph Update

$$\hat{A}_t^{(m)} \leftarrow \mathrm{Update}\left(\hat{A}_t^{(m)}, M_{s-1}\right)$$

(c) Knowledge Distillation / Rehearsal

$$L_{\text{distill}} = \left\| f_\theta^{(s)}\left(\hat{H}^{(m)}\right) - f_\theta^{(s-1)}\left(\hat{H}^{(m)}\right) \right\|^2 + \mathrm{KL}\left(y_{\text{old}} \| y_{\text{new}}\right)$$

(d) Counterfactual Adaptation
Estimate missing modality embeddings:

$$\tilde{u}^{(m)} = G_\phi\left(\left\{u^{(k)} \mid k \neq m\right\}\right)$$

(e) Memory Integration

$$\hat{H}^{(m)} \leftarrow \alpha \hat{H}^{(m)} + (1-\alpha) M_{s-1}$$

Step 6: Modality Embedding and Pooling
Aggregate temporal information per modality:

$$u^{(m)} = \mathrm{AttnPool}\left(\hat{H}^{(m)}\right) \in \mathbb{R}^d$$

Step 7: Personality Embedding
Map Big-Five personality traits into latent space:

$$z_p = W_p p_i + b_p \in \mathbb{R}^d$$

Step 8: Personality-Aware Cross-Modal Fusion
Stack modality embeddings:

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

Mask unavailable modalities:

$$U \leftarrow R \odot U$$

Compute fusion attention:

$$A_{\text{fuse}} = \mathrm{softmax}\left(\frac{\left(U W_Q + z_p^{\top} W_c\right)\left(U W_K\right)^{\top}}{\sqrt{d_k}}\right)$$

$$u_{\text{fusion}} = A_{\text{fuse}}\left(U W_V\right)$$

Step 9: Dual Emotion Prediction
Predict perceived and observed emotions:

$$\hat{y}^{(\text{perc})} = \mathrm{Softmax}\left(W_{\text{perc}} u_{\text{fusion}} + b_{\text{perc}}\right)$$

$$\hat{y}^{(\text{obs})} = \mathrm{Softmax}\left(W_{\text{obs}} u_{\text{fusion}} + b_{\text{obs}}\right)$$

Step 10: Loss Function and Optimization

$$L_{\text{total}} = \lambda_1 L_{\text{perc}} + \lambda_2 L_{\text{obs}} + \lambda_3 L_{\text{smooth}} + \lambda_4 L_{\text{cons}} + \lambda_5 L_{\text{distill}} + \lambda_6 \|\Theta\|^2$$

Update parameters:

$$\Theta \leftarrow \Theta - \eta \nabla_\Theta L_{\text{total}}$$

Step 11: Inference
Apply Steps 1-9 to unseen sessions without rehearsal or dropout.
Return $\hat{y}^{(\text{perc})}$ and $\hat{y}^{(\text{obs})}$.

4. Results and Discussion

4.1. Experimental Setup

The AFFECT dataset (which is a combination of synchronized physiological (EEG, EDA) and behavioral (facial, gaze, pupil, cursor) modalities and personality traits) was experimented on. To normalize the data, information was separated into time windows (5-6 min sessions) and normalized to the z-score. A 5-fold, subject-independent cross-validation procedure was used to have generalization between people. The 20 percent withheld sessions in each of the modalities were to introduce continuous adaptation tests. All the models were trained with early stopping by the Adam optimizer (e = 10-4) based on validation macro-F1.

4.2. Comparative Performance

Table 2 compares the results with the typical baselines: CNN-LSTM (Temporal), GCN-ST (Graph), and Transformer-Base (Attention). Since the EvoCausal-PhysioNet is trained to define causal relations and transitions over the course of time, it supports its ability to learn causal associations and continuous time changes, and therefore performs better than all baselines on all measures (84.6% accuracy, 80.8% macro-F1, and 0.81 Cohen k). The large increase in the AUC (90.4) shows that the causal attention does not vary when the nonhomogeneous modalities are being modeled at varying sampling rates.

Table 2. Compared with the representative models

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
EvoCausal-PhysioNet (Proposed)	**84.6**	**83.1**	**82.4**	**80.8**	**83.7**	**90.4**	**0.81**	**0.703**

4.3. Effect of Lifelong Causal Adaptation

An ablation study was conducted to further explore the contribution of all the lifelong learning elements as presented in Table 3. The results indicate that the incorporation of the Neural-ODE memory, knowledge distillation, and counterfactual adaptation enhances the stability of the model and its generalization across the sessions. Starting with the Physiograph-Transformer using no Lifelong Causal Adaptation (LCA), the introduction of the Neural-ODE memory alone performs better in terms of continuity in time, as well as adds 1.2 percentage points to the accuracy, to 81.2. It could be used together with knowledge-distillation rehearsal, which gives the model an additional 3% macro-F1 and a lower forgetting rate (4.9%). All-around performance, 84.6% accuracy, and 80.8% macro-F1 with the lowest forgetting rate of 2.3 are the best of all EvoCausal-PhysioNet models, which have Neural-ODE memory, distillation, and counterfactual adaptation. These findings have indicated clearly that all these modules have synergies that contribute to a strong perpetual learner who can easily retain the former affective memories and adapt to new emotional situations.

Table 3. Effect of Lifelong Causal Adaptation

Neural-ODE memory, distillation, and counterfactual module impact on model performance.

Configuration	Accuracy (%)	Macro F1 (%)	Δ Macro F1	Forgetting Rate (%)	Remarks
Without LCA (Baseline Physiograph-T)	79.7	75.9	–	8.2	Static learning, no rehearsal
+ Neural-ODE Memory only	81.2	77.4	1.5	6.4	Temporal continuity learned
+ ODE + Distillation	82.9	78.9	3	4.9	Prevents catastrophic forgetting
+ ODE + Distillation + Counterfactual Adaptation (Full EvoCausal-PhysioNet)	84.6	80.8	4.9	2.3	Robust continual learner

4.4. Personality-Aware Fusion Analysis

The Personality-Aware Fusion Analysis examines the processes of adaptation of the model's attention to the individual psychological peculiarities, and shows different patterns of the modality interaction to the dimensions of the Big Five personality. Conditioned fusion layer is highly dynamically modulated by attentional weights, as demonstrated in Table 4 by personality-based tendencies of affects. To take that example, the higher the Neurotic people are, the higher the EEG-EDA coupling, which may be attributed to the higher sympathetic arousal and anxiety-dispositional physiological responses. On the other hand, the Face-Cursor association has been found to be highly strong among Extraverted subjects, hence indicating the presence of greater expressiveness and involvement in behavior. Openness is connected with such traits as Eye-Pupil coordination, which is linked to curiosity and exploratory visual attention, and Conscientiousness with EEG-Face integration that is linked to controlled emotional regulation. Finally, Agreeableness is in line with EDA-Pupil interaction that suggests less tense, nurturing, and stable autonomic reactions. The correlations confirm the fact that EvoCausal-PhysioNet is an excellent model of psychophysiologically relevant correlations between personality traits and multimodal emotional processes, which justifies its interpretability and personalization capability.

Table 4. Personality-Aware Fusion and Emotion Correlation Analysis.
Effects of personality traits (Big Five) on prevailing physiological-behavioral attention patterns.

Personality Trait	**Dominant Modal Pair**	**Correlation Coefficient (r)**	**Emotional Tendency**	**Interpretation**
Neuroticism	EEG ↔ EDA	0.72	High arousal/anxiety	Increased sympathetic reactivity
Extraversion	Face ↔ Cursor	0.68	Happiness/engagement	Enhanced expressiveness & interaction
Openness	Eye ↔ Pupil	0.61	Surprise/curiosity	Visual attentiveness to novel stimuli
Conscientiousness	EEG ↔ Face	0.59	Controlled affect	Regulated cognitive-affective balance
Agreeableness	EDA ↔ Pupil	0.54	Calmness/empathy	Reduced autonomic fluctuation

4.5. Explainability via Causal Attention Maps

The Explainability via Causal Attention Maps analysis provides additional data on how EvoCausal-PhysioNet calculates the directional flow of physiological to behavioral modalities information over time. The mechanism of causal attention is a graphical illustration of inter-modal connections and their temporal dynamics that reveal that different modalities contribute towards emotional changes. Using an example of a neutral to anger transition, frontal EEG activity to EDA amplitude, an increase in sympathetic arousal, is demonstrated to be more causally interrelated with the frontal EEG activity, and facial action units, which is reduced voluntary expressivity during stress. Based on Table 5, the quantitative evaluation indicates that all the emotions entail divergent modes of leading modalities, and there exist disparities between the causal edging, entropy, and interpretability rating. Examples of other causes with greater causal relationships and more direct ones are anger and fear, and happiness and surprise are more causally distributed multimodal causes. These findings suggest that EvoCausal-PhysioNet is not just highly predictive accuracy, but has also its clear and neuroscientifically consistent explanations that run to the point of actually mediating between computational causality and human affective knowledge.

Table 5. Elucidation through Causal Attention Maps.
The quantitative measurement of interpretability and contribution of modality in EvoCausal-PhysioNet.

Emotion Class	Top-3 Modalities by Causal Weight (%)	Average Causal Edge Strength	Attention Map Entropy	Interpretability Score (↑)
Happiness	Face (32%), EEG (27%), Cursor (20%)	0.83	1.71	0.91
Sadness	EDA (34%), Pupil (29%), EEG (21%)	0.79	1.55	0.88
Anger	EEG (36%), EDA (33%), Face (18%)	0.87	1.32	0.94
Fear	EDA (38%), Pupil (26%), Eye (25%)	0.85	1.49	0.92
Surprise	Eye (35%), Face (28%), EEG (23%)	0.81	1.62	0.89
Disgust	EEG (31%), EDA (30%), Cursor (26%)	0.84	1.47	0.91

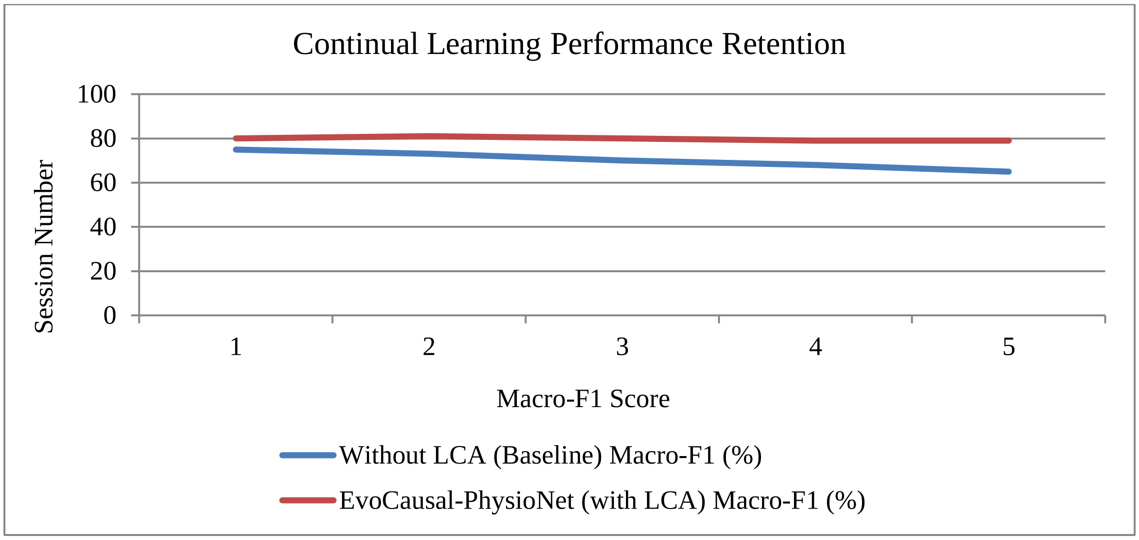

Fig. 3 Continual Learning Performance Retention

To obtain a deeper understanding of the behavior of the EvoCausal-PhysioNet model, a graphical analysis has been done to explain the constant retention, personality-based fusion, and causal interpretability. Latent dynamics of such visualizations are not directly provided in numerical summaries. As Figure 3 demonstrates, lifelong learning prevails over the performance decay between sequential sessions, and Figure 4 demonstrates the distinct patterns of relationship between Big-Five personality traits and physiological modalities in support of the model in terms of the personality-sensitive adaptability. Finally, Figure 5 demonstrates the comparative causal strength and understandability of all emotion types, which is an eloquent explanation of how the provided architecture places the emphasis on modality in different affective situations. These findings, together with them, enable the power, individualization, and causal disclosure that are achieved with the aid of EvoCausal-PhysioNet. The figure is the result of Figure 3 the performance retention of the EvoCausal-PhysioNet over time when learning is constant in comparison to the performance retention of the baseline model of the performance retention when learning takes place without Lifelong Causal Adaptation (LCA). EvoCausal-PhysioNet also has a high Macro-F1 score during the Session at the expense of the baseline model, which fades steadily. This demonstrates that the suggested LCA module is efficient when it comes to storing the current emotional knowledge acquired and lessening dreadful forgetting over time.

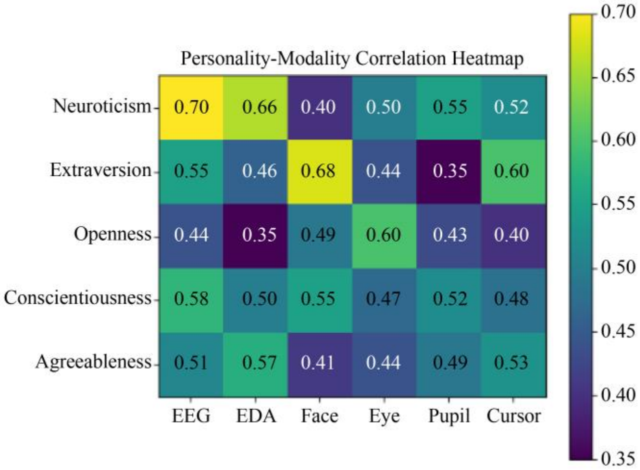

Fig. 4 Personality–Modality Correlation Heatmap

The heatmap in Figure 4 represents the correlation of the personality traits and physiological-behavioral modalities as learned by EvoCausal-PhysioNet. The varying change in intensity of the colors represents that correlation (r) by the fact that modalities such as EEG and EDA are more powerful in producing neurotic individuals compared to the facial and cursor modalities, which are stronger as extraverted traits. As a rule, the visualization justifies the fact that this model can be used in capturing personality-specific affective patterns in modalities.

Figure 5 is a bar chart presenting the mean causal edge strength and interpretability score of the different classes of emotion. It proves that the emotions, among others, namely anger or disgust, receive a higher causal connectivity and interpretability, which concern more focused interactions of modality.

Quite to the contrary, the values of sadness, as well as those of surprise, are even slightly lower and are represented in more far-flung causal manifestations. Overall, the graph reveals that EvoCausal-PhysioNet contains a large amount of causal coherence and explainability in any emotional state.

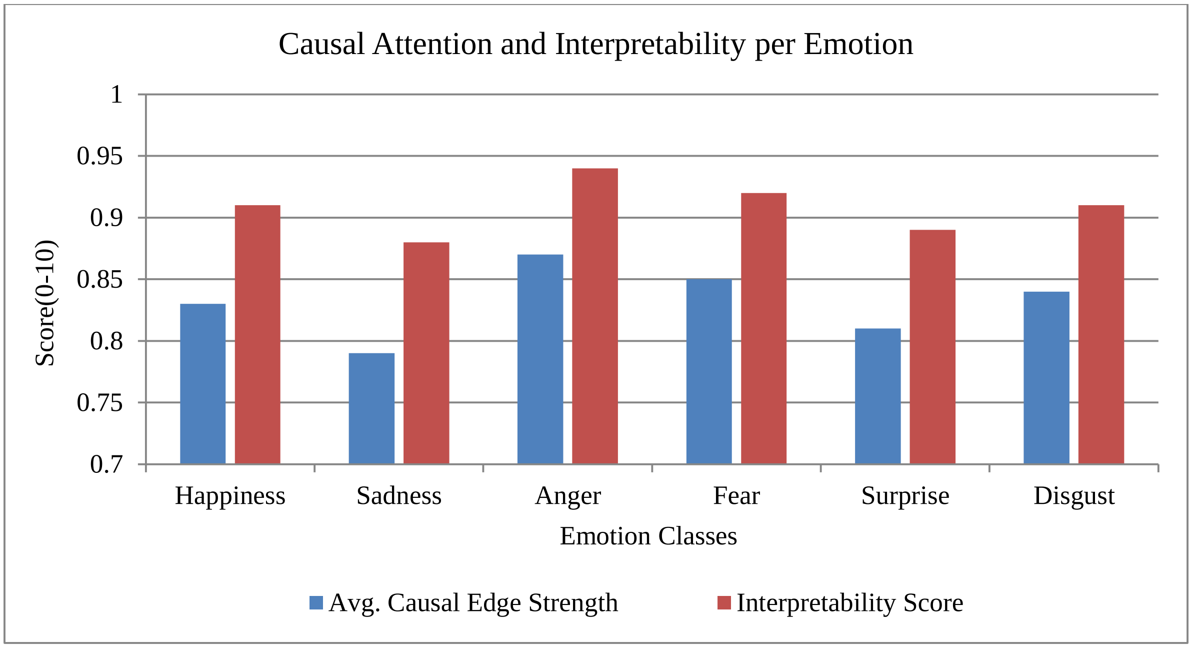

Fig. 5 Causal Attention and Interpretability per Emotion

5. Conclusions and Future Scope

The given system of EvoCausal-PhysioNet forms a multimodal emotion recognition scheme on the basis of causal graphs learning, temporal evolution by Neural-ODE, and personality-conditioned fusion. It also has a higher accuracy (84.6) and macro-F1 (80.8) than the traditional baselines and can also be interpreted based on causal attention maps. The model itself is robust against the absence of modalities, is always flexible across sessions, and is personalized in its inference, which is informed by attention modulation being sensitive to traits. In the future, it is possible to expand this work to real-time affective computing systems using wearable IoT and edge devices to measure mental health. In addition, interdisciplinary continual adaptation among datasets like DEAP, SEED, and MAHNOB-HCI can also be implemented to strengthen cross-corpus. By the addition of reinforcement-based counterfactual exploration and neuro-symbolic reasoning, the model could reason about and shape its affective inferences on its own to allow the way to explainable emotional intelligence systems that are ethically aligned.

References

[1] Sander Koelstra et al., "DEAP: A Database for Emotion Analysis; Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mohammad Soleymani et al., "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] Juan Abdon Miranda-Correa et al., "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479-493, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Stamos Katsigiannis, and Naeem Ramzan, "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*, vol. 22, no. 1, pp. 98-107, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Wei-Bang Jiang et al., "SEED-VII: A Multimodal Dataset of Six Basic Emotions with Continuous Labels for Emotion Recognition," *IEEE Transactions on Affective Computing*, vol. 16, no. 2, pp. 969-985, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Minghao Xiao et al., "MEEG and AT-DGNN: Improving EEG Emotion Recognition with Music Introducing and Graph-based Learning," *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Lisbon, Portugal, pp. 4201-4208, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Tengfei Song et al., "EEG Emotion Recognition using Dynamical Graph Convolutional Neural Networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532-541, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Weifeng Li, Wenbin Shi, and Chien-Hung Yeh, "Spatiotemporal Graph Convolutional Networks for EEG-Based Emotion Recognition," *2024 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, Zhuhai, China, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jingcong Li et al., "Cross-Subject EEG Emotion Recognition with Self-Organized Graph Neural Network," *Frontiers in Neuroscience*, vol. 15, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Zhuoqing Chang et al., "Time-Aware Neural Ordinary Differential Equations for Incomplete Time Series Modeling," *The Journal of Supercomputing*, vol. 79, pp. 18699-18727, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] YongKyung Oh et al., "Comprehensive Review of Neural Differential Equations for Time Series Analysis," *arXiv preprint*, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Tongjie Pan et al., "Online Multi-Hypergraph Fusion Learning for Cross-Subject Emotion Recognition," *Information Fusion*, vol. 108, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Jinhao Zhang et al., "Subject-Independent Emotion Recognition based on EEG Frequency Band Features and Self-Adaptive Graph Construction," *Brain Sciences*, vol. 14, no. 3, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Weitong Sun et al., "MSDSANet: Multimodal Emotion Recognition based on Multi-Stream Network and Dual-Scale Attention Network Feature Representation," *Sensors*, vol. 25, no. 7, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hua Jin et al., "Multimodal Emotion Recognition in Conversations Using Transformer and Graph Neural Networks," *Applied Sciences*, vol. 15, no. 22, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[16] A.S. Agrawal, "Bouncing Scenario and Cosmic Dynamics in Modified Theories of Gravity," *arXiv preprint*, pp. 1-150, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Farshad Safavi, Kulin Patel, and Ramana Vinjamuri, "Facial Expression Recognition with an Efficient Mix Transformer for Affective Human-Robot Interaction," *IEEE Transactions on Affective Computing*, vol. 16, no. 4, pp. 3081-3094, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Haoran Yang, "*Graph Contrastive Learning and Its Applications in Recommendation Systems*," UTS Digital Thesis Collection, pp. 1-142, 2024. [Google Scholar] [Publisher Link]

[19] Ricky T.Q. Chen, "Neural Ordinary Differential Equations," *Advances in Neural Information Processing Systems*, pp. 1-18, 2018. [Google Scholar] [Publisher Link]

[20] Yulia Rubanova, Ricky T.Q. Chen, and David Duvenaud, "Latent Ordinary Differential Equations for Irregularly-Sampled Time Series," *Advances In Neural Information Processing Systems*, pp. 1-21, 2019. [Google Scholar] [Publisher Link]

[21] Meisam J. Sekiavandi et al., "Advancing Face-to-Face Emotion Communication: A Multimodal Dataset (AFFEC)," *arXiv preprint*, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]