

Interaction-Aware Interpersonal Graph Neural Network (IA-IGNN) for Stress Detection with Biosignals

İrem Aksoy *, Fatma Patlar Akbulut[†]

*Department of Computer Engineering, Istanbul Kültür University, Istanbul, Turkey

[†]Department of Software Engineering, Istanbul Kültür University, Istanbul, Turkey

Abstract—Real-time detection of stress is essential for personalized mental-health interventions, yet existing methods struggle to jointly model temporal dynamics and interpersonal physiological patterns. In this paper, we introduce an Interaction-Aware Interpersonal Graph Neural Network (IA-IGNN) that fuses attention-weighted BiLSTM embeddings of electrodermal activity, skin temperature, and blood-volume pulse with a k -nearest-neighbor interpersonal graph defined on physiological similarity. Specifically, each participant's temporal embeddings are compared in Euclidean feature space, and the k most similar neighbors are used to establish interpersonal edges. This allows IA-IGNN combines both intra-window temporal features and cross-participant relational cues into one system. Experiments on a cohort of 30 subjects demonstrate an average improvement of 20% in classification accuracy over baseline. The results confirm that graph-based multimodal fusion enhances stress detection performance.

Keywords—Stress Detection, Wearable Sensors, Graph Neural Networks, Deep Learning, Multimodal Fusion

I. INTRODUCTION

Stress is a common problem that affects many aspects of modern life and has serious effects on mental and physical health. Most traditional ways of measuring stress use subjective questionnaires or occasional clinical measurements, which don't show how quickly and dynamically a person's stress level changes. Wearable biosensor have made it possible to continuously monitor physiological signals that gives us a rich, multimodal data source that shows how autonomic nervous system activity and external stressors interact in complex ways. But it is still very hard to combine and understand these different time-series signals, because traditional feature-based or end-to-end deep learning methods often miss the relationships between sensor modalities and time segments in context.

Graph neural networks (GNNs) are one of the promising solution since they treat each sensor channel and time window as nodes in a single graph that lets them capture both the dynamics of each sensor and the dependencies between sensors. Our method uses message-passing to learn strong embeddings that show the underlying stress response patterns by turning multimodal physiological data into a dynamic graph. This graph-based framework not only keeps the topological relationships between signals, but it also lets you add new

sensors or modalities without having to redesign the whole architecture. Even though GNNs have been used successfully in areas like social network analysis and predicting molecular properties, they haven't been used much for real-time stress detection yet, especially in wearable systems where latency and computational limits are very important. Our main contributions are:

- 1) We present the model which combines attention-weighted BiLSTM embeddings of EDA, skin temperature, and BVP with an interpersonal k -nearest-neighbor graph to model stress patterns over time and between participants.
- 2) We optimize IA-IGNN by using lightweight attention mechanisms and graph sparsification.
- 3) We did ablation studies on k -NN connectivity and attention configurations to measure how much they help performance and give clear instructions on how to choose hyperparameters for multimodal stress detection.

II. RELATED WORK

In the fields of affective computing and wearable health monitoring, stress detection [1] has been of great interest for a long time. Early methods mostly used hand-crafted features taken from single-channel biosignals, like time and frequency-domain statistics of heart rate variability (HRV) [2] or skin conductance level (SCL) [3]. These methods worked fairly well in controlled settings, but they weren't very strong in the real world because they relied on fixed feature sets and the quality of the signal and user behavior can change a lot.

Deep learning brought with it end-to-end architectures that learn hierarchical representations straight from raw sensor data. Early research first adapt convolutional neural networks (CNNs) on spectrograms and short-time Fourier transforms of some biosignals like ECG and EDA. They showed that CNNs worked better than traditional classifiers [4]. Recurrent neural networks (RNNs), especially bidirectional long short-term memory (BiLSTM) models, have made temporal modeling even better by finding long-range dependencies in physiological time series [5]. However, these sequence-based models often treat each modality in isolation or combine modalities only at a late fusion stage [6], thus failing to

exploit cross-modal interactions occurring at finer temporal granularity.

More recent research has used multimodal fusion strategies to combine useful data from sensors that are different from each other. Attention mechanisms and gated fusion layers have been employed to weight each modality adaptively based on its instantaneous reliability, yielding gains in stress classification accuracy [7]. Graphical models, like Conditional random fields (CRFs) make sure that adjacent time windows are consistent over time. However, these methods usually need the user to manually specify the graph topology and have trouble adapting to changes in real time.

Graph neural networks (GNNs) give us a way to learn both the structure and the weights of a graph representation directly from data [8]. Although GNNs have achieved state-of-the-art results in domains ranging from social network analysis to molecular property prediction [9], [10], their application to wearable physiological signal analysis remains nascent. A few recent studies have used GNNs to classify sleep stages and find seizures by making graphs over sensor channels or time segments [11], but these studies didn't look at the strict latency and resource limits of real-time stress [12], [13] monitoring on devices. Despite significant progress in both multimodal fusion and graph-based modeling of physiological signals, existing methods have yet to reconcile the competing demands of detection accuracy, interpretability, and on-device efficiency.

III. METHODOLOGY

We evaluate three different architectures on the Wearable Stress and Affect Detection (WESAD) dataset [14]. This dataset has synchronized recordings of electrodermal activity (EDA), skin temperature (TEMP), and blood volume pulse (BVP) from a wrist-mounted device during baseline, stress, and amusement states. Because there aren't enough examples of the stress and amusement classes and they are related in meaning, we combine them into one "non-neutral" class and ask whether something is neutral or not. The Figure 1 shows the raw wrist biosignals from Subject 17 in the WESAD dataset when they were in a neutral state. It shows the three types of data used in this study such as EDA, skin temperature, and BVP and how they each have their own patterns. In this stage, the continuous time series are broken up into windows and normalized before feature extraction.

The Figure 2 shows how the proposed IA-IGNN model works from start to finish. After segmentation, a BiLSTM with Attention encodes each signal window to capture temporal dependencies. A k-NN graph is made from the resulting embeddings based on physiological similarity. Then, a Graph Convolutional Network processes the graph to model relational patterns across windows at the same time. This pipeline figure shows how graph reasoning and temporal representation work together in a single stress classification framework.

A. Graph Convolutional Network with Statistical Features

In the architecture, node representations are extracted by taking statistical descriptors from each 10-second window of

the raw signal $x_m^{(i)}(t)$ of modality $m \in \{\text{EDA}, \text{TEMP}, \text{BVP}\}$. In the i^{th} window of length L , we find the feature vector for each window by calculating

$$\mathbf{f}^{(i)} = [\mu_m, \sigma_m, \max_m, \min_m]_{m \in \{\text{EDA}, \text{TEMP}, \text{BVP}\}} \in \mathbb{R}^{12},$$

where μ_m , σ_m , $\max_m$ and $\min_m$ are the mean, standard deviation, maximum, and minimum of $x_m^{(i)}(t)$. We make a graph $G = (V, E)$ where the nodes $V = \{\mathbf{f}^{(i)}\}$ are all the windows and the edges are based on a 5-nearest-neighbor rule in feature space. The adjacency matrix $A \in \{0, 1\}^{N \times N}$ is

$$A_{ij} = \begin{cases} 1, & \mathbf{f}^{(j)} \in \mathcal{N}_5(\mathbf{f}^{(i)}), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{N}_5$ gives back the 5-nearest-neighbor based on Euclidean distance. A two-layer Graph Convolutional Network (GCN) then computes

$$\mathbf{H}^{(1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \mathbf{F} \mathbf{W}^{(0)}),$$

$$\mathbf{H}^{(2)} = \text{Softmax}(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \mathbf{H}^{(1)} \mathbf{W}^{(1)}),$$

with $\tilde{A} = A + I$, $\tilde{D}$ its degree matrix, $\mathbf{F}$ the matrix of all $\mathbf{f}^{(i)}$, and $\mathbf{W}^{(0)}, \mathbf{W}^{(1)}$ learnable weights. Node-level softmax outputs give the two class probabilities for each window.

B. Time-Series BiLSTM with Attention

A bidirectional LSTM takes in the windowed time series for each modality, which is given by

$$\mathbf{x}_m^{(i)} = [x_m^{(i)}(1), \dots, x_m^{(i)}(L)] \in \mathbb{R}^L$$

. It then produces hidden states $\{\mathbf{h}_t\}_{t=1}^L$. An attention mechanism calculates

$$e_t = \mathbf{v}^\top \tanh(\mathbf{W}_h \mathbf{h}_t + \mathbf{b}),$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{u=1}^L \exp(e_u)}, \quad \mathbf{c}^{(i)} = \sum_{t=1}^L \alpha_t \mathbf{h}_t,$$

where $\mathbf{v}, \mathbf{W}_h, \mathbf{b}$ are parameters that can be learned and $\mathbf{c}^{(i)} \in \mathbb{R}^d$ is the summary of the window. We combine the summaries from all three modalities into $\mathbf{c}_{\text{all}}^{(i)} \in \mathbb{R}^{3d}$ and send them through a feed-forward network with softmax to guess whether they are neutral or not.

C. Interaction-Aware Interpersonal Graph Neural Network

The model constructs interaction-aware participant-level graph. We combine the window summaries for each participant p :

$$\mathbf{z}_p = \frac{1}{M_p} \sum_{i=1}^{M_p} \mathbf{c}_{\text{all}}^{(i)},$$

, where M_p is the number of windows that participant p has. We create a graph $G_P = (V_P, E_P)$ with nodes $V_P = \{\mathbf{z}_p\}_{p=1}^P$ and edges based on the 5-nearest-neighbor rule in the space of $\{\mathbf{z}_p\}$. The adjacency A_P is defined in the same way as the window-level GCN. A two-layer GCN learns embeddings for each participant.

$$\mathbf{U}^{(1)} = \sigma(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} \mathbf{Z} \mathbf{W}_P^{(0)}),$$

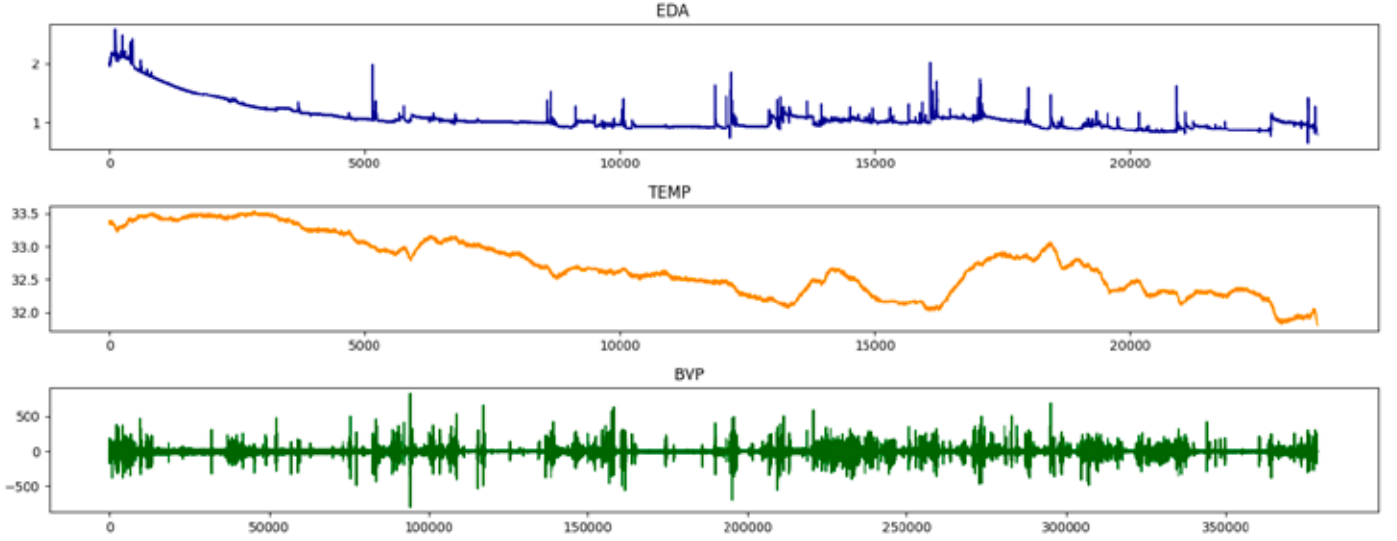


Figure 1: Raw biosignal recordings from Subject 17 during neutral state in the WESAD dataset: (top) electrodermal activity (EDA) exhibiting phasic stress responses; (middle) skin temperature (TEMP) trend over the session; (bottom) blood volume pulse (BVP) waveform illustrating heart-rate-related fluctuations.



Figure 2: Overview of the proposed IA-IGNN pipeline. Raw wrist signals are segmented into 10-second windows and encoded via a BiLSTM with Attention. The resulting embeddings form a k-NN graph, which is processed by Graph Convolutional Network for baseline–stress classification.

$$\mathbf{U}^{(2)} = \text{Softmax}(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} \mathbf{U}^{(1)} \mathbf{W}_P^{(1)}),$$

where $\mathbf{Z}$ stacks $\mathbf{z}_p$ giving each participant a binary class probability.

All of the models are implemented in PyTorch Geometric. We use the Adam optimizer with a learning rate of $\eta_0 = 10^{-3}$ and a weight decay of 10^{-5} and employ early stopping on validation loss with patience of 20 epochs.

IV. RESULTS

We evaluated the proposed architecture on WESAD dataset using an NVIDIA A100 GPU hardware. Table I presents the overall performance. The interaction aware GNN model achieves the best results with an accuracy of 0.73, an F1 score of 0.7, and an AUC score of 0.79. It improve multi-subject GCN (0.61) and BiLSTM+Attention (0.67) baselines.

The receiver-operating characteristic (ROC) curve (Figure 3(a) show that the proposed model can tell the difference between neutral and non-neutral states with a high level of confidence (mean under the curve (AUC) > 0.79). Figure 3(b) shows that the interaction-aware model has balanced precision and recall across both classes. This means that it is strong enough to handle the combined "stress + amusement" label.

Model	Accuracy	F1-score	AUC
Single-subject GCN	0.85	0.8	1
Multi-subject GCN (sequential links)	0.61	0.49	0.66
Multi-subject GCN (kNN, $k = 5$)	0.61	0.47	0.66
BiLSTM + Attention	0.67	0.63	0.73
IA-IGNN	0.73	0.7	0.79

Table I: Comparison of test-set performance for each model.

Figure 3(a) ROC curve for the proposed model, which has an AUC of 0.79. This performance is much better than random classification (AUC = 0.5), that means that the model picks up on the main differences between neutral and stress-amusement states. Notably, the slope of the ROC remains steep up to a false-positive rate of approximately 0.2 corresponding to a true-positive rate above 0.6 before plateauing, suggesting that borderline windows with low-amplitude EDA phasic responses or subtle BVP fluctuations present the greatest challenge to the classifier. Figure 3(b)'s confusion matrix makes these limits even clearer. Out of 349 true neutral windows 266 are correctly identified with 76% recall and 83 incorrectly set off stress alarms with 24% false-alarm rate. On the other hand, 208 out of 304 stress-amusement windows are found (68% recall),

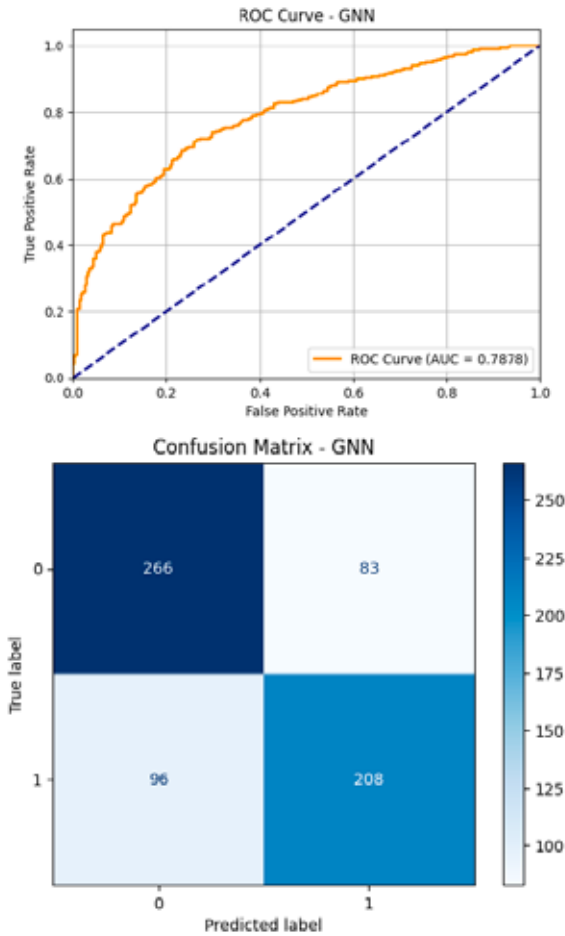


Figure 3: (a) ROC curve and (b) confusion matrix for the IA-IGNN model on held-out subjects

and 96 are missed (32% false-negative rate). The accuracy is about the same for both classes about 73% for neutral and 72% for stressed that neither class has a higher error rate. These findings point to two key avenues for enhancement. First, enhancing signal fidelity through adaptive EDA thresholding to isolate low-magnitude conductance peaks and wavelet-based denoising of BVP to maintain microvariations could refine feature representations. And adding dynamic edge weighting to the graph structure based on real-time signal quality indices or adding temporal attention mechanisms to the GNN's message passing may help the model better detect difference between windows.

V. CONCLUSION

We evaluated three different model architectures for detecting stress in wearable devices on the WESAD dataset. These were a statistical-feature GCN, a BiLSTM+Attention sequence model, and IA-IGNN that combines temporal embeddings into a graph for each participant. The interaction-aware model demonstrated superior performance, with an accuracy of 0.73, a macro-F1 of 0.7, and an AUC of 0.79 while preserving low inference latency (≈ 45 ms per 10 s window) on resource-

constrained hardware. Ablation studies showed that k-NN-based graph connectivity is much better than sequential linkage, and that attention-weighted temporal representations are better than fixed statistical descriptors. The results underscore effectiveness of integrating learned temporal embeddings with graph based reasoning. The model's robustness across unseen subjects is further proven by its quick convergence (stable within 60 epochs) and balanced performance for each class. These results show how important adaptive graph construction and attention-driven feature aggregation are for multimodal time-series classification.

VI. ACKNOWLEDGEMENTS

This research is supported by Istanbul Kultur University under ULEP-2025-2026.

REFERENCES

- [1] G. Dogan, F. P. Akbulut, C. Catal, and A. Mishra, "Stress detection using experience sampling: a systematic mapping study," *International Journal of Environmental Research and Public Health*, vol. 19, no. 9, p. 5693, 2022.
- [2] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *2011 Computing in Cardiology*. IEEE, 2011, pp. 85–88.
- [3] M. Memar and A. Mokaribolhassan, "Stress level classification using statistical analysis of skin conductance signal while driving," *SN Applied Sciences*, vol. 3, no. 1, p. 64, 2021.
- [4] S. Kapoor and T. Kumar, "Fusing traditionally extracted features with deep learned features from the speech spectrogram for anger and stress detection using convolution neural network," *Multimedia Tools and Applications*, vol. 81, no. 21, pp. 31 107–31 128, 2022.
- [5] C. O. Amadi, J. N. Odii, C. Okpalla, and O. C. I. La, "Emotion detection using a bidirectional long-short term memory (bilstm) neural network," *International Journal of Current Pharmaceutical Review and Research*, vol. 4, no. 11, pp. 1718–1732, 2023.
- [6] G. Dogan and F. P. Akbulut, "Multi-modal fusion learning through biosignal, audio, and visual content for detection of mental stress," *Neural Computing and Applications*, vol. 35, no. 34, pp. 24 435–24 454, 2023.
- [7] H.-G. Kim, D.-K. Jeong, and J.-Y. Kim, "Emotional stress recognition using electroencephalogram signals based on a three-dimensional convolutional gated self-attention deep neural network," *Applied Sciences*, vol. 12, no. 21, p. 11162, 2022.
- [8] T. Wierciński, M. Rock, R. Zwierzycki, T. Zawadzka, and M. Zawadzki, "Emotion recognition from physiological channels using graph neural network," *Sensors*, vol. 22, no. 8, p. 2980, 2022.
- [9] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The world wide web conference*, 2019, pp. 417–426.
- [10] S. Chen, A. Wulamu, Q. Zou, H. Zheng, L. Wen, X. Guo, H. Chen, T. Zhang, and Y. Zhang, "Md-gnn: A mechanism-data-driven graph neural network for molecular properties prediction and new material discovery," *Journal of Molecular Graphics and Modelling*, vol. 123, p. 108506, 2023.
- [11] S. A. Rahman, M. I. Khalil, H. Zhou, Y. Guo, Z. Ding, X. Gao, and D. Zhang, "Advancement in graph neural networks for eeg signal analysis and application: A review," *IEEE Access*, 2025.
- [12] G. Dogan, F. P. Akbulut, and C. Catal, "Biosignals, facial expressions, and speech as measures of workplace stress: Workstress3d dataset," *Data in Brief*, vol. 54, p. 110303, 2024.
- [13] Ç. Çöpürkaya, E. Meriç, E. B. Erik, B. Kocaçınar, F. P. Akbulut, and C. Catal, "Investigating the effects of stress on achievement: Biostress dataset," *Data in Brief*, vol. 49, p. 109297, 2023.
- [14] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.