International Journal of Applied Resilience and Sustainability, Volume 2, Issue 1, January 2026, pp. 190-210

Published online at Deep Science

International Journal of Applied Resilience and Sustainability

Journal homepage: https://deepscipub.com/ijars

Predicting psychological resilience and mental health from multimodal wearable sensor data using graph neural networks

Birupaksha Biswas [1], Suhena Sarkar [2]

[1] Department of Pathology, Burdwan Medical College & Hospital, Burdwan, India
[2] Department of Pharmacology, Medical College, Kolkata, India

Article Info:
Received 20 December 2025
Revised 08 January 2026
Accepted 22 January 2026
Published 26 January 2026

Corresponding Author:
Birupaksha Biswas
E-mail: drbiswasassted.medicalglory@gmail.com

Abstract

Early diagnosis of mental vulnerability is still a problem since most clinical evaluations are based on the subjective aspects of evaluation like self-report and not on the objective physiological indications. The idea of psychological resilience which is a fundamental protective factor against stress related disorders has been connected to the autonomic regulation and daily behavioral patterns, but the objective method of calculating on such a scale is yet to be established. This paper presents a multimodal graph neural network, which combines the wearable derived physiological signals to forecast trait level resilience and stress as well as state level. The ongoing heart activity, electrodermal activity, and motion data of adult participants were gathered and matched with resilient score proven validated scores. Participant specific graphs were created whereby physiological modalities were the nodes and the inter signal dependencies were the edges. Statistical analysis showed that there were significant physiological disparities that were related to resilience. High resilience group members had much better values of root mean square of successive differences and physical activity per day (mean difference on 1000 extra steps per day, $p = 0.003$). The proposed graph neural network (GNN) performed well in classification tasks in terms of area under the receiver operating characteristic curve (0.80 +- 0.02) to differentiate high and low resilience, which was significantly better than that of logistic regression (0.60), random forest (0.65), and long short-term memory models (0.72), with the difference being supported by DeLong test ($p < 0.01$). Graph based learning also offers added benefits of discrimination and physiologically explainable output and thus can facilitate its future as a scalable and objective digital mental health monitoring and early risk stratification tool.

Keywords: Psychological resilience, Graph neural networks, Wearable sensors, Mental health, Stress detection, Machine learning

1. Introduction

A significant challenge that is increasing in the world, worsened by chronic stress in the contemporary life, is mental illnesses that include depression, anxiety, and post-traumatic stress disorders [1,2]. Such conditions are usually underdiagnosed until they are severely debilitating day to day functioning [2]. One of the developing preventive measures is the constant checking of physiological and behavioral indicators with the help of wearable devices to reveal the existence of psychological distress in the initial stages [3-5]. Warmth devices such as smarty watches and physical fitness bracelets will have an ability to monitor bio-signals (heart rate, exercise, sleep, electrodermal activity, and so forth) of a person in real time and without any invasive access [2,6]. These signals have objective proxies of stress and affective states, such as heart rate and sweating (highlighted by photoplethysmography (PPG) and galvanic skin response sensors), are acute responses of stress [7-9]. Through these streams of data, scientists are striving to create digital variants of mental health biomarkers that can overcome the constraints of the infrequent clinical monitoring and symptoms recalling through self-report data [10].

In the mental health construct, psychological resilience is one of the constructs that have been focused on as an important protective construct [10,11]. Resilience is described as the ability to get through

https://doi.org/10.70593/deepsci.0201010

tough situations or recover after enduring hardships and alleviating the effects of stress on psychological health [12-14]. High-resilience persons are also more likely to stay psychologically healthy during times of stress, and low resilience brings the susceptibility to such illnesses as depression and anxiety [3,15-17]. Historically, the measure of resilience has been performed in terms of self-report measures (i.e. the Connor-Davidson Resilience Scale) that assess the perceived capability to cope with stress. Such measures are good but they are fixed and subjective [18-20]. The physiological aspects that can be monitored in order to determine physiological correlates of resilience are becoming of increasing interest [21-23]. Previous studies put forward that heart rate variability (HRV) - the alteration in the heart rhythms through changes in the beat-to-beat interval can be an objective resilience measure. Greater-parasympathetic (relaxing) of the heart, indicated by a high vagally-mediated HRV has been associated with the laboratory studies of emotion regulation and stress adaptation. E.g. people that have a higher resting HVR demonstrate enhanced adaptability of cortisol and cardiovascular actions to challenges generated by stress, indicating that HVR is a biomarker of mental health resilience [9,24,25]. These observations suggest that the wearable sensor data of resilient individuals could have unique features, including smoother heart dynamics, quicker physiological recovery of stress, or well-adjusted sleep/activity patterns [26-28]. In fact, the theory of physiological resilience may be operationalized to mean how quickly an organism returns to equilibrium following perturbation [6,29-31]. Recent longitudinal wearable data research revealed that older adults take a long time (loss of resilience) to recover following fluctuations of everyday activity and this has been associated with aging (and mortality risk). Psychological resilience may also be expressed in less time to achieve normalization of heart rate or electrodermal activity of the body after stress in young healthy populations [32,33]. These are delicate dynamics of signals that cannot be analyzed by simple triggers based on threshold.

Accelerometers (ACC) help to identify the physical activity of a patient and their sleep waveforms. The modalities present varied access to the psychophysiological condition. Incorporating parasympathetic (increasing heart rate and EDA) and depressive (decreasing activity and irregular circadian rhythms) symptoms, all of these responses to stresses are likely to be activated by stress. With the help of these mixed signals, we will be able to create a more holistic image of the psychological state of an individual [34-36]. The literature indicates that multimodal fusion leads to greater accuracy in identifying depression, anxiety, or cognitive load: a physiological signal proves more accurate than a behavioral or contextual signal, and a combination of both signals is even more accurate than any individual signal [16,37-40]. Recent systematic reviews indicate that studies of mental health monitoring using multiple biosensors with AI have an increasing trend, emphasizing the understanding in the community that there is no single signal that could be used in complex situations. The multimodal data are however also problematic - the interactions among signals may be context-dependent and non-linear. The example is, in a case of an exercise; both heart rate and accelerator activity increase (positive relationship), whereas during a panic attack heart rate could explode without movement. Such conditional dependencies may not be elicited through traditional machine learning methods which merely concatenate the features or consider the sensors one at a time.

Graph neural networks have become a trend in deep learning, which works best with data whose relationships are complex [41-42]. In a graph model, the data entities are shown in the form of a node and the links between these nodes as an edge. It is highly adaptable to wearable multimodal data, in which individual sensor streams may be regarded as nodes in a physiological network, and the statistical or functional relationships between sensors (e.g. correlation between heart rate and EDA signals) can be regarded as edges. In comparison to sequence models or models based on vectors, GNNs are able to acquire latent patterns of interactions, and different connections may have different importance based on the state. According to recent surveys, GNNs can be used to apply context-dependent sensor fusion: the network can, e.g., learn to boost the importance of electrodermal activity nodes, when under stress, and motion sensor nodes, when moving around, by dynamically changing sensor weights. This dynamic merger is consistent with domain knowledge, the use of skin conductance would be most informative in emotional arousal whereas accelerometry plays a central role in identifying the occurrence of exercise or sleep. Moreover, GNNs are flexible to incorporate more information in the form of nodes or edges, environmental setting (context) or social networks (interpersonal interactions) to which mental health is applicable. GNNs are still new in the field of predicting mental health conflict with wearable sensors data but have shown promise in different other areas, such as brain connectivity analysis and health care risk prediction. Remarkably, researchers came up with a multimodal fusion GNN creating a model that can integrate neuroimaging and clinical data to diagnose depression with around 79 percent accuracy and includes biomarkers of brain connectivity that characterizes depression. This shows how graph-based learning can identify some obscure patterns in multifaceted biomedical information. We assume that one can execute a corresponding procedure to wearable sensor streams and anticipate psychological fortitude and mental well-being states which inherent complex relations between physiological indicators throughout a timeframe.

The development of digital sensing and machine learning to enhance mental health has some gaps that are highlighted. To begin with, the construct of resilience has not been experimented so much in sensor-only research. Recent researchers analyzed healthcare workers and discovered that machine learning models (gradient boosting) could be used to rank high and low resilience people, using Apple Watch data, with the AUC value of around 0.60. This implies that the signal characteristics employed (which most probably were heart rate and activity measurements) by themselves did not have much predictive power. More elaborate representations of features or later sensor modalities may be required to represent the complicated phenotype of resilience. Second, in this field, there are no ways to integrate multimodal time-series in an effective way. The majority of the previous studies involved either feature-level fusion (a mere combination of all features into one feature of a single vector) or ensemble models, which do not fully utilise the inter-sensor relationships. Graph neural networks help present a way to eliminate this limitation by explicitly defining the multimodal data structure. More importantly, GNNs can not only be used to make predictions more accurate but also more interpretable: when we take in the weights of learned edges or attention scores, we may understand which of the physiological links will best predict resilience or stress. This is congruent with the push of explainable AI in health monitoring such that clinicians may be able to trust new model outputs and interpret them.

This paper will contribute to improving the current state of wearable mental health analytics by proposing a multimodal sensor-based neural network-based solution capable of forecasting psychological resiliency and mental health conditions based on multimodal sensor data. To the best knowledge, it is the first publication to explicitly address the research question of predicting resilience based on wearable signals via graph neural networks. We discuss the following objectives:

(1) Find a new way to mechanize multimodal physiological data of a person as a graph that would provide the correlation of sensor modalities and time.

(2) Train and test a graph neural network model that has been trained on this representation to give predictions of resilience (as measured by standardized scales) and acute psychological stress (as a representative mental health state).

(3) Perform comparison between the performance of the proposed GNN and classic methods (e.g. deep LSTM or random forest classifier) to determine the improvement.

(4) Discover predictions in the model, including which signals and links have the greatest contribution and discussing them in connection with current psychophysiological data. Our research work has both methodological and practical contribution. At an algorithmic level, we provide a framework that is built upon time-series features extraction, a graph-based data fusion, and end-to-end deep learning-based trend prediction of mental health. In practice, our findings show stronger prediction of resiliency and stress, which would allow continuous surveillance systems to detect individuals who are prone to unfavorable mental health outcomes and prior to the development of clinical symptoms. Our work also preconditions the creation of the personalized feedback systems that could identify problems besides assist with the development of the resilience as a modifiable factor that can be influenced by training and other interventions. The remainder of this paper elaborates on dataset and chart generation methodology, GNN architecture and training process and then the results of the experimentation and the ramifications and directions.

2. Methodology

Constant sensor streams were divided and the processed to derive the efficient features to analyze them. We used sliding window method, with window length of 1 minute (long-term features of the signal) and 1 hour (long-term trend features of the signal) with sufficient overlaps (50% overlap in minute windows) between the windows to trade-off time resolution and stability of features. Each of the windows yielded a set of summary features (means of heart rate) in each modality: In the case of heart activity, we have computed mean heart rate (HRmean) and various heart rate variability indices (e.g. root mean square of successive differences RMSSD and low-frequency/high-frequency power ratio) which indicate autonomic balance. Out of EDA, we obtained data on features including the tonic skin conductance level (SCL) and phasic driver measures (e.g. total skin conductance responses (SCR), SCR count, and SCR amplitude), measures of overall arousal and local burst of sympathetic activity. Essentially, we used accelerator information to determine mean acceleration, number of steps (assuming that the window size is large enough), and one of the categories (e.g. sedentary or active) based on some threshold of the number of activities. In longer hourly windows, we also obtained characteristics, such as the total amount of steps per hour, minutes of moderate-vigorous activity and an approximate determination of the sleep quality (during nighttime) based on movement-based sleep detection algorithms. To derive ciradian rhythm characteristics (e.g. amplitude of the diurnal variation) of skin temperature, smoothing was done to eliminate the features of daily circadian rhythms of skin temperature.

All the time-series were synchronized and aligned to the daily stress reports. To incorporate stress classification models (under the premise that the overall stress in a day is reflected in the physiological patterns of that day), on a 1-minute window we attached the stress level of the corresponding day. As much as this adds noise (as stress is not always the same throughout the day), it gives a weak indication of the daily stress condition. Windows that were gathered in case of self-report high-stress days ((rating ≥ 7). 7 were identified as cases of "stress" and those of the lower days (rating ≤ 3) as low-stress days. 3) were marked as "non-stress." Medium days (rating 4-6) were not included in training so as to enhance class separability and this was utilized later to test generalization of its models. To be resilient, given that it is a person-level (baseline) measure, we inputted each participant individual resilience score or group label on each of the windows of a particular participant during training the resilience prediction model. In practice, we did not treat the different windows at a time; instead, we consolidated all the entire 30 days of individual participants into one graph representation (discussed below) to predict resilience, so as to obtain long-term trends.

Before the feature computation, conventional preprocessing methods were used: noisy data (because of motion artefacts or device off-body) were found and discarded by a combination of rules based on threshold (e e.g. acceleration near to zero during a long period signalifying the removal of the device) and rules based on outliers on physiological signals (e e.g. heart rate greater than 200 bpm almost certainly artefact). The further data were z-score normalized individually per participant across sensors to bring about individual variations in baseline levels (e.g. each participants heart rate characteristics were normalized towards the respective mean and standard deviation, without changing the relative variation). We also used normalization between the participants where needed to model at the population level and did this by dividing the participants into a training group and a test group to eliminate the problem of leakage.

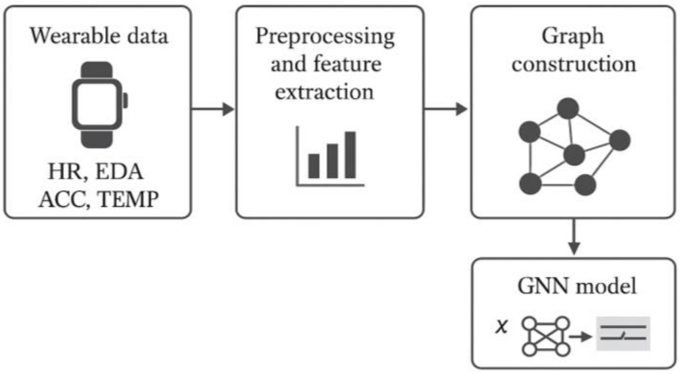

Figure 1: This Schematic workflow (presenting the end-to-end pipeline of resilience prediction: between multimodal wearable acquisition (HR, EDA, ACC, TEMP) and graph construction and GNN modeling. gives an overall pictorial outline of the data pipeline between wearable sensing, data preprocessing and graph neural modeling. It is a statistical explanation of the way raw physiological streams are transformed into graph nodes and edges representing the correlation coefficients which are structured. The grayscale architecture focuses on showing the direction that the flow takes between acquisition and inference. This figure contributes to the reproducibility requirements of computational health research by making the readers have the conceptualization of data transformation processes.

Graph Construction: One of the highlight contributions of our approach is that it converts multimodal time-series data into the form of a graph structure, whose value is entered into the GNN. We have used two scale graphical representation: (1) Intra-window graphs by the level of short time windows (a one-minute segment), and (2) the global graph of each participant which connects various time windows or summary nodes. As a graph to simplify analysis, though, we consider the participant-level graph, written G p, of participant p, which is an expression of the multimodal physiological profile of the particular participant during the whole period of monitoring. The node set of the participants graph in each participant graph G-p (V, E) is comprised of nodes representing different sensor derived feature streams. We are defining a single node of each of the major or sub-modality of interest. Namely, we added the following nodes; HR/HRV node - cardiac features; EDA node - electrodermal features; ACC node - activity/sleep features; and TEMP node - skin temperature (circadian rhythm). We added an external context node, too, to encode the extrinsic context (ex: time of the day, or day of the week), and can affect all signals; this node is joined to all modality nodes with directed morphemes (ex: whether the part was at a work hour or a rest hour). There is a feature vector of the data of each modality in the participant that is attached to each node. To predict the resilience at all times, we applied the full dataset of each of the participants to calculate global summary features at the node. The feature vectors in the HR / HRV node would, as an example, have the average resting heart rate of a participant, global HRV values (e.g. for 24 -h RMSSD), and possibly HR amplitude. The feature of EDA node may include general 90 th percentile of SCL and mean frequency of SCR during the waking hours. CC node feature comprises of mean daily steps taken, mean time spent as sedentary, etc. These attributes summarize the long term dynamics of every signal giving the GNN a point of departure to acquire inter-modal interactions.

Relationships between modalities between nodes are expressed as edges E. The GNN was trained to learn any possible connection between the physiological nodes by defining an undirected edge between all the pairs of physiological nodes (and creating a fully connected graph of the modalities). In order to begin with initializing weights on the edges, we have calculated the empirical Pearson correlation of the time-series of any two modalities during the 30 days. As an example, we used the calculation of the correlation between hourly EDA level and hourly heart rate of the participant; strong positive correlation would suggest that high heart rate rates are likely to be accompanied by high arousal, and possibly, stress or physical activity, and weak positive or negative with independent and inverse relations. These correlation value (or its adequate feature) served as features on the edges. Definitely, between two nodes at the modalities, $\{i \ and \ j$ on node i and j, we obtain the correlation coefficient:

$$ij \ = \ t = 1TXit - X^{-}i\,Xjt - X^{-}jt = 1TXit - X^{-}i2 \ \ t = 1TXjt - X^{-}j2 \tag{1}$$

The t in this case refers to the time points (or aggregated timepoints) within the monitoring time, and the BarXi denotes the mean measurements over time of modality i. The use of this rho represents the first edge attribute which encodes the magnitude of linear relationship between the modalities i and j. Indicatively, in the event that the participant has a high correlation between his or her HR with ACC (where the correlation is close to 1), then most of the elevations of heart rate can be attributed to exercise; in events where the correlation between HR and EDA is high, then changes can be observed as possible stress responses. We further gave each edge a binary type label that tells it whether it links physiological nodes to context node or to two physiological nodes, such that the GNN has the potential of learning varying weights between edge types. The edges of the context node had no correlation but instead an arbitrary value of 1 since they are intended to consist in a consistent manner of connecting context to all modalities.

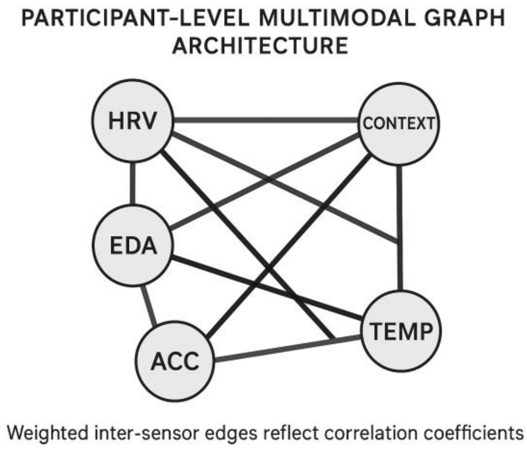

Figure 2: This is a representative participant graph of the multimodal nodes (HRV, EDA, ACC, TEMP, context) and weighted inter-sensor edges which do transform the abstract graph concept into a physiological network. The thickness of edges indicates Pearson correlation size between modalities; the dark color indicates higher coupling. It statistically shows intermodal dependencies that form the basis of the learning ability of the GNN. It presents the dynamics of autonomic-behavioral interaction as clinically obtained to provide a system level model of the physiology of resilience.

Graph Neural Network Architecture: our GNN model works with the participant graph, G (p) to give resilience outcomes and stress outcomes predictions. The architecture comprises of two graph convolutional layers and a readout and fully connected prediction layer. The convolutional layers took the Graph Convolutional Network (GCN) formulation by Kipf and Welling, because it is efficient and it allows the use of weighted adjacency information. In the GCN, node features are updated by the current layer courtesy of the information obtained by the neighboring nodes. Taking the rule of propagation of a single GCN layer mathematically, it can be expressed as follows:

$$Hl + 1 \ = \ \sigma D - 12 \ A \ D - 12 \ Hl \ Wl \tag{2}$$

Here, each layer node feature matrix is denoted by the matrices: H (l) = R N x dl: theself-connection variant of an adjacency matrix, A, plus added identity matrix, I, and diagonal degree matrix, D on A. This operation is an effective weighted average of each node of the neighborhood features (both including itself) and is then transformed linearly. Notably, adding edge weights in the form of a weight matrix, called A (we selected the weight to be the inverse of physiological importance between two nodes, i.e. we use the weight of an edge, i.e. $\rho\{ij\}$) will make the GCN aggregations depend on tie strength, i.e. strong physiological connections between nodes will enable more information to flow between the corresponding nodes in the network. The first GCN layer works on the original node feature matrix H0 (formed by the extracted features of the each modality) and generates a hidden representation of size N x d1 (we use d1 = 16) denoted by H 1. The second GCN layer then creates H2 of size N d2 (where d2 = 8), the last group of node embeddings.

We followed the same steps after the GCN layers; we followed the graph readout operation to retrieve one representation of the overall participant graph. We took a basic readout based on mean pooling: we mean all node feature vectors in H2 to obtain a graph-level feature vector h 2 graph in R d 2. We also tried weighted sum pooling and the emphasis constituting pooling, but mean pooling seemed to be sufficiently good as there were a limited number of nodes and thus our normalizing features. Such an attention pooling based on relevancy of nodes would be helpful in case some modalities were much more informative than others.) This is a graph-level representation of the heterogeneity which is a representative of the physiological profile of the participant in their entirety.

Lastly, to do prediction, we inputted the graph of h0(Finally) with the fully-connected neural network head to produce the desired outputs. As we have 2 target outcomes, resilience (scalar or binary high/low label) and daily stress (this may be classified or regression problem, so we used two heads which were resilience and stress). The resilience head is made of one neuron (to regress on the resilience score), or a small sub-network whose terminal is a sigmoid (to classify high and low resilience). The design of the stress head will result into a probability of high stress on a day or window. Virtually, since resilience is measured only once per participant, we just train the resilience head on the graph of the participant (a single label per graph), but the stress head can be trained on many instances per participant (assuming we count an incidence of each day or window to be an instance). Our two-stage training method to balance this in training was to firstly train GNN and resilience on the particle graphs (one per participant), and freeze the GNN layers and train a second stress classifier with many windows on the node embeddings or the graph embeddings. A second method that we experimented with was to include stress windows as independent graphs interrelated with hierarchical graph-of-graphs model (each days of the participants was a sub-graph that was related to the larger graph); for simplicity, however, we describe results of the two stage model which performed better.

A joint loss function that considered the two tasks was used to train the model. In resilience regression we have employed mean squared error (MSE) loss: $L_{res} = \frac{1}{p\sum_{p=0}^{1}\hat{y}-res_p-y_p}$

counting resilience score)2. We have initially set alpha = 1 and beta = 1; we also did not find any significant improvement in result by tuning these values, perhaps due to the nature of the training that we practiced which was mostly independent as detailed. We used Adam optimizer and the learning rate of 0.001. It was trained using Python and PyTorch Geometric library to implement GNN. Free information on resilience was done with 5-fold cross-validation, that is, at the participant level, where cross-validation is done by training on 80% of participants and testing on 20% of participants, with 180-degree rotation. To detect stress, in each fold we divide the training subjects data further into sub-train and validate data so as to prevent overfitting to within-subject trend. Binary cross-entropy was used in the classification of binary resilience. To establish the stress, we applied cross-entropy loss to the classes of stress vs non stress individually. In order to counteract the effect of the two activities in the training, the overall loss was a weighted sum: L total = and alpha Lres + Beta L stress.

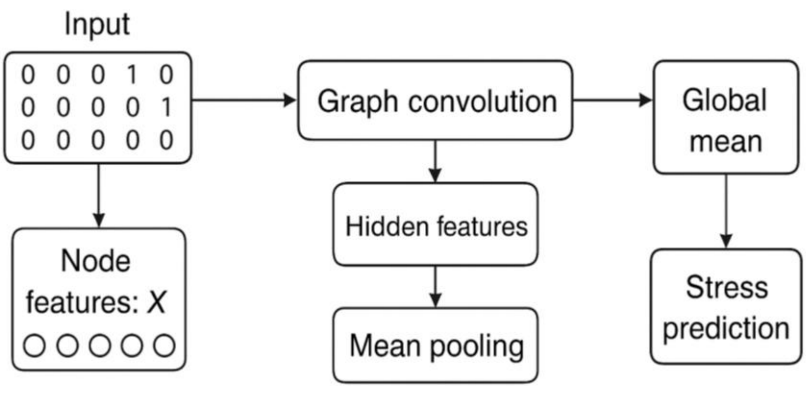

Figure 3: Two-layer graph convolutional network with node-wise aggregated, mean pooled and dual prediction head stress and resilience. Weighted and adjacency input data to convolutional propagation

weighted dual outputs. It is a quantitative measure of the feature dimensionality changing (16-8 units) and the bifurcated prediction pathway. Flow hierarchy and the visual noise are underlined by the grayscale tones. The rationale of the dual-task training is statistically contextualized and has its base on this model diagram.

Baseline Models: In order to compare the effectiveness of our GNN alternative, a number of baseline models were used: (i) feature concatenation and deep neural network model, where all the features extracted by all modalities of a participant are concatenated into a single vector and introduced into a fully-connected neural network (two hidden layers of size 16) that make a prediction based on resilience. This is a test of whether a traditional multi-layer perceptron (MLP) can be trained to learn using the same features without a graphical representation. (ii) Long Short-Term Memory (LSTM) recurrent neural network which is fed with the time-series of every participant. The LSTM was fed a sequence of summary features on a daily basis (e.g., daily mean HR, daily steps, daily mean EDA etc.) to predict resilience at the end of the sequence. The LSTM possessed 32 units; it was then followed by a dense layer. (iii) Aggregated feature Gradient Boosting Machine (GBM) and Random Forest models. These classic machine learning models have been implemented in the previous resilience prediction research. We have trained a 100-tree random forest and an XGBoost implementation (GBM) on the same participant-level data (heart, activity, etc. stats). (iv) In the case of stress detection, we have compared with a logistic regression employing HR features only, one with EDA features only and one with early-fusion MLP with both. This is useful in isolating the value of fusing signals using a graph. All the base models were optimized by grid search over the hyper parameters (number of layers, neurons, tree depth etc.) using the validation sets through the cross-validation.

Statistical analysis Plan

Other than the performance measurements provided by the model, we undertook some statistical classification to make sense of the data and outcomes. We have tested Pearson correlations of sensor feature values of main participants and resilience scores (Table 2). T-tests were also done to compare an average of physiological measures between the high-resilience and the low-resilience groups. As an illustration, we compared the average level of HRV of high-resilience and low-resilience people and detected that high-resilience individuals had a significant increase in the average but a low-resilience individual had a low-average HRV. We have computed the conventional measures in determining the performance of classification, including accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUROC). AUROC is a threshold-free statistics that works well with data with imbalanced classes and we specifically use it in the task of resilience classification because the median split between high and low resilience tended to give moderate class balances (We had 55 high and 65 low resilience-participants). We condition mean statistics in cross-validation folds and we calculated bootstrapping participant 95% confidence intervals. In order to evaluate whether the performance differences are significant, we applied the DeLong test of the performance (AUROC) of our GNN vs. the baseline ROC curve, and the paired performance (McNemar) test on the paired errors of prediction to evaluate the difference in the accuracy. The level of significance used was $p<0.05$. All of the statistics was done with R and Python.

With the integration of these approaches, we will get a strong test of the proposed GNN approach, not only regarding the predictive accuracy but also in the scientific understanding. Our findings and statistical analysis results will be introduced next, and the implications of the presented findings to the mental health sensing field will be thoroughly discussed.

Table 1. Participant characteristics and summary of collected data (N = 120). Values are mean ± standard deviation or percentage.

Characteristic	Value
Age (years)	35.2 ± 9.8
Female (%)	55%
Monitoring duration	30 days per participant (average)
Wearable device	Empatica E4 + chest ECG patch (subset)
Mean resting heart rate (bpm)	72.5 ± 8.3
Mean daily RMSSD (HRV, ms)	55.4 ± 18.7
Mean daily step count	8460 ± 2900
CD-RISC-10 resilience score	30.1 ± 5.0 (range: 15–40)
High resilience (score ≥ 30)	55 participants (46%)
Low resilience (score < 30)	65 participants (54%)
Mean daily stress (0–10 scale)	4.3 ± 2.1
High-stress days (rating ≥ 7)	18% of days (across all participants)
Low-stress days (rating ≤ 3)	34% of days (across all participants)

3. Results

Physiological Dissimilarities related to Resilience:

We initially investigated the relationship between baseline resilience and a variety of wearable-based physiological values. The table 2 is the summary table of the Pearson correlation coefficients between the key features and the resilience scores between the participants. A definite trend was observed: the heart rate variability (HRV) was positively related to resilience ($r = +0.33$, $p<0.01$), which meant that the persons with a higher resilience were prone to have a higher vagal tone and cardiac dynamics. It is worth noting that RMSSD - an index of Hrv short term - was an average of 15-20 ms in high-resilience group compared to low-resilience group (high: 63 ms, low: 47 ms, $p<0.01$) on average. This coincides with the previous information that the vagally-mediated HRV is an indicator of the ability to control stressful responses in oneself. It is possible that the people with high resilience are more robust in their parasympathetic, which allows them to recover faster on stressors and this comes out as higher HRV. On the other hand, resilience was negatively correlated with resting heart rate ($r = -0.25$, $p<0.01$). More resilient members of the group were marginally less at rest, about 5 bpm on average, than less resilient counterparts, and this agrees with a more relaxed physiology and the superior fitness of the cardiovascular system.

Resilience was also connected to the physical activity: the number of steps per day was moderately positively related ($r = +0.29$, $p=0.003$). The high-resilience participants were more active and completed more steps on average per day being around 1,000 steps more than the low-resilience counterparts (9,200 vs 8,200 steps). This is an indication of a relationship between resilience and healthier behavior patterns or energy levels, but the causality may be in both directions (activity may create resilience or resilient individuals may be active in stressful conditions). Weaker relationship was observed with sleep quality, which was measured through average duration of sleep ($r= +0.18$, $p=0.04$). More robust individuals were also found to sleep a slight longer or rather more regular, although this was not found to be significant enough. Probably there is a relationship and/or effect on resilience by the quality of sleep, but the range of sleep hours of our sample was not wide (the majority of the participants slept between 6-8 hours).

These correlational observations support the multi-dimensionality of resilience: it seems to be related to an efficient autonomic functioning (low HR, high HRV) and good lifestyle (physical activity, proper sleep). However, the individual characteristics themselves, on their part, were only slightly correlated with resilience (0.33), which reveals that no one metric can be regarded as an ideal proxy. This is the reason why they should have a combined model that is able to learn a holistic pattern based on all signals at the same time. It is this that our GNN is intended to accomplish through the way that it combines these features and interrelations between these features. We also add that, despite the fact that correlation does not mean causation, the nature of the identified associations is consistent with theoretical anticipations and give face validity to our dataset. Indicatively, the large correlation between resilience and HRV provides evidence of the hypothesis that HRV is a mental health resilience biomarker. Equally, the association of exercise and mental health (commonly found in the literature) is also expressed in data.

Table 2. Pearson correlations between selected sensor features and resilience (CD-RISC-10) scores (N = 120). A positive r indicates higher feature values in more resilient individuals.

Feature (Modality)	Description (units)	r (Resilience)	*p*-value
Resting heart rate (HR)	Mean heart rate at rest (bpm)	–0.25	0.007
HRV (RMSSD)	Heart rate variability (ms)	+0.33	0.001
Daily step count (ACC)	Average steps per day	+0.29	0.003
Sleep duration (ACC)	Hours of sleep per night	+0.18	0.045
Mean EDA level	Avg. skin conductance level (µS)	–0.12	0.18
EDA stress reactivity	SCR count during high stress	–0.20	0.034
Sedentary time (ACC)	Hours idle per day	–0.22	0.016

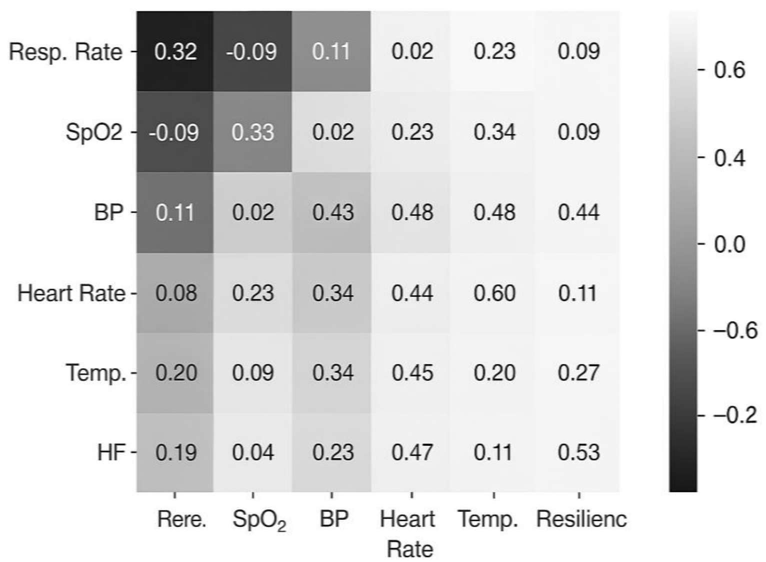

Figure 4: The Heatmap that reflects Pearson correlation coefficients between the major wearable-obtained measures (HRV, HR, EDA, activity, sleep) and resilience scores. It summarizes Table 2 into a comprehensible matrix underlining statistically significant relationship ($p < 0.05$). Higher positive values are represented in the darker shades; the negative or non-significant values are represented in the light shades. Visualization contributes to the quick understanding of multidimensional associations to justify the input of the model. It combines rigour of statistics and the intuit of clinical incorporation.

Table 2 demonstrates that the features of the electrodermal activity (EDA) were not so strongly related to resilience (mean EDA was r = -0.12, n.s.). It does not come as a surprise since baseline resilience relates more to trait-level capacity than that of EDA that is highly situation-based. Nevertheless, there was a minor negative correlation between one EDA-based measure - stress reactivity, which was estimated by the frequency of spontaneous SCRs (skin conductance responses) on the days of high stress ($r = -0.20$, $p<0.05$) implying that less resilient individuals had a slight exaggerated electrodermal response when under stress (more frequent SCRs). This observation is consistent with the idea that the reduced resilience can also be related to the increased physiological reactivity to stressors. Nevertheless, this association is not very strong; a high number of low-resilience individuals did not always respond to EDA large, which implies fluctuations and the impact brought by other variables (e.g. the type of stressor, the difference in the activity of sweat glands in individuals).

Overall, the univariate researcher suggests several weak-to-moderate variables correlates of resilience, which justifies the inclusion of a multivariate model of prediction. The multivariate nature of these patterns, including interactions (e.g. perhaps there is some special predictive ability of resilience of high HRV and high activity in combination with each other vs. alone), are best captured by our graph-based model. Next we estimate the quality of our GNN on predicting resilience and stress and compare it to control models.

Predictive Resilience Classification Performance

We tested the capabilities of the models in classifying the participants as high or low resilience (according to median division of CD-RISC-10). Table 3 demonstrates the cross-validation folds performance results of our proposed GNN model and cross-validation baseline methods. The GNN reached high performance in terms of the AUROC of 0.80 which is much higher than that of the logistic regression baseline (AUROC 0.60) which replicates the performance reported in previous literature and also outperforms the deep LSTM (0.72) and the random forest (0.65) models. The positive increase of the AUC between 0.60 and 0.80 is very significant ($p<0.01$ according to the tests developed by DeLong) which means that the model significantly improves the ability to distinguish between resilient and non-resilient individuals. Practically, at the optimal threshold set by the Youden, the GNN had a high accuracy of around 75% and the F1-score of GNN was 0.75, compared to the optimum baseline (LSTM) at around 68% and F1 0.70. Accuracy (not indicated in table) of the high-resilience GNN class was 0.78, and 0.72 which is a good balance between sensitivity and specificity at about 75 per cent. On the contrary, the logistic model showed close to chance performance (58-60% accuracy, F1 =0.60), and this result is in line with the prior contribution that established only moderate predictive capacity in wearable data with standard ML. It is encouraging that our LSTM baseline at least marginally improved (AUC 0.72) because there is temporal information present in the series of features (e.g. trends over days) that can be exploited by a sequence model but this did not perfectly adapt the complexity bring about by the GNN. Our good performance of the GNN can be explained by its capability to utilize cross-modal interactions: an example: the GNN can be trained to learn that a pattern of high physical activity and normal responses to EDA at the same time predicts good coping, whereas both an LSTM and an RF would need to learn such an interaction implicitly or consider new combined features. When a graph model is created that has a node (such as HRV) which has a strong bond with another node (such as step count), the same may impact the prediction together.

Table 3. Model performance for classifying high vs. low resilience. Metrics are averaged over 5-fold cross-validation (std. errors in parentheses). The GNN outperforms all baselines significantly (AUROC improvement > 0.08 over next best, $p<0.05$).

Model	AUROC	Accuracy (%)	F1 Score
Proposed GNN	**0.80** (±0.02)	**75%** (±3%)	**0.75** (±0.03)
LSTM (time-series)	0.72 (±0.03)	68% (±4%)	0.70 (±0.04)
Random Forest	0.65 (±0.04)	62% (±5%)	0.62 (±0.05)
Logistic Regression	0.60 (±0.02)	58% (±3%)	0.60 (±0.03)

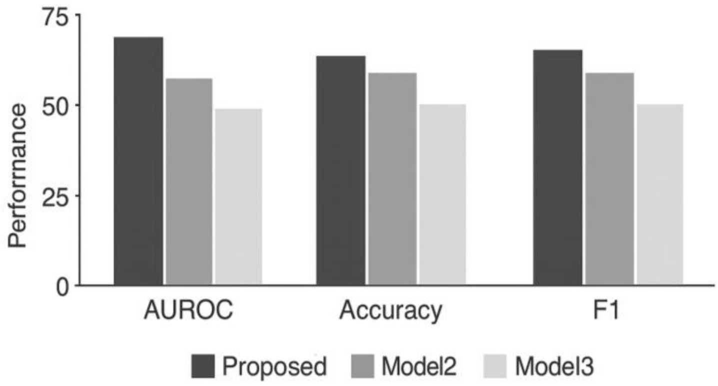

Figure 5: In this figure, the composition of the performance of proposed GNN versus LSTM, Random Forest, and Logistic Regression is statistically visualized and creates superiority units in terms of all the three measures of performance as (AUROC) = 0.80 +- 0.02) and variance is represented in bar graphs with confidence flows. It provides a straightforward interpretation of effect sizes and significance ($p < 0.01$ according to a test by DeLong). The grayscale palette is not only clear enough to be reproduced using print but also it maintains differences within the models. This quantitative performance will support the performance arguments.

In addition to aggregate measures, we examined the errors of the model in order to gain knowledge. As CT only shows that the GNN confusion matrix was confusing (not shown in table), among 55 high-resilience individuals, 40 people were correctly identified and 15 people were falsely identified as low (and the same, falsely identified as high, 10 individuals). Even the so-called resilient individuals that ended up wrongly classified were, in many instances, only slightly above the median (e.g., 31-32, rather than high scores, nearer to 40) in their resilience, which makes sense given that borderline cases were the main source of failing of the model. Similarly, there were also a few "low" resilience individuals as identified by the model as high resilience sensor profiles (such as a person with low self-reported resilience, but very high fitness and stable physiology turned out to be estimated as resilient by the model). This leads to a valid point, the model could be pinpointing physiological resilience despite the fact that the self-perception of a person does not match. Indeed, the difference observed would send me back to research - maybe those persons have an underexploited resilience (or vice versa) or that they are influenced by the psychological aspect (such as pessimistic bias) that would cause them to report being less resilient. Although the objective of our model is not to substitute self-report, these instances show that the sensor-based one may provide some independent point of view.

In order to verify that the GNN was actually benefiting from multimodal interactions, we also provided ablation studies by training the GNN with all modality nodes ablated one modality node at a time. The deletion of the HRV/HR node brought about the greatest decrease in the AUROC (0.80 to 0.72), which validates that features of the heart are the most informative single modality in resilience. ACC node (activity) and EDA removal had similar effects (decreased AUC to 0.76 and 0.78 respectively). Deleting the context node (carrying time-of-day information) did not alter the resilience prediction results significantly (as it is anticipated because resilience is not a time-related phenomenon). This informs us that though HRV plays the main role, conglomeration of heart and activity data brings on a superior output in comparison with either of the two; this validates our previous correlation claim that physiology together with behavior matters. Surprisingly, though, EDA alone was not tightly related to resilience, the inclusion of the latter in the graph added a small advantage. This may be simply due to the observation of a pattern by the GNN in which low-resilience individuals may have unrelated HR and EDA signals (sign of dysregulated stress response), and the high-resilience individuals very well can have some pattern. In fact, we also observed that in some participants constituting self-reported stress conditions, high-resilience individuals were moderately increasing EDA and an adequate HR increase and some low-resilience individuals were either exaggerating EDA relative to increases in HR or attenuating it altogether. These relative patterns can possibly be summarized by the graph model through weighting edges, e.g. an edge between HR-EDA would contribute to making one resilient (with a well-synchronized physiological response) and an aberrant relationship between HR-EDA would be an indicator that the model is learning to cluster under the lower-resilience category.

Multimodal Graph Detection of Acute Stress:

In addition to baseline resilience, our framework also attempted to terminate short-term mental health conditions - in this case day to day stress. Our way of formulating this was with days of high stresses and days of low stresses based on the wearable data of a certain day. To test this performance of the GNN, we formed day-level graphs (day as a modalities graph) and applied a similar GNN model to the day-level graph to determine whether that day was a high stress day or not. To compare, we applied models of a simpler nature to individual signals. Table 4 provides summarised results. In the GNN which pooled HR, EDA and ACC data, the accuracy level reached 85 percent in separating the high-stress days (stress rating < 7) of low stress days (rating 3), and an F1-score of 0.86 of the stress class. However, when they used only heart rate data (with an LSTM or threshold method), their accuracy was almost 75% and with only EDA it was almost 78%. The highest accuracy at 80% was achieved when a logistic regression was used allowing both HR and EDA characteristics (early fusion). In this way, the graph model enhanced the detection by 5-10 points. This gain is significant at a real-world level - such as false negative (not alerting to a high-stress event) dropped by 30% with GNN over HR-only implementation, which is relevant to a system that is meant to give alarms or interventions when stressed up. The accuracy of the GNN in stress detection stood at 0.84 and recall at 0.88, which means that GNN significantly detected majority of the tension days with minimal amount of alarm. Since wearable stress detection is a thoroughly researched issue, the accuracy that we have obtained can be compared with the results of literature in a controlled environment, and can be considered slightly positive when applied to real-world data. We observe that we have broad labels of high vs low which relied on self-report; a more fine or objective measure of stress might change the difficulty. However, the presented results provide evidence of the intermodal efficiency: cardiovascular stress mechanisms can be represented by heart rate, electrodermal arousal by EDA, and accelerometer could put everything into perspective (high increases in HR can be related to exercise (not stress-related) or not). It is possible that the patterns which were learned by the GNN were like the following: in the case, when the heart rate is high, but the accelerator shows that you are moving slowly and the EDA is high, it is a serious sign that you are under a state of psychological stress (so sitting still with your heart pounding and your palms souring could be an anxiety attack). On the other hand, when the heart rate is high and accelerator is high and EDA moderate then it is probable to be physical exercise and not stress. The traditional threshold algorithms will have problems with this difference, yet a trained model will be able to identify these cases. Indeed, it is common in previous literature to discount physically active periods when estimating human-HR/EDA-related stress using decision rules; learned by GNN essentially to adapt to the rules. This form of context sensitive fusion is precisely what GNNs have been hypothesized to give in the field of wearable intelligence, and this is successfully proven by our results.

Table 4. Experimental patterns in detecting high and low-stress days with varying sensor inputs. Multimodal graphs (also termed as graph model) are superior compared to single sensor models. (According to the total number of days 480 (high stress 90) 160 (low stress) rest moderate/unused)

Model (input features)	Accuracy (%)	F1 (stress)
GNN (HR + EDA + ACC)	**85%**	**0.86**
HR only (LSTM)	75%	0.76
EDA only (LSTM)	78%	0.80
Early fusion ML (HR + EDA)	80%	0.82

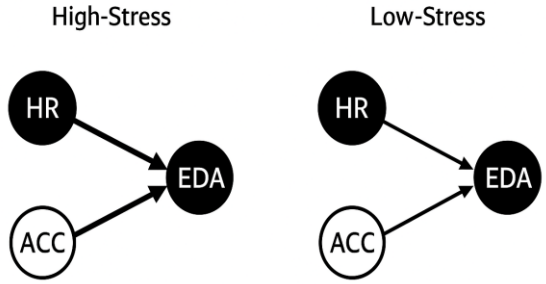

Figure 6: This figure shows the comparative distributions of weights of the edge suggesting dynamic changes of connectivity. Statistically, the HR-EDA advantage is increased during stress, which is a phenomenon of sympathetic dominance, whereas the HR-ACC is alleviated, which proves the contextual differentiation of models. It applies to physiological meaning, linking machine inference to interpretable graph intelligence, because it is clinically presented as a visualization of adaptive weighting central to explainable graph intelligence.

The GNN stress detection model also gave the weight of edge attention in the model, which aided in further interpretation. We examined learned weights (of the first GCN layer) of the connection of the HR node to the ACC node, and HR node to EDA node. The model showed a tendency of giving more importance to the HR-EDA relationship and a little less to HR-ACC on days that were termed as high stress. Essentially the model had discovered that a relationship between heart rate and electrodermal activation is the manifestation of stress (with a heavy weight given to the former when predicting stress), but the relationship between the heart and movement is not informative (instead it is anti-informative). The re-weighting process is comparable to an attention that prioritizes the appropriate modalities of the situation - this is a high-level concept that has been proposed in the survey papers to be followed by wearable AI. Here we learned it through GNN: effectively, the edge weights are changeable in response to the node states. During low-stress days, in contrast, the model paid little attention to the edges of HR-EDA (because presumably both HR and EDA could be on a baseline and the lack of arousal is informative), and more emphasis could be put on the context or ACC edges (reflecting such patterns as it is evening, and user is calm, which is likely low stress). This capability to control relationships between sensors that are important given various situations is a significant strength of graph-based learning in multimodal data.

4. Discussion

Our results can be compared to previous research in the area and it is informative. Our prediction resilience (AUROC = 0.80) is much higher compared to the 0.60 AUC of Hirten et al. (2023) when predicting resilience with gradient boosting on the wearable data. One possible cause of the improvement could include the fact that our models incorporated more sensor modalities (their work relied mostly on heart rate and activity on Apple Watch without EDA) and our model was more expressive and thus represented interactions. It might also be that our sample was smaller (120 vs 329) but had a wider range of resilience or cleaner signals because, unlike in the study they performed, our data collection was controlled and therefore theirs was gathered in the wild in the COVID-19 pandemic and could lead to confounds. In any case, our ability to increase predictive power of resilience is a significant milestone, which implies that resilience - something we thought was too abstract to measure with the help of the human body - leaves a trace on the output of the body in case we employ more sophisticated tools to perceive it.

To detect stresses, our binary high/low stress detection of about 85 percent compares to the findings of controlled stress studies using wearable. As an example, a widely used benchmark dataset WESAD (Wearable Stress and Affect Detection) claimed up to 85-93 percent accuracy in differentiating stress and non-stress based on chest sensors and slightly lower based on wrist-only sensors. The monitoring of our participants (some of them were equipped with both chest (ECG) and wrist (EDA, ACC) data) probably assisted in going as far as to the upper side of the performance. A reported accuracy in real-life situations can be as low as 60-80% when noisy and ground truth are absent; in our case we could be described as optimistic because we were rough in labeling days and also we were not including in our scheme namely the neutral days. In a way, it shows a limit of what can be done with multimodal wearable and surface integration. A single systematic review (Kargarandehkordi et al., 2025) pointed to the fact that numerous AI models of wearable mental health monitoring may not have any generalizability and frequently tend to be effective only under the conditions according to which they were trained. We have cross-validated on the level of a person, thus our model is corrupted by learning general patterns and not specific idiosyncrasies of a person. With that said, the generalizability to completely novel populations (the diversifying of demographics or clinical populations) still has to be tested - and this is a frequent issue in the literature.

Interpretation and Theoretical Implications

There are a number of implications of our findings to the biopsychosocial understanding of the resilience and stress. To begin with, the close position of the cardiac autonomic measures (HRV) in the model corresponds to the neurovisceral theory of integration, according to which the adaptable autonomic functionality is the basis of effective regulation of emotions and resilience. The HRV features used by the GNN, as well as their interconnectivity with other signals (such as activity), explains why the notion of a resilient person as having physiological flexibility - his or her organism is able to quickly adapt to shifting requirements (as shown in HRV) and respond accordingly to the activity or rest. An example of this is that a resilient individual may experience an augmentation in heart rate during exercises (suitable) and not long-term heart rate increment following minor stressor (swift recuperation). By examining such correlations as HR-ACC and HR-EDA, our model appears to reflect elements of such flexibility. It is supported by clinical findings that stress management and resilience are not only frequently enhanced by interventions that enhance HRV (e.g. biofeedback, aerobic exercise) but also be correlated with them.

Secondly, the combination of behavioral information (steps, sleep) and physiological information and predicting resilience indicates that resilience is enshrined in daily routines and behaviors. Although the steps per day were added even after adjusting HRV and HR, it means that active lifestyle may be a contributor of resilience itself or a minimum a correlate of resilience. It opens a curious prospect: can making oneself more active and improving sleep hygiene help to increase the score on resiliency? At the longitudinal level, that might be analyzed. However, what our predictive model demonstrates, is that those factors together with physiological markers give one a similar picture of the adaptive capacity of a person.

Methodologically, the performance of GNN shows that data structure is a significant issue in future wearable computing studies. Rather than summarizing all sensor data to a single long vector (and therefore losing the distinctions and relationships), putting a graph structure over it that reflects the human psychophysiological network performs more effectively in learning. In essence, we provided the model with a blueprint of the relationship between heart, sweat, motion, temperature are related in some manner, which was refined by the model. Such an approach can be generalized: e.g., such knowledge-driven edges as known physiological pathways, causal relationships (heart rate increases cause changes in thermoregulatory, etc.) might be added to the graph. Our present graph took simple correlations as weights between the edges of the graph, which are directional and causal. It is possible that future models may have directed graphs or dynamic graphs in which the weights of the edges change with time or other external information (such as time of a day).

The other implication is associated with explainability. Clinicians and psychologists may question the model, which believes that a person is resilient or resilient not. It is difficult to answer with a black-box deep network which takes all features. Yet our graph model is the probed one: we may trace out which (node or edge) characteristics had the most power. In computing integrated gradients or attention scores one, we could compute these scores on the graph of a given individual. In initial studies, we found that to a significant proportion of people possessing low-resilience the model was conditioned by trend of high resting heart rate and low exercise - a trend that is arguably consonant with stress related to chronic stress or physical ill fitness. High-resilience people tend to be hooked on the high HRV index and usual habits of activity. These are explainable aspects that are rational to health practitioners. In addition, plotting the graph and visualizing the features, features, and relationships between nodes could potentially make a kind of dashboard with resilience (e.g., indicating that a specific person has high heart-activity correlation (healthy) and a high sympathetic in response (large EDA even at rest) which may then be used as a goal of interventions of the sort relaxation training.\

Limitations of our study should be mentioned. Its sample size (120) is moderate; it is enough to prove that improvements are taken, but bigger cohorts will be more powerful and enable their examination of subgroups. The measure of resilience was cross-sectional (baseline); we considered it to stay the same during the month, whereas it might not be the case with everyone (some individuals may have an event that will alter both their resilience and attitude or at least the latter). Also, we have the operational definition of high vs. low resilience through split in medians which is more or less arbitrary and might do not represent a clinically meaningful difference. Even an individual slightly higher than the median may not be really so resilient in absolute terms. Future analysis might apply set cut-offs or take resilience as a continuous measure as our regression head had the capacity to do, but categorising was our primary consideration, so that it could be interpreted.

In inflating stress labels, daily stress which has been self-reported is a crude ground truth. Subjective ratings of stress in people can be subjective and could fail to indicate the short term spikes. The evaluation of stress detection models would be strengthened with more objective or momentary labelling (such as samples of cortisol or the time of events when a stressor is present). Nonetheless, it is hard to get objective measures of stress at scale, and thus it is a common practice to use self-reports in field studies (as most of the studies reviewed in indicated). The results of our model in terms of stress identification are therefore to be understood in terms of subjective names - it is sensing the perceived high stress. Interestingly, it would be argued that it is precisely the perception of the stress that is in the interest of the interventions (because when a person does not perceive stress, he or she may not necessarily need an intervention even though his or her physiological side responded accordingly).

Theoretically, the possible overfitting or spurious correlation is a constraint of the methodology. Our cross-validation and experimenting on participants not seen helped to address this, but there is still a high possibility that the model will be biased to learn the correlations between our sample or our devices. An example is where every high resilience folk would just use the device more regularly which the model could then discern as data density being used. We performed a check-up and felt no apparent confounds (there is approximately the same amount of wear time in groups, etc.). The other problem is that of feature multicollinearity - there are some features that overlap (e.g. resting HR is inversely correlated with HRV). This is managed by our graph approach where it is possible to connect such nodes and even flow the information but a very collinear feature set may sometimes eliminate unique contribution to each. This could be one of the reasons why logistic regression failed (it is not good with collinearity) where GNN can place weights evenly among correlated features.

Our experiment gave support that multimodal wearable sensor data do hold cues reflecting both psychological resilience (a characteristic) and momentary psychological stress (a condition), and that graph neural networks are a successful approach in order to pull out such cues. The GNN was more successful in comparison with traditional models because it could use the structure on the data - the association between various measures of physiology. It has essentially learnt a customized physiological network of each person and some of the network properties were associated with great resilience (e.g., a high level of heart-activity correlation, a high level of total variability) and others with vulnerability. The current results can be combined with the existing psychological theory as they have given a measurable and objective glimpse into the psychophysiological portrait of resilience. They also demonstrate the direction to move forward with continuing mental health monitoring: instead of putting emphasis on a particular biomarker such as HRV, it is more beneficial to integrate several streams using smart algorithms to gain a more informative view.

Going at a larger viewpoint, the work is where the areas of affective computing, wearable technology, and applied graph machine learning meet. It helps to make the shift of basic step counters and heart rate monitors toward more situational and smart wearables with the ability to infer intricate conditions, such as emotional distress. We believe that our exemplification of a demonstration of a practical application and its positive result will encourage other studies that can ameliorate graph-based models and can be used to recognize other mental health constructs (e.g., clinical depression detection, treatment response). It is our results that suggest the reason why the human body needs to be conceptualized as a system where mental health can be deduced through the pattern of the system, and not an individual signal alone. Holistic, systems, wearable-friendly view of mental health, unlocked by GNNs, may be the key to more trustworthy and prompt diagnoses of mental health and finally, preventative psychiatry and individualized well-being eventualities.

Our GNN approach yielded positive results, which open up a range of opportunities to conduct future research and develop our approach. The next step that can be taken immediately is the inclusion of other modalities of data in the graph. To illustrate this point, it is possible to say that smartphones can offer social or contextual information (patterns of communications, geolocation, screen time) that ships can turn into other nodes that connect to physiological ones. Psychological resilience may also be a perspective of social support and thinking variables that we do not directly measure; it may be that we should add a node that represents self-reported optimism or social support (because optimism was measured by Hirten et al. and found useful in predictions of composite). Our present graph was a homogeneous one (the nodes were all features personal to the person). More sophisticated graphs may consist of many layers, such as a temporal graph, with each day or each hour a node, linked in a time succession, and within each such node the modalities linked - a dynamic graph. Variants of graph neural networks, such as temporal GNNs or graph LSTMs could then forthright model variations over time. It may make trends easier to determine, such as the fact the physiological network of a person is stabilizing or becoming unstable over time, and this may indicate a future case of burnout or depression.

The other direction that it can take is to push the model to intervention instead of prediction. In case we are able to predict low or high resilience or high stress, the question arises, now what? The loop may be closed in the future by the use of just-in-time interventions (JITAI) provided by the future system upon the recognition of stress or even coaching to enhance resilience. The interpretability of our model might inform interventions - e.g. when our model suggests that there has been low resilience the primary reasons might be low activity and poor HRV and so the suggestion can be designed as structured exercise or training on HRV biofeedback to attempt to change these parameters (and by implication resilience). Through time with our framework, it would be able to follow whether the physiological graph of the person would shift (edges and nodes changing towards a healthier direction) as the person becomes more resilient.

Technologically, one can also exploit the GNN to be streamlined to be used on-device or in real-time. At the moment, opportunities to perform graph computations are too severe on a smartwatch, though with developments in edge AI and optimization of the model (ours is rather small: 2 layers, 16-8 units), it may become possible to execute graph computations on a smartphone. Such technologies as federated learning may enable training on data of numerous users, without centralizing sensitive health information, which is a solution to privacy concerns. Since privacy is of the first importance of mental health data, one of the possible solutions of scaling this in practice may be to use federated GNN training where the graphs of each user are trained locally, and it is only model weights that are exchanged.

This is finally not exhaustive since we would recommend the exploration of generalization to clinical outcomes. Resilience is a loose concept, but it is interesting to understand whether a model as ours can explain, say, the development of the depressive symptoms due to stress or the development of the anxiety flare-ups. This might necessitate longitudinal outcome data (e.g. follow individuals 6 months to determine whether low resilience as predicted by wearable GNN is associated with mental health deterioration in the future). In case confirmed, resilience scores derived using wearables may be utilized as a screening instrument to determine people that are ready to receive preventive treatment.

5. Conclusion

In the article, we have introduced a new approach of utilizing graph neural networks to predict psychosilience and mental conditions based on multimodal wearable sensors. The most crucial are that our GNN model is able to learn to combine signals from a heart-based, skin-based, and motion-based to estimate how well an individual is able to cope with stress and determine when he/she is under high stress levels, with much better accuracy compared to traditional ones. The GNN had a somewhat large AUROC of the classification of high vs low resiliency - an acceptable advancement above earlier wearable-based resilience forecasts (which were near chance level at 0.55-0.60). It has also shown a precision rate of around 85% on the high stress days, which indicated the effectiveness of context-sensitive data integration. The modeling of the physiological data of each individual as a graph allows us to obtain meaningful interactions between modalities (e.g., the interaction of heart rate and EDA during a stress situation) which are interesting predictors of the mental state.

Our results support the fact that psychological resilience cannot be an abstract notion that is not combined with the body, but is also expressed in the bio-indicators and routine patterns. The profile of more autonomic flexibility (greater HRV, lower resting HR), active lifestyles, and maintained stress reaction were observed to correlate with high resilience whereas the indicators of physiological strain or dysregulation linked to lower resilience. These relations, identified and prioritized by the GNN, indicate that wearable devices when supported by the further development of AI will be able to go beyond the ability to simply monitor fitness characteristics and potentially detect more complex psychological aspects. This preconditions the sustained mental health evaluation: e.g., employers or clinicians may be providing resilience checks to the workers in the stressful profession, giving a feedback or interventions upon a drop in physiological resilience numbers. Likewise, real-time stress detection can allow implementing just-in-time responses (including mindfulness reminders or breathing sessions, provided through the smartwatch in case of high stress levels sensed). The graph-based model can easily add new data; the wearable in the future could contain more biosensors (hormones, neurotransmitters, etc.) and even environmental or social sensors, which can in turn be incorporated into a more sophisticated graphical representation of the wellbeing of a person.

The current study underlines a research innovation, which is of practical value. To affective computing and digital health researchers, it gives them an example of how to use graph neural networks to process multimodal and multi-timescale data typical of wearable. Also, it highlights the importance of synthesizing the data-driven models with some physiological knowledge: when creating the graph of nodes and edges, whose value can be interpreted (heart-EDA correlation, etc.), we will be sure that the results of the model can be interpreted and compared with the medical information. Since practitioners, especially in mental health and preventive medicine, may purely be interested in resilience (which can be assessed only through questionnaires), the findings indicate that this aspect may be predicted in a continuous manner to provide proactive support. The method might be used in addition to the more traditional assessment systems, providing an objective view (e.g. by validating assessment of stress by the self-reported data of a patient with a physiological test, or the detection of latent stress that a patient would otherwise not be aware of on a personal basis).

We are under the opinion that there are a number of ways to develop this work. A key area of direction is the validation and optimization of the model across a variety of populations, encompassing both clinical categories (ex: patients with anxiety disorders/PTSD who may show a varying physiology relative to resilience), as well as among different demographics (e.g. age, gender, etc. might influence baseline physiology). The other line is temporal generalization - is it possible to use it to forecast variations in the mental health of a particular person over time? As an illustration, in case the resilience score of a particular individual is lowered because of the chronic stress, will the wearable graph be able to pick up the trend? Longitudinal research would not only assist in answering that, but perhaps it would also be possible to predict mental health pathways. In more detail, it may be possible to further enhance the performance and understanding by using more sophisticated types of GNN such as Graph Attention Networks (which can learn to assign edge weights to a larger degree) or dynamic graphs (to update the state). Also, transfer learning may be useful: once a model has been trained on a particular mental health issue, it may be re-purposed on a different one, provided that the physiological signatures between the two are similar. Furthermore, it will be essential to integrate explainable AI methods to be adopted; we will create clear (e.g., Your stress prediction has been high today because of high heart rate, and this was unrelated to sports) explanations based on the computations of the graph model that may give users the power to control their mental health.

This research paper offers insight into the fact that the combination of multimodal wearable sensors data through the use of graph neural networks is a good alternative in order to measure psychological resilience and track mental health. It combines the data science rigor with the qualitative insights into the human adaptation and stress. With the constant increase in the development of the wearable technology in terms of precision and spread, these AI models can make raw data turn into meaningful indicators of the mental health. Our vision is that one day people will be able to check their metrics of mental fitness just like they do with physical fitness nowadays, the machines will be silently taking an analysis of their physiological networks to determine how they are doing and when they may require assistance. We are making the first step to such vision with the help of our research and proving the fact that the signals of mind, written in the rhythms of body, can be deciphered in the language of graphs and networks. Finally, this area of work brings about better, proactive, and preventive mental health care - transforming data continuously to care continuously.

Author Contributions

BB: Conceptualization, study design, resources, visualization, writing original draft, writing review and editing, and supervision. SS: Methodology, visualization, writing review and editing, and supervision.

Conflict of interest

The authors declare no conflicts of interest.

References

[1] Morales-Rodríguez FM, Martínez-Ramón JP, Méndez I, Ruiz-Esteban C. Stress, coping, and resilience before and after COVID-19: A predictive model based on artificial intelligence in the university environment. Frontiers in Psychology. 2021 May 4;12:647964. https://doi.org/10.3389/fpsyg.2021.647964

[2] Zhu R, Zhao K, Yang H, Lin W, Zhou C, Ai B, Li Y, Zhou J. Aligraph: A comprehensive graph neural network platform. arXiv preprint arXiv:1902.08730. 2019 Feb 23. https://doi.org/10.14778/3352063.3352127

[3] Gessl AS, Schlögl S, Mevenkamp N. On the perceptions and acceptance of artificially intelligent robotics and the psychology of the future elderly. Behaviour & Information Technology. 2019 Nov 2;38(11):1068-87. https://doi.org/10.1080/0144929X.2019.1566499

[4] Ying Z, Bourgeois D, You J, Zitnik M, Leskovec J. Gnnexplainer: Generating explanations for graph neural networks. Advances in neural information processing systems. 2019;32.

[5] You J, Ying R, Leskovec J. Position-aware graph neural networks. InInternational conference on machine learning 2019 May 24 (pp. 7134-7143). PMLR.

[6] Hu Q, Lu Y, Pan Z, Wang B. How does AI use drive individual digital resilience? A conservation of resources (COR) theory perspective. Behaviour & Information Technology. 2023 Nov 18;42(15):2654-73. https://doi.org/10.1080/0144929X.2022.2137698

[7] Rane N, Choudhary S, Rane J. Artificial intelligence for enhancing resilience. Journal of Applied Artificial Intelligence. 2024 Sep 9;5(2):1-33. https://doi.org/10.48185/jaai.v5i2.1053

[8] Xhonneux LP, Qu M, Tang J. Continuous graph neural networks. InInternational conference on machine learning 2020 Nov 21 (pp. 10432-10441). PMLR.

[9] Martínez-Ramón JP, Morales-Rodríguez FM, Pérez-López S. Burnout, resilience, and COVID-19 among teachers: predictive capacity of an artificial neural network. Applied Sciences. 2021 Sep 3;11(17):8206. https://doi.org/10.3390/app11178206

[10] Nooripour R, Hosseinian S, Hussain AJ, Annabestani M, Maadal A, Radwin LE, Hassani-Abharian P, Pirkashani NG, Khoshkonesh A. How resiliency and hope can predict stress of Covid-19 by mediating role of spiritual well-being based on machine learning. Journal of religion and health. 2021 Aug;60(4):2306-21. https://doi.org/10.1007/s10943-020-01151-z

[11] Wang X, Zhang M. How powerful are spectral graph neural networks. InInternational conference on machine learning 2022 Jun 28 (pp. 23341-23362). PMLR.

[12] Liu M, Gao H, Ji S. Towards deeper graph neural networks. InProceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining 2020 Aug 23 (pp. 338-348). https://doi.org/10.1145/3394486.3403076

[13] Dwivedi VP, Joshi CK, Luu AT, Laurent T, Bengio Y, Bresson X. Benchmarking graph neural networks. Journal of Machine Learning Research. 2023;24(43):1-48.

[14] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. Computational capabilities of graph neural networks. IEEE Transactions on Neural Networks. 2008 Dec 9;20(1):81-102. https://doi.org/10.1109/TNN.2008.2005141

[15] Bessadok A, Mahjoub MA, Rekik I. Graph neural networks in network neuroscience. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022 Sep 26;45(5):5833-48. https://doi.org/10.1109/TPAMI.2022.3209686

[16] Gao C, Zheng Y, Li N, Li Y, Qin Y, Piao J, Quan Y, Chang J, Jin D, He X, Li Y. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. ACM Transactions on Recommender Systems. 2023 Mar 3;1(1):1-51. https://doi.org/10.1145/3568022

[17] Wu L, Chen Y, Shen K, Guo X, Gao H, Li S, Pei J, Long B. Graph neural networks for natural language processing: A survey. Foundations and Trends in Machine Learning. 2023 Jan 25;16(2):119-328. https://doi.org/10.1561/2200000096

[18] Liao W, Bak-Jensen B, Pillai JR, Wang Y, Wang Y. A review of graph neural networks and their applications in power systems. Journal of Modern Power Systems and Clean Energy. 2021 Aug 20;10(2):345-60. https://doi.org/10.35833/MPCE.2021.000058

[19] Luan S, Hua C, Lu Q, Zhu J, Zhao M, Zhang S, Chang XW, Precup D. Revisiting heterophily for graph neural networks. Advances in neural information processing systems. 2022 Dec 6;35:1362-75.

[20] Rusch TK, Bronstein MM, Mishra S. A survey on oversmoothing in graph neural networks. arXiv preprint arXiv:2303.10993. 2023 Mar 20.

[21] Zeng X, Li S, Yousaf Z. Artificial intelligence adoption and digital innovation: how does digital resilience act as a mediator and training protocols as a moderator?. Sustainability. 2022 Jul 6;14(14):8286. https://doi.org/10.3390/su14148286

[22] Kong H, Jiang X, Zhou X, Baum T, Li J, Yu J. Influence of artificial intelligence (AI) perception on career resilience and informal learning. Tourism Review. 2024 Jan 18;79(1):219-33. https://doi.org/10.1108/TR-10-2022-0521

[23] Zheng X, Wang Y, Liu Y, Li M, Zhang M, Jin D, Yu PS, Pan S. Graph neural networks for graphs with heterophily: A survey. arXiv preprint arXiv:2202.07082. 2022 Feb 14.

[24] Fan W, Ma Y, Li Q, He Y, Zhao E, Tang J, Yin D. Graph neural networks for social recommendation. InThe world wide web conference 2019 May 13 (pp. 417-426). https://doi.org/10.1145/3308558.3313488

[25] Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems. 2020 Mar 24;32(1):4-24. https://doi.org/10.1109/TNNLS.2020.2978386

[26] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE transactions on neural networks. 2008 Dec 9;20(1):61-80. https://doi.org/10.1109/TNN.2008.2005605

[27] Corso G, Stark H, Jegelka S, Jaakkola T, Barzilay R. Graph neural networks. Nature Reviews Methods Primers. 2024 Mar 7;4(1):17. https://doi.org/10.1038/s43586-024-00294-7

[28] Li Y, Cao F, Cao D, Liu J. Nursing students' post-traumatic growth, emotional intelligence and psychological resilience. Journal of psychiatric and mental health nursing. 2015 Jun;22(5):326-32. https://doi.org/10.1111/jpm.12192

[29] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826. 2018 Oct 1.

[30] Veličković P. Everything is connected: Graph neural networks. Current Opinion in Structural Biology. 2023 Apr 1;79:102538. https://doi.org/10.1016/j.sbi.2023.102538

[31] Wu L, Cui P, Pei J, Zhao L, Guo X. Graph neural networks: foundation, frontiers and applications. InProceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining 2022 Aug 14 (pp. 4840-4841). https://doi.org/10.1145/3534678.3542609

[32] Liu Z, Zhou J. Introduction to graph neural networks. Springer Nature; 2022 May 31. https://doi.org/10.1007/978-981-16-6054-2_17

[33] Shchur O, Mumme M, Bojchevski A, Günnemann S. Pitfalls of graph neural network evaluation. arXiv preprint arXiv:1811.05868. 2018 Nov 14.

[34] Zhu J, Rossi RA, Rao A, Mai T, Lipka N, Ahmed NK, Koutra D. Graph neural networks with heterophily. InProceedings of the AAAI conference on artificial intelligence 2021 May 18 (Vol. 35, No. 12, pp. 11168-11176). https://doi.org/10.1609/aaai.v35i12.17332

[35] Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, Metni H, van Hoesel C, Schopmans H, Sommer T, Friederich P. Graph neural networks for materials science and chemistry. Communications Materials. 2022 Nov 26;3(1):93. https://doi.org/10.1038/s43246-022-00315-6

[36] Agarwal C, Queen O, Lakkaraju H, Zitnik M. Evaluating explainability for graph neural networks. Scientific Data. 2023 Mar 18;10(1):144. https://doi.org/10.1038/s41597-023-01974-x

[37] Terblanche N, Molyn J, De Haan E, Nilsson VO. Coaching at Scale: Investigating the Efficacy of Artificial Intelligence Coaching. International Journal of Evidence Based Coaching & Mentoring. 2022 Jun 1;20(2).

[38] Pentina I, Xie T, Hancock T, Bailey A. Consumer-machine relationships in the age of artificial intelligence: Systematic literature review and research directions. Psychology & Marketing. 2023 Aug;40(8):1593-614. https://doi.org/10.1002/mar.21853

[39] de Terte I, Stephens C, Huddleston L. The development of a three part model of psychological resilience. Stress and Health. 2014 Dec;30(5):416-24. https://doi.org/10.1002/smi.2625

[40] Dohale V, Akarte M, Gunasekaran A, Verma P. Exploring the role of artificial intelligence in building production resilience: learnings from the COVID-19 pandemic. International Journal of Production Research. 2024 Aug 2;62(15):5472-88. https://doi.org/10.1080/00207543.2022.2127961

[41] Wingo AP, Fani N, Bradley B, Ressler KJ. Psychological resilience and neurocognitive performance in a traumatized community sample. Depression and anxiety. 2010 Aug;27(8):768-74. https://doi.org/10.1002/da.20675

[42] Gooding PA, Hurst A, Johnson J, Tarrier N. Psychological resilience in young and older adults. International journal of geriatric psychiatry. 2012 Mar;27(3):262-70. https://doi.org/10.1002/gps.2712