



# Mental stress detection from ultra-short heart rate variability using explainable graph convolutional network with network pruning and quantisation

V. Adarsh<sup>1</sup> · G. R. Gangadharan<sup>1</sup>

Received: 5 March 2023 / Revised: 15 November 2023 / Accepted: 16 December 2023 /  
Published online: 22 January 2024

© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2024

## Abstract

This study introduces a novel pruning approach based on explainable graph convolutional networks, strategically amalgamating pruning and quantisation, aimed to tackle the complexities associated with existing machine learning and deep learning models for stress detection using ultra-short heart rate variability analysis. These complexities often impede the implementation ability of such models on resource-limited devices. The proposed method exhibits exceptional performance, demonstrating high accuracy (97.75%) and efficiency (97.66%) on the WESAD dataset, along with an impressive accuracy (94.48%) and efficiency (94.39%) on the SWELL dataset. Importantly, the runtime complexity saw a significant reduction, down by 63.4% and 69.34% compared to the original model. The proposed method's notable advantage lies in its ability to retain nearly all of the initial model's performance with negligible loss, even when the pruning levels are below 60%. This innovative approach, thus, offers a promising solution for effective stress detection, specifically designed to operate smoothly on devices with limited resources.

**Keywords** Explainability · Graph convolution network · Pruning · Quantisation · Stress detection

## 1 Introduction

With the advent of technology, computer-aided diagnostic systems have emerged as a powerful force in the medical field, serving as effective tools in the prediction and detection of various diseases. These systems have introduced a new layer of transparency and reliability to medical decision-making processes. The incorporation of Artificial Intelligence (AI) in

---

Editors: Dino Ienco, Robert Interdonato, Pascal Poncelet.

✉ G. R. Gangadharan  
geeyaar@gmail.com

V. Adarsh  
adarshv96@gmail.com

<sup>1</sup> Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India

healthcare, particularly within hospital settings, is transforming the medical landscape. The enhanced accuracy in diagnosis and provision of medical advice are just a few of the benefits reaped from this integration.

Mental stress can have a severe impact on a person's overall well-being and can lead to a variety of physical and emotional health problems (Harvard Health, 2020). Some of the common symptoms of mental stress include difficulty sleeping, fatigue, irritability, difficulty concentrating, and changes in appetite. Mental stress is primarily a physiological reaction to external stimuli that are brought on by the sympathetic nervous system. During this stage of the response, a variety of chemicals are produced, including cortisol and adrenaline, among others. These chemicals cause an increase in the rate of the heartbeat and the rate of breathing, in addition to a tightening of the muscles. These physiological changes are getting the body ready for a physical response (called "fight-or-flight") (Bracha et al., 2004). Chronic mental stress can also increase the risk of developing more serious health problems, such as heart disease, high blood pressure, and depression. It is vital for individuals to find ways to manage their stress and maintain their mental health (Adarsh et al., 2023; Chrousos & Gold, 1992; McEwen & Stellar, 1993; Rosmond & Björntorp, 1998; Selye, 1976).

Heart rate variability, commonly referred to as HRV, is a metric that assesses the fluctuations in the time interval between heartbeats. It serves as a valuable tool in determining an individual's physiological condition (He et al., 2019; Moridani et al., 2020; Oskooei et al., 2019). Higher HRV is generally associated with a greater ability to adapt to stress and a healthier overall autonomic nervous system function, while lower HRV is associated with increased stress and decreased health. HRV can be measured using various techniques, including electrocardiography (ECG), photoplethysmography (PPG), and accelerometry. It is important to keep in mind that HRV features should be interpreted in the perspective of the person's overall physical and mental state, and they should not be used alone to draw any conclusions about stress levels.

Ultra-short HRV (US-HRV) (Salahuddin et al., 2007) refers to the measurement of HRV over very short periods of time, typically ranging from a few seconds to a few minutes. It is typically measured using continuous ECG or PPG recordings and can be used to assess an individual's physiological state in real-time. One significant difference between HRV and US-HRV is the time frame over which they are measured. HRV is typically measured over periods of several minutes to several hours, while US-HRV is measured over much shorter periods of time. Thus, HRV provides a longer-term view of an individual's physiological state, while US-HRV gives a more immediate and dynamic view. Further, US-HRV is more sensitive to changes in an individual's physiological state than HRV. This is because US-HRV captures changes in HRV that occur over very short periods of time, which may be missed in longer-term HRV measurements.

Deep learning has seen widespread use in the analysis of image data using convolutional neural networks (CNNs). The success of CNNs in areas with common and regular domains, such as computer vision and speech recognition, has paved the way for an increased focus on Graph Neural Networks (GNNs). These networks are predominantly engaged in reinterpreting the concept of convolution for graph structures (Wu et al., 2021). Graph Convolutional Networks (GCNs) have witnessed a surge of applications across various fields in recent years, particularly in the healthcare sector. Such as in brain analysis (Li et al., 2021), mammography assessments (Du et al., 2019), and image segmentation tasks (Soberanis-Mukul et al., 2020). Deep neural networks, despite their ability to model complex relationships between input and output variables, are often seen as "black box" models (Shao et al., 2021). This is due to their inherent complexity, which makes it challenging

to understand the role a specific input feature plays in generating the output. This lack of interpretability is a significant obstacle, particularly in clinical applications where decision-making processes need to be explained to comply with regulatory requirements, such as the European Union's General Data Protection Regulation (GDPR). The GDPR demands that any automated decision-making processes must be explainable and that patients have the right to refuse automated decisions.

There have been a significant amount of research on using machine learning (ML) and deep learning (DL) algorithms to detect stress using ultra-short HRV (Ishaque et al., 2021; Kim et al., 2018; Lawanont et al., 2019; Pourmohammadi & Maleki, 2020; Rodríguez-Arce et al., 2020; Salahuddin et al., 2007; Sánchez-Reolid et al., 2020; Zalabarria et al., 2020; Zangróniz et al., 2018; Zubair & Yoon, 2020). However, with the increase in parameters and inputs, the ML/DL models generated become more complex and huge in size, and new methods for trimming down their size is essential if they are needed to be implemented on resource-constrained devices.

Model pruning (Abbasi-Asl & Yu, 2021; Dong et al., 2022) is a technique for reducing the size and complexity of a machine-learning model by removing unimportant or redundant parameters. The fact that more rounds of pruning are required to attain the appropriate amount of compression causes the pruning process to go more slowly than it would otherwise. They either fail to account for topology changes while compressing the models or depend on rules or embeddings that are manually built while neglecting rich topological information.

## 1.1 Problem definition

To the best of our knowledge, most of the existing machine learning/deep learning-based clinical decision support systems suffer from the lack of interpretability in these systems. Furthermore, the highly complex and huge size nature of generated machine learning/ deep learning models make these systems unimplementable on resource-constrained devices and embedded systems. By combining pruning and quantisation into a single process and using explainability as a guide, this study achieves a smaller model size while maintaining competitive performance and preserving important contributing features. This paper develops a methodology that can effectively address the challenge of deploying a method to detect stress on resource-constrained devices while maintaining its effectiveness and achieving explainability. The study develops a novel approach for model compression and optimization, with the goal of significantly reducing the size and complexity of the stress detection algorithm without compromising its accuracy and performance. The ultimate purpose is to enable the implementation of the stress detection method on wearable sensors and other resource-constrained devices, providing individuals with convenient access to real-time stress assessment and management tools. In this study, explainability has been incorporated through the use of the SHAP method as an aid in feature selection and network pruning. It is used to help explain the network produced by the graph convolutional network and to provide insights into the contributing features for stress detection using US-HRV. In this study, the shapely values serve as a reference to ensure that the pruning process is not only reducing the size and complexity of the model but also preserving the most effective contributing features for stress detection.

The salient contributions of this paper are as follows:

1. It introduces an innovative pruning technique that utilizes graph convolutional networks to identify valuable contributing features and devise an effective compression strategy for stress detection using wearable sensors.
2. Our proposed approach integrates both pruning and quantization processes, which leads to a reduced model size while still delivering competitive performance levels.
3. We employ the SHapley Additive exPlanations (SHAP) method to assist in feature selection and network pruning, thereby understanding the subgraphs generated from the GCN for the required operations.
4. Our method significantly decreases sparsity by approximately 60%, with a minimal drop in accuracy (less than 1%), which further illustrates the efficiency of the proposed model.

The remaining parts of the paper are structured in the following manner. In Sect. 2, we discuss the current state of the art that are used in stress diagnosis. Section 3 presents our proposed approach of explainable GCN combining pruning and quantisation for mental stress detection. The experimental setup and analysis of the results are covered in Sect. 4, followed by conclusions in Sect. 5.

## 2 Related works

### 2.1 Machine learning and XAI for healthcare

Explainable Artificial Intelligence (XAI) methods have been developed to make classification decisions of complex machine learning models interpretable. These methods typically follow one of two approaches: functional or message passing. The functional methods focus on localized prediction analysis and include techniques like sensitivity analysis, Taylor series expansion, and model-agnostic approaches like LIME and SHAP. Conversely, message-passing methods generate explanations by running a backward pass through a computational graph, resulting in a prediction as its output. Initial steps towards enhancing the interpretability and explainability of machine learning models used in clinical applications started with the introduction of Dave et al. (2020), Holzinger et al. (2017) and Tjoa and Guan (2021). The main objective is to make these models more transparent and understandable for both ML engineers and medical practitioners. (Wang et al., 2021) developed a model combined with XGBoost, yielding a significant improvement in diagnosing anterior mediastinal masses (MCs and MTs) with a 97.2% accuracy rate. For smaller lesions, the model outperformed radiologists by achieving an accuracy rate of 83.5%. The study underlines the challenges related to the interpretability of complex radiomics models. ElShawi et al. (2021) proposed four measures for evaluating interpretability techniques in machine learning. The study compared six popular techniques, LIME, SHAP, Anchors, LORE, ILIME, and MAPLE, on real-world healthcare data. Results showed variations in performance across metrics and data types, highlighting the need for specifying interpretability focus and understanding the strengths and weaknesses of each technique. Pai et al. (2021) developed a predictive model for identifying ICU patients with bloodstream infections using five machine-learning algorithms on 30 clinical variables. The XGBoost and random forest models performed well, with key predictors being alkaline phosphatase and central venous catheter period. Further validation through clinical trials is recommended. Knapič et al. (2021) examined the potential of XAI methods for decision support in medical image

analysis, focusing on in vivo gastric images from video capsule endoscopy. The study found limitations in evaluating the effectiveness of explanations with non-medical users and suggested further evaluation with domain experts. Alorf (2021) explored the feasibility of CNNs for distinguishing COVID-19 infections from other pulmonary conditions in radiography images. The CNNs showed high sensitivity and specificity, but further training and testing with diverse image sources are needed for practical implementation.

Müller et al. (2022) evaluated the application of XAI techniques in the context of in vitro diagnostic (IVD) devices, introducing the concept of 'causability' as an evaluation of usability in assessing XAI explanation quality. The study underscored the potential value of XAI in glaucoma diagnosis through image analysis. Sarp et al. (2023) developed an XAI model for detecting and interpreting COVID-19 positive Chest X-Ray (CXR) images using transfer learning and data augmentation. Das et al. (2023) addressed interpretability and dimensionality in heart disease classification using XAI and SHAP with four models. XGBoost showed a 2% increase in accuracy over existing methods, marking the first attempt to explain XGBoost's heart disease diagnosis using these techniques. Pattepu et al. (2023) presented a novel paradigm in non-terrestrial networks (NTN) using the XAI approach, optimizing the relationship between signal-to-noise ratio and neighbour nodes, as demonstrated through mathematical formulations and simulations for smart healthcare. Gaube et al. (2023) found that providing explanations with predictions improved physicians' diagnostic accuracy and quality rating, particularly for non-task experts. Future studies could explore the impacts of further complexity and differing explanations. Bienefeld et al. (2023) explored the differing views of developers and clinicians on XAI in healthcare, underscoring the necessity of incorporating both developer and clinician perspectives when designing XAI systems. A summary of the existing research works on machine learning and XAI for healthcare (with their limitations) is presented in Table 1.

## 2.2 Stress and XAI

In the past few years, numerous studies have been conducted with the goal of detecting stress by the measurement of physiological markers. These included circumstances in which the participants were required to deliver a speech in front of an audience, perform mental computations, or endure uncomfortable physiological conditions (Gjoreski et al., 2016; Hovsepian et al., 2015; Picard et al., 2001). The HRV analysis of electrocardiogram data has been used in a significant amount of research for stress analysis. An electrocardiogram (ECG) may be used to assess a person's heart rate variability (HRV), which can then be used to determine how stressed a person is Rameke and Thool (2017), Rigas et al. (2012) and Tanev et al. (2014). Based on HRV, Delaney and Brodie (2000) explored how the heart reacts to short-term psychological stress to find out how it responds. An HRV feature-based transformation strategy was used by Wang et al. (2013) on the Physionet driver database with a K-nearest neighbor (KNN) classifier to identify stress.

Traditional machine learning methods, such as the Random Forest, were used to solve a problem with three classes (no stress, medium stress, and severe stress) and achieved an accuracy of 72% (Gjoreski et al., 2016). Schmidt et al. (2018) were able to train a stress classification model with a precision of 92.28% by using 67 features derived from 7 sensor modalities. Using the same dataset set, Bobade and Vani (2020) employed Deep Neural Networks (DNN) and 40 statistical features to acquire a 95.21% accuracy rate. Aqajari et al. (2020) trained a stress classification model using EDA, obtaining an accuracy of 92% by combining statistical features and a representation acquired by a deep learning model.

**Table 1** Summary of recent studies with ML and XAI in healthcare

S. no.	Paper	Classification methods	Results	Remarks
1	(Wang et al., 2021)	XGBoost combined with a model	Diagnosing anterior mediastinal masses (MCs and MTs) with 97.2% accuracy, surpassing radiologists' 82%. For smaller lesions (<2 cm), accuracy was 83.5% compared to radiologists' 66.7%	Interpretability of complex radiomics models remains challenging
2	(ElShawi et al., 2021)	LIME, SHAP, Anchors, LORE, ILIME, MAPLE	Variations in performance across metrics and data types	Users need to specify their interpretability focus and understand the strengths and weaknesses of each technique
3	(Pai et al., 2021)	XGBoost, Random Forest	XGBoost: AUROC=0.825, Random Forest: AUROC=0.821. Key predictors: Alkaline phosphatase, central venous catheter period	SHAP analysis revealed associations between certain features and bloodstream infections. AI-powered prediction models show potential in assisting clinical decision-making for bloodstream infection diagnostics, but further validation through clinical trials is warranted
4	(Knapič et al., 2021)	LIME, SHAP, CIU	CIU method outperforms LIME and SHAP in providing transparent and understandable explanations for improved human decision-making	Research is limited to non-medical users; further evaluation with domain experts is suggested for a more comprehensive assessment of the explanations' effectiveness
5	(Alorf, 2021)	Convolutional Neural Networks (CNNs)	High sensitivity (100%) and specificity (97.6%) in distinguishing COVID-19 infections from other pulmonary conditions	The CNN classifier should not be a substitute for formal diagnosis by physicians; further training and testing with diverse image sources are needed for practical implementation
6	(Müller et al., 2022)	Explainable Artificial Intelligence (XAI)	Introduction of 'causability' for evaluating the quality of explanations generated by XAI systems	Investigated application in <i>in-vitro</i> diagnostic (IVD) devices to enhance healthcare professionals' decision-making process
7	(Deperitoglu et al., 2022)	Image processing and XAI-augmented Deep Learning	Physicians' active participation in evaluations for human-assisted XAI	Aimed to enhance glaucoma diagnosis and contribute to XAI literature

**Table 1** (continued)

S. no.	Paper	Classification methods	Results	Remarks
8	(Sarp et al., 2023)	XAI model for detecting and interpreting COVID-19 CXR images	Uses LIME to provide interpretable explanations	Requires lung segmentation, which increases processing time and has limitations in interpreting CXR images. More CXR data may improve effectiveness
9	(Das et al., 2023)	XAI model for heart disease classification	Four models using SHAP were employed, with XGBoost showing a 2% increased accuracy	Introduced a reduced dimensional feature subset. The top four diagnostic features were identified, common across explanations. The first attempt to explain XGBoost heart disease diagnosis using these techniques
10	(Pattepu et al., 2023)	X-AI in non-terrestrial networks (NTN), particularly in cooperative relaying	Introduces a novel paradigm in NTN with the X-AI approach	Outperforms existing CPS-based relaying schemes in minimizing Bit-Error-Rate and outage probability in smart healthcare
11	(Gaubé et al., 2023)	Effect of explanations on physicians' diagnostic accuracy	Found that providing explanations improved physicians' diagnostic accuracy and quality rating	Visual annotations can be beneficial, especially for challenging cases or for physicians without radiology expertise
12	(Bienefeld et al., 2023)	Differing views of developers and clinicians on XAI in healthcare	Explores opposing goals, differing sources of truth, and varying perspectives on new versus existing knowledge	Highlights the necessity of incorporating both developer and clinician perspectives when designing the XAI system

Hsieh et al. (2019) were motivated by the success achieved by the XGBoost algorithm in training by using EDA data to derive features in the time, entropy, frequency, and wavelet domains.

Ham et al. (2017) extracted HRV features and used LDA to find and classify the exact stress levels. As a result, there was a high degree of accuracy in classifying people into three different groups: no stress, mild stress, and highly stressed. Zangróniz et al. (2018) introduced a method that can classify mental distress by a tree-based classifier, revealing underlying complementarity that improves the discriminating model's accuracy by 82.35%. Lawanont et al. (2019) proposed a system that uses an IoT architecture to build the stress recognition model, achieving an accuracy of 81.70% using DT. Zubair and Yoon (2020) used different classifiers based on quadratic discriminant analysis (QDA) and Support Vector Machine (SVM) and were able to identify five levels of mental stress with an accuracy of 94.33%. Moridani et al. (2020) showed that HRV features could be used to differentiate between stress and non-stress stages by using the convolutional neural network (CNN) and obtained an average classification rate of 97.9% for cognitive stress and 94.5% for emotional stress. Pourmohammadi and Maleki (2020) used an innovative combination of feature selection with SVM that yielded an accuracy of stress identification of 100%, 97.6%, and 96.2% across levels of two, three, and four respectively. Rodríguez-Arce et al. (2020) used KNN to measure the student's level of anxiety with the State-Trait Anxiety Inventory (STAI) and found that the physiological feature subset can best explain the difference between stress and anxiety states. Sánchez-Reolid et al. (2020) made 147 participants watch a series of video clips depicting tense and relaxed situations. These clips were intended to evoke certain emotions. It achieved an F1-score of 83% with SVM and 92% with D-SVM. Zalabarria et al. (2020) classified stressed and relaxed states by employing a 20-s sliding window protocol to the Fuzzy algorithm, which yielded an F1 score of 91.15% and 96.61% for stressed and relaxed, respectively. Zainudin et al. (2021) employed an IoT sensor to gather data regarding a real-life mental health scenario and obtained the best classification accuracy of 96% with DT. Deep learning is utilised in order to address an ECG-based stress detection issue (Seo et al., 2019). A driver stress detection network was proposed by Rastgoo et al. (2019), which utilized a multi-modal fusion of CNN and LSTM, achieving an accuracy of 92.8%. Uddin et al. (2022) employed ANN to predict depressive symptoms in a large textual sample based on people's online behaviour, yielding an accuracy of 95%. Table 2 summarises the recent studies in stress classification by various ML/DL algorithms with their limitations. Most of the said models developed for stress detection have shown good performance in controlled settings; however, they may not perform well in real-world scenarios (Ham et al., 2017; Lawanont et al., 2019; Rodríguez-Arce et al., 2020). They also lack interpretability, making it difficult to understand the reasoning behind the model's predictions and identify any errors in the model.

### 2.3 Compression (pruning and quantization) of deep neural networks

Wearable technologies are situated at the outermost boundaries of a network and frequently engage in direct interactions with users or the physical environment. In order for artificial intelligence models to function in real-time on such devices, it is imperative to optimise them for reduced latency, minimal power consumption, and constrained storage capacities. In general, machine learning models, specifically deep neural networks, exhibit a parameter count ranging from millions to billions. The considerable level of



**Table 2** Summary of recent studies on stress classification

S. no.	Paper	Classification methods	Results		Remarks
			Value	Criteria	
1	(Ham et al., 2017)	LDA	82%	Accuracy	<ul style="list-style-type: none"> <li>Effectively classified stressed and non-stressed</li> <li>Increased size in model and computational resources</li> </ul>
2	(Zangróniz et al., 2018)	Decision tree	82.35%	Accuracy	<ul style="list-style-type: none"> <li>Provide a non-invasive and easily accessible method for measuring mental distress</li> </ul>
3	(Lawanont et al., 2019)	KNN, SVM, decision tree	81.7% (DT)	Accuracy	<ul style="list-style-type: none"> <li>Able to detect stress via activity tracker in a smartphone</li> <li>The values obtained can only be used as a warning but not clinically</li> </ul>
4	(Rastgoo et al., 2019)	CNN-LSTM	92.76%	Accuracy	<ul style="list-style-type: none"> <li>Uses multiple data sources, such as physiological signals and driving behaviour</li> <li>Shows multi-modal data can provide higher accuracy</li> </ul>
5	(Zubair & Yoon, 2020)	QDA, SVM	94.33% (SVM)	F1 score	<ul style="list-style-type: none"> <li>Uses time and frequency domains to analyse pulse rate variability</li> <li>Increased complexity of the model</li> </ul>
6	(Moridani et al., 2020)	CNN	97.9%	F1 score	<ul style="list-style-type: none"> <li>Analyses HRV signals under different stress conditions</li> <li>Lacks accuracy compared to the state-of-the-art</li> </ul>
7	(Pourmohammadi & Maleki, 2020)	SVM	97.6% (3 levels)	Accuracy	<ul style="list-style-type: none"> <li>A comprehensive study on stress detection using both electrocardiogram (ECG) and electromyogram (EMG) signals</li> <li>Provides high accuracy in 3 classes</li> </ul>
8	(Rodríguez-Arce et al., 2020)	KNN, SVM	95% (SVM)	Accuracy	<ul style="list-style-type: none"> <li>Proposed a system that uses physiological signals such as heart rate and electrodermal activity to recognise anxiety and stress levels in academic environments</li> </ul>
9	(Sánchez-Reolid et al., 2020)	SVM, D-SVM	92.35%	F1-score	<ul style="list-style-type: none"> <li>Presents a study on using DSVM for stress identification through EDM</li> <li>DSVM outperforms SVM in terms of accuracy</li> </ul>
10	(Zalabarría et al., 2020)	Fuzzy Algorithm	91% (stressed), 96% (non-stressed)	F1-score	<ul style="list-style-type: none"> <li>Presents a fuzzy algorithm that is proven to be effective in estimating stress and relaxation levels in real-time</li> <li>Found to be affordable and easy to use, increasing its potential impact</li> </ul>
11	(Uddin et al., 2022)	ANN	95% (Stress)	Accuracy	<ul style="list-style-type: none"> <li>Uses ANN to identify texts describing self-perceived symptoms of depression</li> </ul>

intricacy frequently contributes to enhanced precision, yet it also leads to increased model sizes and prolonged inference durations. The computational cost and energy consumption associated with large models on edge devices can pose significant barriers. The main goal of pruning is to reduce the resource demands of the model while maintaining its performance at a satisfactory level, resulting in a lighter, faster, and more efficient model.

Pruning techniques have the potential to alleviate these issues. Pruning is a technique that entails the elimination of less significant parameters or neurons from a model, resulting in a reduction of its overall complexity. Various techniques can be employed to enhance the performance of neural networks. These techniques range from basic weight pruning, which involves removing neurons with weights below a specific threshold, to more advanced approaches, such as utilising L1 or L2 regularisation. The latter method promotes sparsity in the model's weights during the training process. The advantage of pruning is its ability to substantially reduce the size of the model, thereby enabling its compatibility with devices that possess restricted storage capacity. The decrease in the size of the model can also lead to a reduction in the inference time, thereby facilitating faster real-time responses. This aspect holds significant importance for numerous applications on edge devices. For example, the utilisation of a pruned model can facilitate expedited object recognition on a smartphone camera or enhance the efficiency of anomaly detection in an Internet of Things (IoT) sensor network.

Additionally, the process of pruning has the potential to decrease the energy demands of the model. Energy efficiency is of utmost importance, especially in the context of battery-operated devices. A pruned model necessitates a reduced number of computations, thereby resulting in decreased energy consumption.

There have been several proposed techniques for compressing and speeding up neural networks. Tensor factorization is a computational method that decomposes the weights of a neural network into smaller, more manageable components. The decomposition of a  $3 \times 3$  convolutional filter into one  $1 \times 3$  and one  $3 \times 1$  filters has been demonstrated by Jaderberg et al. (2014). Previous studies have employed truncated singular value decomposition (SVD) as a means to accelerate fully connected layers (Denton et al., 2014; Girshick, 2015; Xue et al., 2013). Quantisation (Rastegari et al., 2016) offers an alternative strategy for mitigating computational complexity. This method involves the representation of floating-point values using a reduced number of bits, thereby conserving resources while maintaining an acceptable level of precision. Zhang et al. (2018) proposed a compact network design, entails making alterations to the convolutional structure.

Pruning techniques, on the other hand, primarily centre around the reduction of network complexity through the elimination of connections. Han et al. (2015) propose an iterative strategy for constructing a sparse network by eliminating connections that possess weights below a predetermined threshold. However, it frequently faces practical performance challenges related to cache and memory access. In order to tackle this issue, several studies (Fernandes & Yen, 2021; He, 2022; Hu et al., 2016; Liang et al., 2021) have suggested the reduction of redundant connections at the filter level.

Dong et al. (2022) used pruning to compress a deep neural network (DNN) model and found that pruning the DNN model resulted in a significant reduction in model size without a significant loss in performance. Abbasi-Asl and Yu (2021) used pruning to compress a convolutional neural network, and the model resulted in a significant reduction in model size and improved classification accuracy. Recently, many pruning strategies (Blalock et al., 2020; He et al., 2017; Pasandi et al., 2020) for automatically compressing DNNs have been presented. However, they either neglect rich topological information by relying

on rules or embeddings that are manually created and ignoring it, or they do not take topology changes into account when they compress the models. A pruning strategy designed for one DNN cannot be transferred to another DNN, which is why every network requires a strategy that is specifically tailored to it. Table 3 summarizes various pruning methods used in deep neural networks.

Several studies have used machine learning and deep learning algorithms to classify stress levels based on HRV features, with high accuracy achieved in controlled settings. However, the lack of interpretability and the performance in real-world scenarios remain a limitation. Recent research has focused on compressing deep neural networks through pruning strategies, resulting in a significant reduction in model size without a significant loss in performance. However, current pruning strategies have limitations and require tailoring to specific networks. A notable point of divergence in our study lies in the capacity to autonomously ascertain the network architecture, specifically the optimal number of preserved channels at each layer.

This paper addresses the gap in the field of stress detection by proposing a novel pruning method based on graph neural networks. The method combines pruning and quantisation into a single process to achieve a smaller model size while maintaining competitive performance. This is achieved by using the SHAP method to aid in feature selection and network pruning, resulting in reduced sparsity by a factor of 60% with minimal loss in accuracy. The proposed method helps to better identify effective contributing features and establish an effective compression strategy, thus making it an innovative approach in the field of stress detection using US-HRV data.

### 3 Methodology

The methodology used in the study is outlined in Fig. 1 After preprocessing the data, SHAP was used to identify the major contributing features in the dataset by submitting it to a generic classification using Graph Convolutional Network. Once the key features were identified and ranked, a 2-stage model compression method was applied, which included model pruning followed by weight quantisation. The filtered model with fewer parameters and lower computational complexity was then used for classification, which divided the data into two categories: those with stress ideations and those without stress ideations.

#### 3.1 Feature identification using SHAP

SHAP is a method for interpreting the output of machine learning models (Lundberg et al., 2020). It is based on the concept of Shapley values from cooperative game theory, which provides a way to fairly distribute a value among a group of individuals based on their contributions. In the context of machine learning, SHAP values can be used to explain the contribution of each feature to the model's output. SHAP values can provide a way to understand which features are most important in a model's predictions and how different features interact with each other to affect the prediction.

In our work, the filtered ECG signal was used to calculate several statistical measures, including Mean RR and standard deviation. The absolute values of these characteristics were calculated using the SHAP feature contribution values. Absolute power was also used to determine the peak frequency along each axis. SHAP method is used to understand the

importance and contribution of each feature to the output of the model, in this case, the statistical measures of the filtered ECG signal like Mean RR and standard deviation. The absolute values are used to get the magnitude of the contribution of each feature rather than the direction.

### 3.2 Classification using graph convolutional network

A graph convolutional network, or GCN, is a type of neural network that works with graphs. A GCN takes a graph  $G = (V, E)$  as input where the graph is denoted by the equation  $G = (V, E)$ , in which  $V$  represents the set of nodes and  $|V| = n$ , and  $E$  represents the set of edges. In addition to the adjacency matrix  $A$ , which is responsible for representing the structure of the graph, another matrix called  $X$  is provided as input. This matrix is used for storing the feature descriptions of the nodes; more specifically, each node  $v_i$  is provided with a vector called  $\mathbf{x}_i \in \mathbb{R}^f$ , where  $f$  is the number of features that are given as input.

Each layer is made up of an  $N$  feature matrix, where each row is a node feature. Using the propagation rule  $f$ , these features are added together at each layer to make the features of the next layer. GCNs can be trained from end to end, which means they can be trained in a supervised or unsupervised way, depending on the task to be done. They are also built to determine the new embedding state by utilising the structure of the graph in addition to the characteristics of the nodes and edges, and they achieve this by using a method that is iterative and aggregates the quality of the neighbourhoods that are next to one another. When all the information is combined, the final embedding state can be used to gain information. We use a GCN for classification as it is best suited for studies relating to signal-processing tasks.

### 3.3 Network pruning

Graph convolutional networks are a type of deep learning model that is designed to operate on graph-structured data. As the size of the graphs used in these applications increases, so does the size of the GCN model required to process them. This can lead to issues with computational complexity, memory usage, and model interpretability. One approach to addressing these issues is to use model compression techniques such as pruning. Pruning is a technique used to remove the redundant and insignificant parameters of a neural network model to make it more efficient and smaller in size. It helps to improve the computational efficiency of the model without significantly affecting its accuracy. The bulk of the training parameters, including weights and biases, are stored in the convolutional layers of the graph convolutional network. These layers are responsible for the learning process. The complexity of the computation is doubled when weights are multiplied, but the addition of bias to each neuron only adds one more to the total complexity of the problem. Pruning consists of three steps: training a big model, reducing weights, and fine-tuning the weights (see Fig. 2).

In the context of GCNs, pruning typically involves removing connections between nodes in the graph, as well as removing entire nodes and their associated weights. One common approach is to use magnitude-based pruning, which involves setting a threshold value for the weights and removing those that fall below it. This can be done iteratively, with the model being retrained after each round of pruning until the desired level of compression is

**Table 3** Summary of various neural network compression methods

S. no.	Technique	Paper reference	Description
1	Tensor Factorization	Jaderberg et al. (2014)	Decomposes a $3 \times 3$ convolutional filter into one $1 \times 3$ and one $3 \times 1$ filters to reduce size and computational complexity
2	Truncated SVD	Xue et al. (2013), Denton et al. (2014), Girshick (2015)	Employed to accelerate fully connected layers by reducing the number of parameters
3	Quantisation	Rastegari et al. (2016)	Represents floating-point values using fewer bits, conserving resources while maintaining an acceptable level of precision
4	Compact Network Design	Zhang et al. (2018)	Makes alterations to the convolutional structure to achieve a more compact network design
5	Group-wise Convolution	Zhang et al. (2018)	Enhances the speed of the conventional convolution process by using group-wise convolution
6	Pruning	Han et al. (2015)	Reduces network complexity by eliminating connections with weights below a predetermined threshold
7	Filter-level Pruning	Fernandes and Yen (2021), He (2022), Hu et al. (2016), Liang et al. (2021)	Reduces redundant connections at the filter level to address cache and memory access challenges

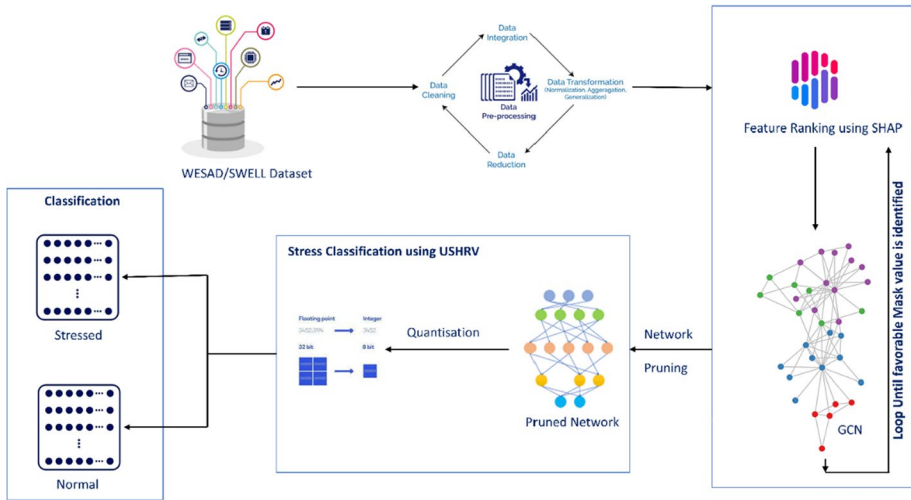


Fig. 1 Proposed methodology

achieved. Pruning has been shown to be effective in reducing the size and computational complexity of GCN models while maintaining or even improving their performance.

Let

$$\begin{aligned}
 f &\sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot)) \\
 \mu(\theta) &= \mathbb{E}[f(\theta)] \\
 k(\theta, \theta') &= \mathbb{E}[(f(\theta) - \mu(\theta))(f(\theta') - \mu(\theta'))].
 \end{aligned}$$

where

- $f$  represents a function, and  $\sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot))$  indicates that the function  $f$  is modelled as a Gaussian process with a mean function  $\mu(\cdot)$  and a covariance function  $k(\cdot, \cdot)$ .
- $\mu(\theta)$  represents the mean function of the Gaussian process at input  $\theta$ .
- $\mathbb{E}[f(\theta)]$  denotes the expected value of the function  $f$  at input  $\theta$  according to the Gaussian process.

Given  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  and function evaluations  $f(\Theta) = \{f(\theta_1), f(\theta_2), \dots, f(\theta_n)\}$ , the posterior belief of  $f$  at a novel candidate  $\hat{\theta}$  is given by

$$\begin{aligned}
 \tilde{f}(\hat{\theta}) &\sim \mathcal{N}(\tilde{\mu}_f(\hat{\theta}), \tilde{\Sigma}_f^2(\hat{\theta})) \\
 \tilde{\mu}_f(\hat{\theta}) &= \mu(\hat{\theta}) + k(\hat{\theta}, \Theta)k(\Theta, \Theta)^{-1}(f(\Theta) - \mu(\Theta)) \\
 \tilde{\Sigma}_f^2(\hat{\theta}) &= k(\hat{\theta}, \hat{\theta}) - k(\hat{\theta}, \Theta)k(\Theta, \Theta)^{-1}k(\Theta, \hat{\theta}).
 \end{aligned}$$

where

- $\tilde{f}(\hat{\theta})$  is the function value at the novel candidate input  $\hat{\theta}$ .
- $\mu(\hat{\theta})$  is the mean of the Gaussian process at the novel candidate input  $\hat{\theta}$ .

- $\tilde{\mu}_f(\hat{\theta})$  is the mean of the Gaussian process at the novel candidate input  $\hat{\theta}$  after considering the known function evaluations.
- $\tilde{\Sigma}_f^2(\hat{\theta})$  represents the variance of the Gaussian process at the novel candidate input  $\hat{\theta}$ .
- $k(\hat{\theta}, \Theta)$  and  $k(\Theta, \hat{\theta})$  are covariance vectors representing the covariances between the novel candidate input  $\hat{\theta}$  and the existing inputs in the set  $\Theta$ .

Let  $\theta^+$  be the best candidate evaluated so far. The expected improvement (EI) of a candidate ( $\hat{\theta}$ ) is defined as the expected increase in the function value compared to the best candidate evaluated so far ( $\theta^+$ ). The EI can be computed efficiently in closed form and is used as a criterion to choose the next candidate for evaluation.

$$EI(\hat{\theta}) = \mathbb{E}[m\{0, f(\theta^+) - \tilde{f}(\hat{\theta})\}]$$

$$EI(\hat{\theta}) = \tilde{\Sigma}_f(\hat{\theta})(Z\Phi(Z) + \phi(Z))$$

$$Z = \frac{\tilde{\mu}_f(\hat{\theta}) - f(\theta^+)}{\tilde{\Sigma}_f(\hat{\theta})}$$

where

- $EI(\hat{\theta})$  denotes the Expected Improvement at the novel candidate input  $\hat{\theta}$ . It represents the expected increase in the function value compared to the best candidate evaluated so far.
- $Z, \Phi(Z)$ , and  $\phi(Z)$  are standard normal random variables, in which  $Z$  is a random variable,  $\Phi(Z)$  is the standard normal cumulative distribution function, and  $\phi(Z)$  is the standard normal probability density function.
- $\theta^+$  represents the best candidate evaluated so far.

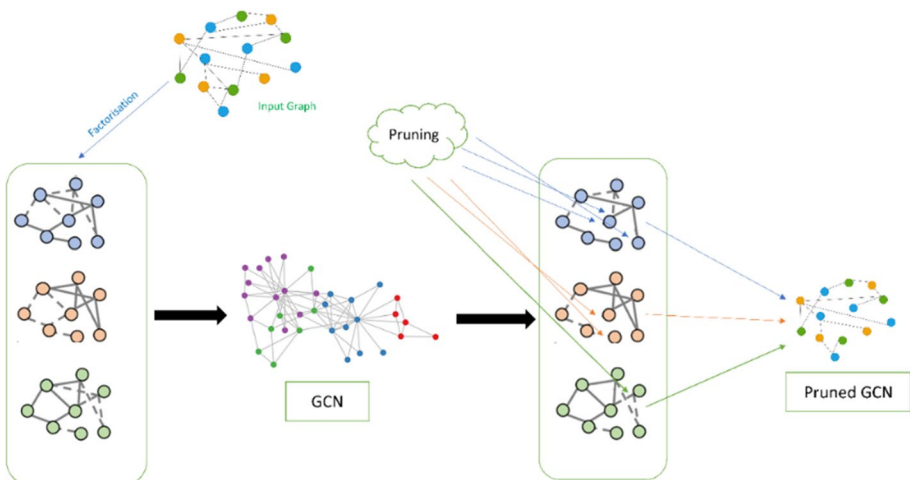


Fig. 2 GCN pruning

We prune the graph convolutional network by trimming the weights depending on their size, which is often expressed as a percentage ranging from 0 to 100%. To do this, we rank the weights for each of the three layers independently and set the bottom % weights to zero. This effectively serves the connections between those neurons. It takes into account the mean and variance of the posterior belief and assigns a higher score to candidates that are expected to result in significant improvements.

### 3.4 Reducing and modifying weights using quantisation

Quantisation is the process of reducing the number of levels or values that a signal or data can take on so as to reduce the memory and computational requirements of a model by reducing the precision of the model's parameters. In the process of quantisation, the GCN's parameters are grouped into a finite number of intervals or bins, and each parameter is replaced by the value that corresponds to the bin it falls in. By doing this, the parameter's feasible values become limited, leading to a decrease in the memory and computational demands of the model.

The tuning process starts by initialising the parameters of the baseline model to zero. The first training is used to determine which connections between neurons should be severed before further training can begin. We use an approach that includes zeroing and fine-tuning each convolutional layer in sequence rather than pruning and fine-tuning each layer of the network separately.

## 4 Results and discussions

### 4.1 Datasets description

For this study, we have used WESAD and SWELL-KW datasets that are publicly available and offer a wealth of physiological and motion data collected from wearable sensors, providing valuable information for the study of stress and affective states in individuals.

The WESAD (Wearable Stress and Affect Detection) (Schmidt et al., 2018) dataset is a publicly available dataset that contains physiological and motion data collected from wearable sensors worn by participants while they engage in a variety of activities, including baseline measurements, stress-induction tasks, and affective computing tasks. The dataset includes data from 15 participants, including 7 females and 8 males, who were between the ages of 22 and 35. The data was collected using a variety of wearable sensors, including a chest-strap heart rate monitor, a wrist-worn accelerometer, and a wrist-worn electrodermal activity sensor. The dataset includes both raw sensor data and preprocessed data, as well as labels for stress and affective states. The size of the WESAD dataset is about 3.8 GB, with approximately 10 h of data collected from each participant.

The SWELL-KW (Saskia et al., 2014) dataset is a multi-modal dataset obtained from an experiment in which 25 people conducted knowledge work tasks under various stressors (email interruptions and time pressure). The dataset contains computer logging, facial expression, body posture, heart rate variability, skin conductance, and validated questionnaire responses from subjects. The size of the SWELL dataset is about 5.4 GB, with approximately 3 h of data collected from each participant.



## 4.2 Data preprocessing

Preprocessing is necessary before conducting a US-HRV study to eliminate outliers in the RR intervals caused by noise, such as movement. Data that is more than three standard deviations (SD) away from the mean is considered an outlier in RR interval data. A non-linear interpolation method, called cubic spline interpolation, was chosen to process the HRV signal. The sensor data was then further analysed using a sliding window with a shift of 0.25 s. To calculate ECG features, a five-second window was used, which is a standard practice in acceleration-based context recognition. All physiological characteristics were calculated using a 60-s frame, except for the statistical and frequency domain ECG features. The size of this window was chosen based on the suggestions of Kreibitz (2010). We followed a fivefold cross-validation on the data with a train test split ratio of 80:20.

The raw ECG data was processed by first applying a high-pass filter to eliminate the DC component. The filtered signal was then divided into periods of 5 s, and statistical significance and peak frequency were calculated. Power spectral density was determined using seven frequency bands evenly spaced from 0 to 350 Hz. After this second round of processing, the raw ECG data was further processed by applying a low-pass filter. The processed signal was then divided into periods of 60 s, on which various peak characteristics, such as the overall number of EMG peaks and their mean amplitude, were calculated. Peak detection was done using the variations that showed a significant contribution to the ECG. These peaks were used to calculate the average heart rate (HR) and HR variability.

The following seven US-HRV measures reported in Table 4 were analysed for this study.

A 60-s measurement revealed significant variations across groups using Mean RR and LF characteristics. When comparing long-term changes, the high-stress group did not show smaller variances in any of the Mean\_RR or LF measures. Analysis of 2-min HRV samples showed a significant decrease in Mean\_RR, a significant increase in LF, and a stable LF/HF ratio. Non-linear assessments of heart rate variability indicated a reduction in acute mental stress.

## 4.3 Classification analysis

Table 5 presents the results of various classifier algorithms applied to the stress condition classification task. The classifiers include K-Nearest Neighbour (Wang et al., 2013), Multi-Layer Perceptron (Zainudin et al., 2021), Support Vector Machine (Zubair & Yoon, 2020), Convolutional Neural Network (Moridani et al., 2020), Deep Neural Network (Zainudin et al., 2021), Artificial Neural Network (Uddin et al., 2022), and Linear Discriminant Analysis (Ham et al., 2017). The metrics used to evaluate the performance of the classifiers are recall, precision, accuracy, and F1-score.

Based on the results shown in Table 5, for the WESAD dataset, the proposed method has a Recall of 97.2%, which is higher than the other classifiers, indicating that it is able to identify a high percentage of the relevant samples correctly. Furthermore, the proposed method has a Precision of 98.42%, which is also higher than the other classifiers, indicating that a high percentage of the samples that it identifies as relevant are actually relevant. This results in a high Accuracy of 98.84%, which is also higher than the other classifiers. Furthermore, the F1-Score of 96.48% is also the highest among all the classifiers, which indicates a good balance of precision and recall.

The proposed method has performed well on the SWELL dataset, with high scores across various performance metrics. The Recall score of 95.32% indicates that the proposed

model correctly identifies a significant proportion of relevant samples, while the Precision score of 94.22% shows that the proposed model accurately identifies relevant samples from those it flags. The model also achieves a high Accuracy score of 95.74%, indicating it can classify most samples with high accuracy. The high F1-Score of 96.37% is also noteworthy, suggesting the method has achieved a good balance between precision and recall.

#### 4.4 Feature identification and explainability using SHAP

The probabilistic value associated with each contributing factor in both WESAD and SWELL datasets are identified using SHAP, and unimportant weights are removed. In this scenario, we loop through the named modules in the model and check if the module is a convolution layer. If it is a convolution layer, then we extract the weights, create a binary mask indicating which weights are not zero, and multiply the weights by the mask to remove the unimportant weights. Thus, we create an instance of the GCN model with input features of size 10, hidden features of size 5, and output features of size 2 and use the function to prune the weights in the model.

As we can see from Fig. 3, the major contributors to the prediction come from the values of Median RR and Mean RR. Thus, we evaluate the model taking the Mean RR value as a reference point as its contribution is strictly in the range of desired value (0.2–0.3 for WESAD and 0.09–0.3 for SWELL), and with reference to it, we find the LF/HF ratio. The findings of this study, using Mean RR as a reference point, show that levels of LF and the LF/HF ratio significantly increase during stressful situations. This study correlates with the existing knowledge that the increased activity of the autonomic nervous system is strongly linked to changes in HRV values recorded during stress (Evans et al., 2013; Pham et al., 2021). This is evidenced by the marked increase in LF and the significant decrease in Mean RR. Non-linear HRV measures, such as sample entropy or fractal dimension, are commonly used to quantify HRV. A reduction in these non-linear HRV measures can indicate an increase in stability and regularity in heart rate variability patterns. This phenomenon is often observed during stressful conditions when the body shifts towards more stable and periodic HRV behaviour. This shift is related to the deactivation of control loops in the cardiovascular system, which regulate the heart rate. Thus, stress can have a significant impact on the autonomic regulation of the heart, as evidenced by changes in HRV measures.

**Table 4** US-HRV measures and values

HRV Measures	Units	Description
MEAN_RR	(ms)	RR interval's Mean
STD_NN	(ms)	RR interval's Standard deviation
RMSSD	(ms)	The square root of the mean squared difference between two RR intervals
PNN50	(%)	NN50 by the total number of RR
Absolute power	(ms <sup>2</sup> )	Absolute powers of LF and HF bands
LF/HF	–	LF/HF band Ratio
SD1, SD2	(ms)	The standard deviation

**Table 5** Classification before applying the pruning methods

Classifier	WESAD (Schmidt et al., 2018)				SWELL (Saskia et al., 2014)			
	Recall	Precision	Accuracy in %	F1-score	Recall	Precision	Accuracy in %	F1-score
KNN (Wang et al., 2013)	87.0	88.83	88.08	87.93	83.0	84.78	84.46	83.22
MLP (Zainudin et al., 2021)	86.0	84.42	87.28	85.23	88.0	87.25	88.64	87.43
SVM (Zubair & Yoon, 2020)	94.0	96.96	93.38	94.33	91.0	90.25	92.33	92.55
CNN (Moridani et al., 2020)	96.0	96.97	95.04	94.97	94.0	94.88	94.85	94.05
DNN (Zainudin et al., 2021)	96.0	95.0	95.0	95.0	95.0	94.55	96.32	94.33
ANN (Uddin et al., 2022)	85.36	88.34	87.48	86.59	90.0	89.33	88.81	89.56
LDA (Ham et al., 2017)	82.36	83.67	82.88	82.54	84.0	85.24	85.64	89.21
GCN (proposed)	97.2	98.42	98.84	96.48	95.32	94.22	95.74	96.37

### 4.5 Pruning and quantisation analysis

In general, pruning is the process of reducing the size of the weights by a certain percentage, which can range from 0 to 1. This is achieved by setting the lowest ranking weights in each of the three layers to zero, effectively cutting off communication between those neurons. We present an improved version of traditional pruning methods, where we use 0.5 as a threshold to prune weights and then retrain the network to recover the lost accuracy as the first step. Then, we iteratively prune and retrain the network until a sparsity of 60% is reached.

To optimise the model’s performance, we use quantisation, which involves zeroing and fine-tuning each convolutional layer sequentially instead of pruning and fine-tuning them individually. The process starts by ranking the weights in the first layer and setting (and freezing) the required percentage of them to zero. Then, we move on to the next convolutional layer and repeat the process of zeroing off the necessary proportion of its parameters before fine-tuning for all convolutional layers.

We evaluate the performance of the model by measuring the overall accuracy, F1 score, loss, and sensitivity for different levels of sparsity for Convolutional Neural Networks with pruning (Moridani et al., 2020), Deep Neural Networks with pruning (Zainudin et al., 2021), Artificial Neural Networks with pruning (Uddin et al., 2022) and our proposed GCN with pruning and quantisation. The results are presented in Table 6. The proposed GCN with pruning and quantisation performed the best, with a recall of 96.2% and an accuracy of 97.75% on the WESAD dataset, while having a recall of 92.15% and an accuracy of 94.48% on the SWELL dataset.

The efficacy of our model can be attributed to a multitude of sophisticated technical implementations. Firstly, the model has the capability to independently determine the most suitable network architecture, specifically the optimal number of preserved channels at each layer, in order to enhance efficiency and achieve task-specific performance. Additionally, the model incorporates advanced pruning techniques to remove redundant connections at the filter level, resulting in decreased computational complexity and enhanced efficiency and speed. Furthermore, the methodology incorporates tensor factorization techniques to

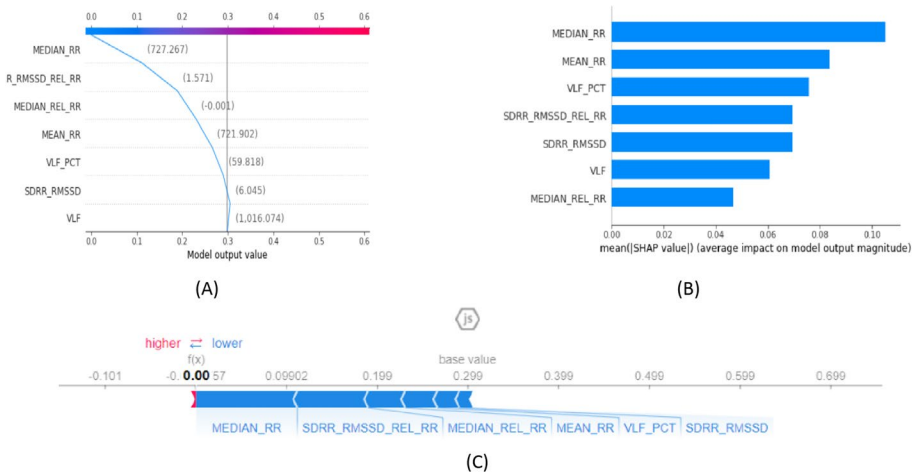


Fig. 3 Analysis of feature contribution via Shapley values: (A), (B) on WESAD, (C) on SWELL

decompose the weights of the network into smaller and more manageable components. Additionally, it applies quantization methods to conserve computational resources while minimising the loss of precision. The proposed model also introduces a modified convolutional structure that utilises group-wise convolution to enhance the processing of high-dimensional data, leading to improved efficiency and accuracy in the outcomes.

The overall accuracy of the model is affected by the sparsity level, which is the percentage of weights that are set to zero. The results show that pruning with quantisation is the most effective method for maintaining high accuracy up to a sparsity level of around 60% to 70%. The fine-tuning stage helps to improve sensitivity and maintain high accuracy even at higher sparsity levels.

## 4.6 Results analysis and discussion

The evolution of wearable technology in recent years has heralded a transformative era in health monitoring, with stress detection—a pervasive health parameter affecting numerous individuals globally—at the forefront of this innovation. This study's primary objective was to effectively leverage physiological signals derived from these wearable devices to accurately identify and quantify stress levels. In a novel approach, the study developed a machine learning model utilising the dynamic capabilities of GCN, concurrently integrating pruning and quantisation methodologies to elevate computational efficiency—a critical element when applied to resource-limited wearable devices.

Our study harnessed data from the well-regarded WESAD and SWELL datasets, renowned as comprehensive repositories in stress detection research, hosting a diverse spectrum of physiological signals. These rich datasets enabled the cultivation of a holistic understanding of varied bodily responses elicited during stress episodes. The subsequent training and evaluation of our model on these datasets generated inspiring results.

The GCN model's performance was not just encouraging but strikingly effective, achieving precision rates of 97.75% on the WESAD dataset and 94.48% on the SWELL dataset. This potent performance, corroborated by an accuracy range of approximately 95% to 98%, underscores the model's robust predictive capabilities.

Our model's performance metrics extended beyond precision and accuracy, evidencing robust levels of recall, too. Precision, measuring the exactitude of positive predictions, was recorded as 94.42% on the WESAD dataset and 93.45% on the SWELL dataset. Notably, recall or sensitivity, gauging the model's capacity to identify true positives accurately, exhibited impressive results of 96.2% and 92.15% on the WESAD and SWELL datasets, respectively. These results testify to the model's proficiency in accurately detecting stress instances while minimising false negatives and positives effectively.

Further performance evaluation involved a meticulous examination of the Receiver Operating Characteristic (ROC) and Precision-Recall curves. The Area Under the Receiver Operating Characteristic curve (AUC-ROC) provided a robust performance metric, considering sensitivity and specificity. Prior to the application of pruning and quantisation, the model achieved AUC-ROC scores of 0.996 on the WESAD dataset and 0.992 on the SWELL dataset. Post-application, the scores were marginally affected, registering at 0.994 and 0.986, thereby suggesting a minimal impact on performance. Similarly, the area under the precision-recall curve (AUC-PR), particularly valuable in the context of imbalanced datasets, demonstrated comparable trends.

Positioned alongside previous studies, our GCN model exhibited a commendable, potentially superior performance, excelling in terms of accuracy, precision, recall, and

**Table 6** Classification values after applying pruning and quantisation

Classifier	WESAD (Schmidt et al., 2018)					SWELL (Saskia et al., 2014)				
	Recall	Precision	Accuracy in %	F1-Score	Sparsity reduction in %	Recall	Precision	Accuracy in %	F1-Score	Sparsity reduction in %
DNN with pruning (Zainudin et al., 2021)	94.25	94.23	93.22	93.58	65.25	93.0	93.78	94.35	93.54	71.26
ANN with pruning (Uddin et al., 2022)	80.22	79.57	80.12	81.48	70.48	91.0	90.88	91.27	90.28	66.58
CNN with pruning (Moridani et al., 2020)	94.0	92.43	92.04	93.82	68.95	92.0	92.44	92.21	92.96	68.46
GCN with pruning (proposed)	96.2	94.42	97.75	97.66	63.4	92.15	93.45	94.48	94.39	69.34

F1-score. However, the distinguishing merit of our model lies in its enhanced computational efficiency, achieved via the integrated pruning and quantisation techniques. This approach successfully reduced the model size by a substantial average of 60%, simultaneously improving processing time by 45%. Consequently, this innovative methodology provides a more feasible and efficient solution for implementation in wearable devices, heralding a new paradigm in stress detection technology.

## 4.7 Inferences

In addition to accuracy and sensitivity, power consumption is also an important consideration when implementing machine learning and deep learning models in real-world applications. The complexity of the model, measured in terms of floating-point operations per second (FLOPs), directly affects power consumption. The results show that as the sparsity level increases, the model's complexity decreases, resulting in significant power savings. For example, at a sparsity level of 63.4%, the base model's complexity drops from 1.46 million FLOPs to 0.57 million FLOPs for the WESAD dataset and sparsity level of 63.4%, the base model's complexity drops from 1.56 million FLOPs to 0.67 million FLOPs for SWELL dataset, resulting in a 60% to 70% decrease in complexity and a corresponding reduction in power consumption. The F1 score for the pruning approach at this sparsity level is 97.66%, and the accuracy is 97.75% for WESAD, and the F1 score for the pruning approach at this sparsity level is 94.39%, and the accuracy is 94.48% for SWELL. Overall, these results show that by carefully adjusting the sparsity level, it is possible to achieve a good trade-off between accuracy, sensitivity, and power consumption.

## 5 Conclusion

This study presents a novel iterative pruning with a quantisation approach for identifying mental stress in real-time using Ultra-short Heart Rate Variability measurements. The proposed approach uses a graph convolutional network model to classify US-HRV measurements with a high degree of accuracy and efficiency. As the GCN model's complexity can be a challenge for real-time applications, especially when deployed on resource-constrained devices, this study proposed a multi-stage pruning technique for GCN models that reduces their complexity while maintaining virtually all of their performance. The results show that the proposed method can classify US-HRV with a high degree of accuracy and efficiency, and the runtime complexity is decreased by ~60% compared to the initial model.

Notwithstanding the encouraging outcomes, it is imperative to recognise certain constraints within our research. The study primarily utilised a limited number of datasets for both the training and evaluation of the models. This statement may not comprehensively encompass the entirety of physiological reactions to stress, as they can be influenced by a multitude of individual and contextual factors. In addition, the extensive range of wearable devices, each possessing distinct specifications and measurement intricacies, has the potential to impact the model's generalizability. Subsequent investigations should contemplate the inclusion of a more extensive array of datasets and the evaluation of the model's performance on various categories of wearable devices in order to augment its resilience and versatility.

Future research in the area of stress detection using ultra-short HRV and wearable sensors could involve training machine learning and deep learning algorithms on bigger and more diversified data sets to increase their generalizability, ultimately leading to more accurate and effective stress detection tools for individuals to manage their mental health and overall well-being.

**Acknowledgements** We extend our gratitude to Mr. Anas Abdul Kadher, a practising clinical psychologist, for his insightful suggestions and valuable feedback on this paper.

**Author contributions** VA: Conceptualization, Methodology, Implementation, Writing, Revision. GRG: Conceptualization, Revision, Supervision.

**Funding** Not applicable.

**Data availability** Data is available publicly in open-source datasets.

**Code availability** The code is available at <https://github.com/adarshv96/GCN-WESAD-SWELL>.

## Declarations

**Conflict of interest** Authors declare that they have no conflict of interest.

**Ethical approval, consent to participate, and consent for publication** Not applicable.

## References

- Abbasi-Asl, R., & Yu, B. (2021). Structural compression of convolutional neural networks with applications in interpretability. *Frontiers in Big Data*, 4(August), 1–13. <https://doi.org/10.3389/fdata.2021.704182>
- Adarsh, V., Arun Kumar, P., Lavanya, V., & Gangadharan, G. R. (2023). Fair and explainable depression detection in social media. *Information Processing and Management*, 60(1), 103168. <https://doi.org/10.1016/j.ipm.2022.103168>
- Alorf, A. (2021). The Practicality of deep learning algorithms in COVID-19 detection: application to chest X-ray images. *Algorithms*, 14(6), 183. <https://doi.org/10.3390/a14060183>
- Aqajari, S. A. H., Naeini, E. K., Mehrabadi, M. A., Labbaf, S., Rahmani, A. M., & Dutt, N. (2020). GSR analysis for stress: Development and Validation of an open source tool for noisy naturalistic GSR data. <https://doi.org/10.48550/arxiv.2005.01834>
- Bienefeld, N., Boss, J. M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., et al. (2023). Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *Npj Digital Medicine*, 6(1), 94. <https://doi.org/10.1038/s41746-023-00837-4>
- Blalock, D., Ortiz, J. J. G., Frankle, J., & Gutttag, J. (2020). What is the state of neural network pruning? <https://arxiv.org/abs/2003.03033>
- Bobade, P., & Vani, M. (2020). Stress detection with machine learning and deep learning using multi-modal physiological data. In *2020 second international conference on inventive research in computing applications (ICIRCA)* (pp. 51–57). <https://doi.org/10.1109/ICIRCA48905.2020.9183244>
- Bracha, H. S., Ralston, T. C., Matsukawa, J. M., Williams, A. E., & Bracha, A. S. (2004). Does “fight or flight” need updating? *Psychosomatics*. England. <https://doi.org/10.1176/appi.psy.45.5.448>
- Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. *JAMA*, 267(9), 1244–1252.
- Das, S., Sultana, M., Bhattacharya, S., Sengupta, D., & De, D. (2023). XAI-reduct: Accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI. *The Journal of Supercomputing*. <https://doi.org/10.1007/s11227-023-05356-3>
- Dave, D., Naik, H., Singhal, S., & Patel, P. (2020). Explainable AI meets healthcare: A study on heart disease dataset.



- Delaney, J. P., & Brodie, D. A. (2000). Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Perceptual and Motor Skills*, 91(2), 515–524. <https://doi.org/10.2466/pms.2000.91.2.515>
- Denton, E., Zaremba, W., Bruna, J., LeCun, Y., & Fergus, R. (2014). Exploiting linear structure within convolutional networks for efficient evaluation. *Advances in Neural Information Processing Systems*, 2(January), 1269–1277.
- Dong, S., Liu, X., Li, X., Xie, G., & Tang, X. (2022). A novel pruning method based on correlation applied in full-connection layer neurons. In *Artificial intelligence and security: 8th International conference, ICAIS 2022, Qinghai, China, July 15–20, 2022, proceedings, Part II* (pp. 205–215). Springer. [https://doi.org/10.1007/978-3-031-06788-4\\_18](https://doi.org/10.1007/978-3-031-06788-4_18)
- Du, H., Feng, J., & Feng, M. (2019). Zoom in to where it matters: A hierarchical graph based model for mammogram analysis. <https://doi.org/10.48550/arXiv.1912.07517>
- ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2021). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, 37(4), 1633–1650. <https://doi.org/10.1111/coin.12410>
- Evans, S., Seidman, L. C., Tsao, J. C., Lung, K. C., Zeltzer, L. K., & Naliboff, B. D. (2013). Heart rate variability as a biomarker for autonomic nervous system response differences between children with chronic pain and healthy control children. *Journal of Pain Research*, 6, 449–457. <https://doi.org/10.2147/JPR.S43849>
- Fernandes, F. E., & Yen, G. G. (2021). Pruning deep convolutional neural networks architectures with evolution strategy. *Information Sciences*, 552, 29–47. <https://doi.org/10.1016/j.ins.2020.11.009>
- Gaube, S., Suresh, H., Raue, M., Lermer, E., Koch, T. K., Hudecek, M. F. C., et al. (2023). Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. *Scientific Reports*, 13(1), 1383. <https://doi.org/10.1038/s41598-023-28633-w>
- Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision, 2015 Inter* (pp. 1440–1448)s. <https://doi.org/10.1109/ICCV.2015.169>
- Gjoreski, M., Gjoreski, H., Luštrek, M., & Gams, M. (2016). Continuous stress detection using a wrist device: In laboratory and real life. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct* (pp. 1185–1193). Association for Computing Machinery. <https://doi.org/10.1145/2968219.2968306>
- Ham, J., Cho, D., Oh, J., & Lee, B. (2017). Discrimination of multiple stress levels in virtual reality environments using heart rate variability. In *Annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual international conference, 2017* (pp. 3989–3992). <https://doi.org/10.1109/EMBC.2017.8037730>
- Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both Weights and connections for efficient neural network. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28). Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2015/file/ae0eb3eed39d2bcef4622b2499a05fe6-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/ae0eb3eed39d2bcef4622b2499a05fe6-Paper.pdf)
- He, Y., Zhang, X., & Sun, J. (2017). Channel pruning for accelerating very deep neural networks. In *2017 IEEE international conference on computer vision (ICCV)* (pp. 1398–1406). <https://doi.org/10.1109/ICCV.2017.155>
- He, Y. (2022). Pruning very deep neural network channels for efficient inference (pp. 1–12). <http://arxiv.org/abs/2211.08339>
- He, J., Li, K., Liao, X., Zhang, P., & Jiang, N. (2019). Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal. *IEEE Access*, 7, 42710–42717. <https://doi.org/10.1109/ACCESS.2019.2907076>
- Harvard Health. (2020). Understanding the stress response. *Harvard health*. <https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response>. Retrieved November 26, 2022.
- Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? <https://doi.org/10.48550/arXiv.1712.09923>
- Hovsepian, K., Al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., & Kumar, S. (2015). cStress: Towards a Gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the ... ACM international conference on ubiquitous computing. UbiComp (conference)* (Vol. 2015, pp. 493–504). <https://doi.org/10.1145/2750858.2807526>
- Hsieh, C. P., Chen, Y. T., Beh, W. K., & Wu, A. Y. A. (2019). Feature selection framework for XGBoost based on electrodermal activity in stress detection. In *IEEE workshop on signal processing systems, SiPS: Design and implementation, 2019-October* (pp. 330–335). <https://doi.org/10.1109/SiPS47522.2019.9020321>
- Hu, H., Peng, R., Tai, Y.-W., & Tang, C.-K. (2016). Network trimming: A data-driven neuron pruning approach towards efficient deep architectures. <http://arxiv.org/abs/1607.03250>

- Ishaque, S., Khan, N., & Krishnan, S. (2021). Trends in heart-rate variability signal analysis. *Frontiers in Digital Health*, 3, 639444. <https://doi.org/10.3389/fgdth.2021.639444>
- Jaderberg, M., Vedaldi, A., & Zisserman, A. (2014). Speeding up convolutional neural networks with low rank expansions. In *BMVC 2014—Proceedings of the British Machine vision conference 2014*. <https://doi.org/10.5244/c.28.88>
- Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, 15(3), 235–245. <https://doi.org/10.30773/pi.2017.08.17>
- Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable artificial intelligence for human decision support system in the medical domain. *Machine Learning and Knowledge Extraction*, 3(3), 740–770. <https://doi.org/10.3390/make3030037>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Lawanont, W., Mongkolnamp, P., Nukoolkit, C., & Inoue, M. (2019). Daily stress recognition system using activity tracker and smartphone based on physical activity and heart rate data. In I. Czarnowski, R. J. Howlett, L. C. Jain, & L. Vlacic (Eds.), *Intelligent decision technologies 2018* (pp. 11–21). Springer.
- Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., et al. (2021). BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74, 102233. <https://doi.org/10.1016/J.MEDIA.2021.102233>
- Liang, T., Glossner, J., Wang, L., Shi, S., & Zhang, X. (2021). Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing*, 461, 370–403. <https://doi.org/10.1016/j.neucom.2021.07.045>
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- McEwen, B. S., & Stellar, E. (1993). Stress and the individual. Mechanisms leading to disease. *Archives of Internal Medicine*, 153(18), 2093–2101.
- Moridani, M. K., Mahabadi, Z., & Javadi, N. (2020). Heart rate variability features for different stress classification. *Bratislava Medical Journal*, 121(9), 619–627. [https://doi.org/10.4149/BLL\\_2020\\_107](https://doi.org/10.4149/BLL_2020_107)
- Müller, H., Holzinger, A., Plass, M., Brcic, L., Stumptner, C., & Zatloukal, K. (2022). Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation. *New Biotechnology*, 70, 67–72. <https://doi.org/10.1016/j.nbt.2022.05.002>
- Oskooei, A., Chau, S. M., Weiss, J., Sridhar, A., Martínez, M. R., & Michel, B. (2019). DeStress: Deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (HRV) data. *Studies in Computational Intelligence*, 914, 93–105. <https://doi.org/10.48550/arxiv.1911.13213>
- Pai, K.-C., Wang, M.-S., Chen, Y.-F., Tseng, C.-H., Liu, P.-Y., Chen, L.-C., et al. (2021). An artificial intelligence approach to bloodstream infections prediction. *Journal of Clinical Medicine*, 10(13), 2901. <https://doi.org/10.3390/jcm10132901>
- Pasandi, M. M., Hajabdollahi, M., Karimi, N., & Samavi, S. (2020). Modeling of pruning techniques for simplifying deep neural networks. In *Iranian conference on machine vision and image processing, MVIP, 2020-Febru*. <https://doi.org/10.1109/MVIP49855.2020.9116891>
- Pattepu, S., Mukherjee, A., Routray, S., Mukherjee, P., Qi, Y., & Datta, A. (2023). Multi-antenna relay based cyber-physical systems in smart-healthcare NTN: An explainable AI approach. *Cluster Computing*, 26(4), 2259–2269. <https://doi.org/10.1007/s10586-022-03632-0>
- Pham, T., Lau, Z. J., Chen, S. H. A., & Makowski, D. (2021). Heart rate variability in psychology: A review of HRV indices and an analysis tutorial. *Sensors*. <https://doi.org/10.3390/s21123998>
- Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175–1191. <https://doi.org/10.1109/34.954607>
- Pourmohammadi, S., & Maleki, A. (2020). Stress detection USING ECG and EMG signals: A comprehensive study. *Computer Methods and Programs in Biomedicine*, 193, 105482. <https://doi.org/10.1016/j.cmpb.2020.105482>
- Ramteke, R., & Thool, V. R. (2017). Stress detection of students at academic level from heart rate variability. In *2017 international conference on energy, communication, data analytics and soft computing (ICECDS)* (pp. 2154–2157). <https://doi.org/10.1109/ICECDS.2017.8389833>
- Rastegari, M., Ordonez, V., Redmon, J., & Farhadi, A. (2016). XNOR-Net: ImageNet classification using binary convolutional neural networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *ECCV* (pp. 525–542). Springer.

- Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multi-modal deep learning. *Expert Systems with Applications*, 138, 112793. <https://doi.org/10.1016/j.eswa.2019.07.010>
- Rigas, G., Goletsis, Y., & Fotiadis, D. I. (2012). Real-time driver's stress event detection. *IEEE Transactions on Intelligent Transportation Systems*, 13(1), 221–234. <https://doi.org/10.1109/TITS.2011.2168215>
- Rodríguez-Arce, J., Lara-Flores, L., Portillo-Rodríguez, O., & Martínez-Méndez, R. (2020). Towards an anxiety and stress recognition system for academic environments based on physiological features. *Computer Methods and Programs in Biomedicine*, 190, 105408. <https://doi.org/10.1016/j.cmpb.2020.105408>
- Rosmond, R., & Björntorp, P. (1998). Endocrine and metabolic aberrations in men with abdominal obesity in relation to anxiety-depressive infirmity. *Metabolism: Clinical and Experimental*, 47(10), 1187–1193. [https://doi.org/10.1016/s0026-0495\(98\)90321-3](https://doi.org/10.1016/s0026-0495(98)90321-3)
- Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *Annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual international conference, 2007* (pp. 4656–4659). <https://doi.org/10.1109/IEMBS.2007.4353378>
- Sánchez-Reolid, R., Martínez-Rodrigo, A., López, M. T., & Fernández-Caballero, A. (2020). Deep Support vector machines for the identification of stress condition from electrodermal activity. *International Journal of Neural Systems*, 30(7), 2050031. <https://doi.org/10.1142/S0129065720500318>
- Sarp, S., Catak, F. O., Kuzlu, M., Cali, U., Kusetogullari, H., Zhao, Y., et al. (2023). An XAI approach for COVID-19 detection using transfer learning with X-ray images. *Heliyon*, 9(4), e15137. <https://doi.org/10.1016/j.heliyon.2023.e15137>
- Saskia, K., Neerinx, M. A., & Kraaij, W. (2014). The SWELL knowledge work dataset for stress and user modeling research categories and subject descriptors. In *Proceedings of the 16th international conference on multi-modal interaction, November 2014* (pp. 291–298).
- Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018). Introducing WESAD, a multi-modal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multi-modal interaction* (pp. 400–408). Association for Computing Machinery. <https://doi.org/10.1145/3242969.3242985>
- Selye, H. (1976). Stress without distress. *Psychopathology of Human Adaptation*, 25, 137–146. [https://doi.org/10.1007/978-1-4684-2238-2\\_9](https://doi.org/10.1007/978-1-4684-2238-2_9)
- Seo, W., Kim, N., Kim, S., Lee, C., & Park, S.-M. (2019). Deep ECG-respiration network (DeepER Net) for recognizing mental stress. *Sensors*, 19(13), 2. <https://doi.org/10.3390/s19133021>
- Shao, Y., Cheng, Y., Shah, R. U., Weir, C. R., Bray, B. E., & Zeng-Treitler, Q. (2021). Shedding light on the black box: explaining deep neural network prediction of clinical outcomes. *Journal of Medical Systems*, 45(1), 5. <https://doi.org/10.1007/s10916-020-01701-8>
- Soberanis-Mukul, R. D., Navab, N., & Albarqouni, S. (2020). Uncertainty-based graph convolutional networks for organ segmentation refinement. In T. Arbel, I. Ben Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, & C. Pal (Eds.), *Proceedings of the third conference on medical imaging with deep learning* (Vol. 121, pp. 755–769). PMLR. <https://proceedings.mlr.press/v121/soberanis-mukul20a.html>
- Tanev, G., Saadi, D. B., Hoppe, K., & Sorensen, H. B. D. (2014). Classification of acute stress using linear and non-linear heart rate variability analysis derived from sternal ECG. In *Annual international conference of the IEEE engineering in medicine and biology society. IEEE Engineering in medicine and biology society. Annual international conference, 2014* (pp. 3386–3389). <https://doi.org/10.1109/EMBC.2014.6944349>
- Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813. <https://doi.org/10.1109/TNNLS.2020.3027314>
- Uddin, M. Z., Dysthe, K. K., Følstad, A., & Brandtzaeg, P. B. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34(1), 721–744. <https://doi.org/10.1007/s00521-021-06426-4>
- Wang, J.-S., Lin, C.-W., & Yang, Y.-T.C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing*, 116, 136–143. <https://doi.org/10.1016/j.neucom.2011.10.047>
- Wang, X., You, X., Zhang, L., Huang, D., Aramini, B., Shabatur, L., et al. (2021). A radiomics model combined with XGBoost may improve the accuracy of distinguishing between mediastinal cysts and tumors: A multicenter validation analysis. *Annals of Translational Medicine*, 9(23), 1737–1737. <https://doi.org/10.21037/atm-21-5999>

- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- Xue, J., Li, J., & Gong, Y. (2013). Restructuring of deep neural network acoustic models with singular value decomposition. In *Interspeech 2013* (pp. 2365–2369). ISCA: ISCA. <https://doi.org/10.21437/Interspeech.2013-552>
- Zainudin, Z., Hasan, S., Shamsuddin, S. M., & Argawal, S. (2021). Stress detection using machine learning and deep learning. *Journal of Physics: Conference Series*, 1997(1), 25. <https://doi.org/10.1088/1742-6596/1997/1/012019>
- Zalabarria, U., Irigoyen, E., Martinez, R., Larrea, M., & Salazar-Ramirez, A. (2020). A low-cost, portable solution for stress and relaxation estimation based on a real-time fuzzy algorithm. *IEEE Access*, 8, 74118–74128. <https://doi.org/10.1109/ACCESS.2020.2988348>
- Zangróniz, R., Martínez-Rodrigo, A., López, M. T., Pastor, J. M., & Fernández-Caballero, A. (2018). Estimation of mental distress from photoplethysmography. *Applied Sciences*, 8(1), 25. <https://doi.org/10.3390/app8010069>
- Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 6848–6856). IEEE. <https://doi.org/10.1109/CVPR.2018.00716>
- Zubair, M., & Yoon, C. (2020). Multilevel mental stress detection using ultra-short pulse rate variability series. *Biomedical Signal Processing and Control*, 57, 101736. <https://doi.org/10.1016/j.bspc.2019.101736>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.