State-of-the-Art of Stress Prediction from Heart Rate Variability Using Artificial Intelligence

Yeaminul Haque[1] · Rahat Shahriar Zawad[1] · Chowdhury Saleh Ahmed Rony[1] · Hasan Al Banna[2] · Tapotosh Ghosh[3] · M. Shamim Kaiser[4] · Mufti Mahmud[5,6,7]

Abstract

Recent advancements in the manufacturing and commercialisation of miniaturised sensors and low-cost wearables have enabled an effortless monitoring of lifestyle by detecting and analysing physiological signals. Heart rate variability (HRV) denotes the time interval between consecutive heartbeats.The HRV signal, as detected by the sensors and devices, has been popularly used as an indicative measure to estimate the level of stress, depression, and anxiety. For years, artificial intelligence (AI)-based learning systems have been known for their predictive capabilities, and in recent years, AI models with deep learning (DL) architectures have been successfully applied to achieve unprecedented accuracy. In order to determine effective methodologies applied to the collection, processing, and prediction of stress from HRV data, this work presents an in depth analysis of 43 studies reporting the application of various AI algorithms. The methods are summarised in tables and thoroughly evaluated to ensure the completeness of their findings and reported results. To make the work comprehensive, a detailed review has been conducted on sensing technologies, pre-processing methods applied on multi-modal data, and employed prediction models. This is followed by a critical examination of how various Machine Learning (ML) models, have been utilised in predicting stress from HRV data. In addition, the reported reseults from the selected studies have been carefully analysed to identify features that enable the models to perform better. Finally, the challenges of using HRV to predict stress are listed, along with some possible mitigation strategies. This work aims to highlight the impact of AI-based stress prediction methodologies from HRV data, and is expected to aid the development of more meticulous techniques.

Keywords HRV · Physiological sensors · Stress prediction · Multi-modal data · Machine Learning · Deep Learning

✉ Mufti Mahmud
mufti.mahmud@ntu.ac.uk; mufti.mahmud@gmail.com

Yeaminul Haque
yeaminbijoy1997@gmail.com

Rahat Shahriar Zawad
rahatshahriar06@gmail.com

Chowdhury Saleh Ahmed Rony
rony.ict.bup@gmail.com

Hasan Al Banna
Hasan.banna@bup.edu.bd

Tapotosh Ghosh
tapotosh@cse.uiu.ac.bd; tapotoshghosh@gmail.com

M. Shamim Kaiser
mskaiser@juniv.edu

1   Department of ICT, Bangladesh University of Professionals, Dhaka, Bangladesh

2   Department of CSE, Bangladesh University of Professionals, Dhaka, Bangladesh

3   Department of CSE, United International University, Dhaka, Bangladesh

4   Institute of Information Technology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

5   Department of Computer Science, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK

6   Computing and Informatics Research Centre, Nottingham Trent University, Clifton Lane, NG11 8NS Nottingham, UK

7   Medical Technologies Innovation Facility, Nottingham Trent University, Clifton Lane, NG11 8NS Nottingham, UK

Introduction

In the human body, numerous receptors such as skin and eyes receive external environmental stimuli, transmit the signal to the brain for processing, and then produce a corresponding response. The harmful stimuli modify the human body's internal or external steady-state conditions (both physical and chemical). To correct this imbalance, the human body develops stress in order to maintain a steady state condition, also known as homeostasis [1]. This stress is detected by the body's sympathetic nervous system, which results in the secretion of hormones such as cortisol. The stress hormone increases the blood sugar, alertness, and blood pressure to supply additional blood flow in the body [2].

The heart rate (HR) is defined as the number of heartbeats per minute. The discrepancy in the time intervals between consecutive heartbeats (interbeat intervals (IBIs)) is considered heart rate variability (HRV). The autonomic nervous system (ANS), a rudimentary nervous system component, regulates unconscious body effects such as HR, ventilation, metabolism, mental stress, and hypertension [3]. Non-invasive monitoring of HRV offers a numerical metric to evaluate blood pressure [4]. Numerous HRV-derived parameters are used to diagnose mental stress and are indeed a critical indicator to evaluate body and mind conditions.

The resting HR of a person critically ranges from 60 to 90 beats per minute. When a person gets stressed, their HR rises dramatically. Increased HR causes a considerable increase in blood pressure, which is linked to low HRV. Thus, low HRV is a well-known indicator of stress. Therefore, it clearly shows that stress is closely related to the neurological system and the balance of the human body [5]. A growing amount of data shows a rising incidence of stress-related health problems connected to today's hectic lifestyle. Therefore, predicting stress has become a priority in order to maintain a productive and healthy lifestyle [6].

There is a growing demand for fast and efficient stress detection systems that can effectively help people understand and manage their stress levels. There have been many models developed for the prediction of stress from physiological parameters (such as electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), blood pressure (BP), HRV), behavioural features (such as facial expression, speech, posture), and self-reported questionnaires. In addition, current pieces of research have emphasised the significance of monitoring physiological signals in order to provide user's brief and effective feedback during regular tasks [7].

Collecting relevant data from cell phones is quite convenient and straightforward in this age of technological progress. Behavioural patterns, as well as physiological data (GSR, EEG, HR), can be collected through smartphones, and by combining these sensor data and smartphone records (calls, locations), stress can be predicted [8]. Video cameras, accelerometers, and touch displays based on data can also be a stress predictor and used for model construction [9]. However, smartphone-based data is not that accurate as the sensors are not medical grade. Furthermore, only device-based systems have some situation-based constraints where it gives poor predictions.

Figure 1 shows a possible representation of the AI framework for predicting stress from multi-modal sensor data. The EEG parameters and stress questionnaires are the most used mental stress detectors for participants in a contained environment. The feature sets by combining EEG measurements (distraction, levels of engagement, cognitive state) with statistical characteristics (mean, median, mode, and variance) are used to categorise stress levels into high and low categories [10]. But the heterogeneously collected self-reported stress questionnaires are susceptible to missing values and the halo effect, which results in a defective prediction model [7]. In addition, from a psychological point of view, self-reports are more related to current feelings.

Figure 2a presents the year-wise, and Fig. 2b shows the algorithm-wise distribution of the research works in this article.

In order to conduct this review, stress prediction studies that incorporate AI-based techniques, which predict HRV, were searched in sources such as the IEEE Xplore digital library, Science Direct, PubMed, and Google Scholar. For this purpose, 242 papers were initially found. After removing duplicates and reviewing the abstracts, 102 publications were chosen for full-text review. After reviewing the entire text of these publications, 56 were eliminated since they were not stress prediction-focused studies that incorporated both HRV and AI. Finally, we have thoroughly reviewed 43 articles in this research. Figure 3 depicts the process of selecting articles for this study.

A population pyramid depicting the distribution of male and female participants in available datasets is presented in Fig. 4. A word cloud representing the keywords extracted from article titles is presented in Fig. 5.

Related Works

Many researchers make use of machine learning (ML) and/or rule-based (RB) methods to infer the mental state of an individual based on HRV. The HRV can be estimated using a variety of physiological measures, including heart rate, galvanic skin response, body temperature, and blood pressure. In this section, we have presented several works that provide

Fig. 1 A possible representation of AI framework for predicting stress from multi-modal sensor data. Data collected from different parts of the human body (such as the brain, chest, and hands). The prediction model receives these data as input signals. AI (rule-based, shallow machine learning, deep machine learning)-based algorithms were taken into account in this research to detect stress levels
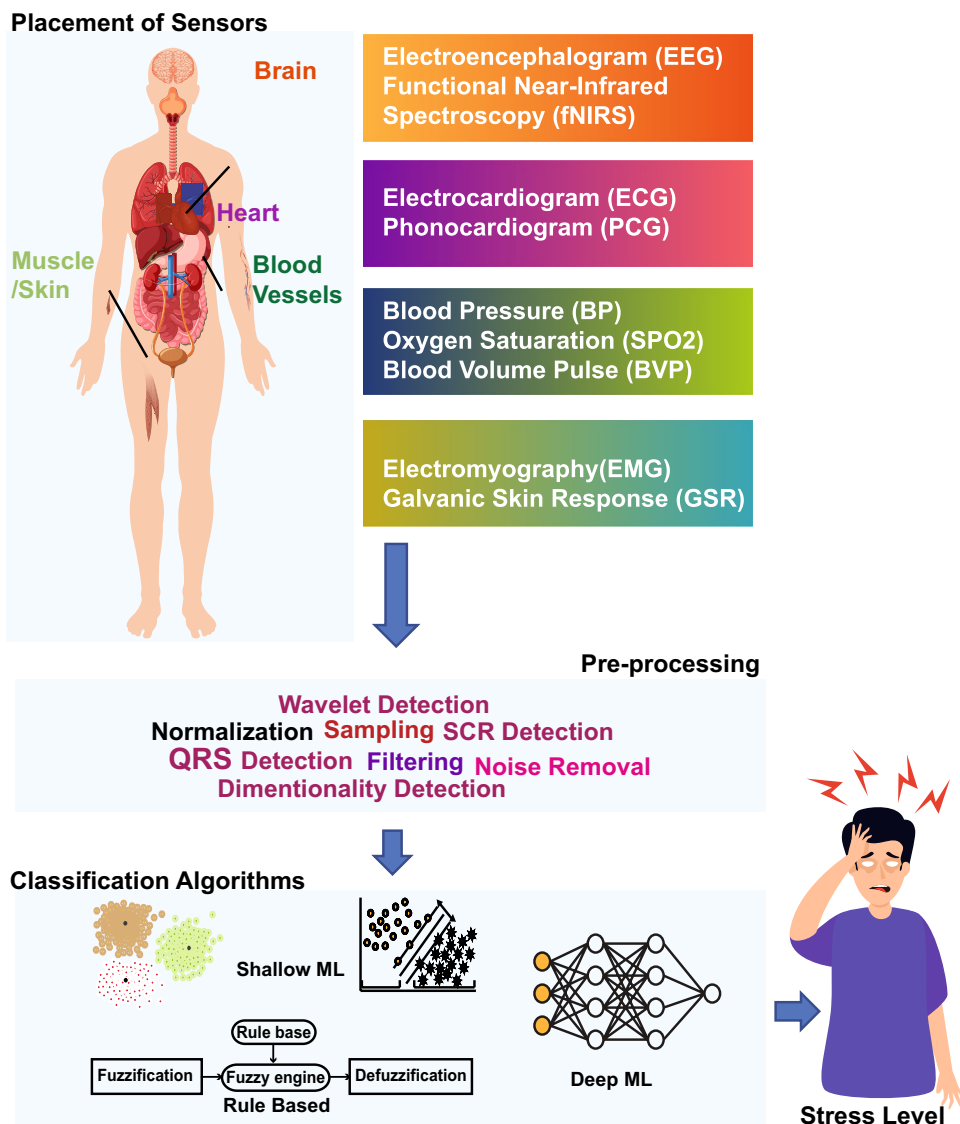

Placement of Sensors

Brain

Electroencephalogram (EEG)
Functional Near-Infrared Spectroscopy (fNIRS)

Heart

Electrocardiogram (ECG)
Phonocardiogram (PCG)

Muscle /Skin

Blood Vessels

Blood Pressure (BP)
Oxygen Satuaration (SPO2)
Blood Volume Pulse (BVP)

Electromyography(EMG)
Galvanic Skin Response (GSR)

Pre-processing

Wavelet Detection
Normalization Sampling SCR Detection
QRS Detection Filtering Noise Removal
Dimentionality Detection

Classification Algorithms

Shallow ML

Rule base
Fuzzification → Fuzzy engine → Defuzzification
Rule Based

Deep ML

Stress Level

(a)

(b)

Fig. 2 **a** The year-wise distribution of studies. We presented the number of research papers on stress prediction using AI approaches that occurred between 2016 and 2021. **b** The algorithm-wise distribution of the articles. The bar diagram shows the most popular algorithms found in the reviewed studies
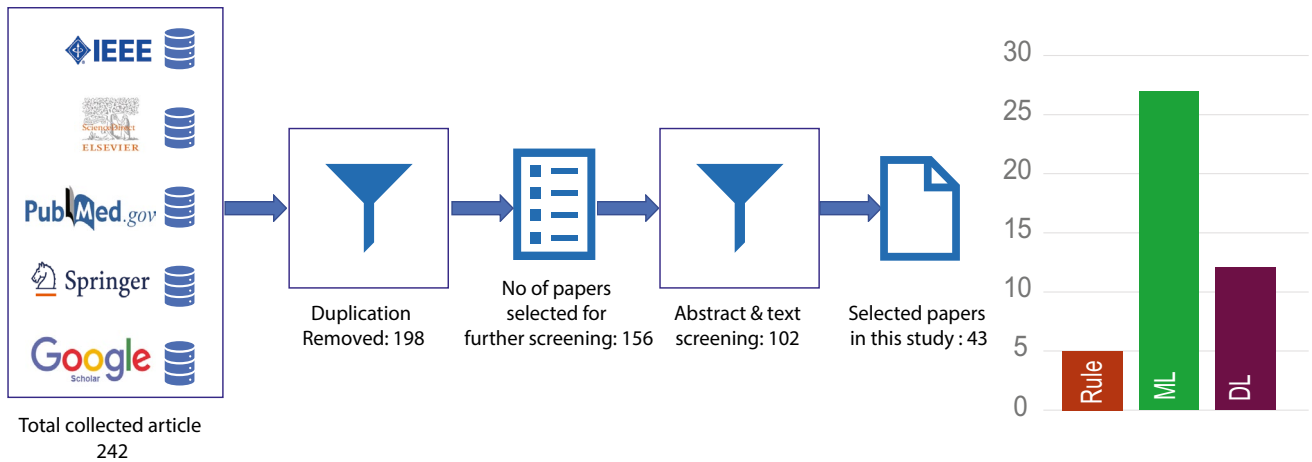
Fig. 3 Reviewing research articles, we identified 242 research publications in Science Direct, PubMed, Google Scholar, and the IEEE Xplore digital library at first. Ultimately, 43 research articles were chosen for this review after the screening process

a review of HRV-based stress prediction models with ML algorithms and rule-based approaches.

Panicker and Gayathri [11] proposed extensive reviews on different ML algorithms such as support vector machine (SVM), *K*-nearest neighbour (KNN), multilayer perceptron (MLP), long short-term memory (LSTM), decision tree (DT), linear discriminant analysis (LDA), Naïve Bayes (NB), logistic regression (LR), and probabilistic neural network (PNN) to predict various emotions (fear, anger, sadness) and stress using physiological data. These data were collected using EEG, ECG, GSR, and skin conductivity sensors. They investigated the connections between the biological characteristics of persons with emotional and mental stress. The authors have ignored different state-of-the-art RB methods in their survey. They did not incorporate the pre-processing

of data properly. These RB systems attracted researchers due to their explainability and better performance for the small dataset.

Piotrowski and Szypulska [12] provided a comprehensive overview of KNN, NB, and neural network (NN)-based drowsiness detection methods relying on HRV data extracted from ECG, EEG, and electrooculography (EOG) readings. They reviewed several ML techniques as well as pre-pressing approaches for this purpose. However, RB techniques were not included in their research.

Can et al. [13] investigated various stress detection approaches using data from smartphones and wearable sensors like ECG, EMG, electrodermal activity (EDA), EEG, GSR, and PPG. They classified the outputs into stress levels and classes. In their review, the authors focused more

Fig. 4 Population distribution of datasets from the selected articles showing the number of male and female participants
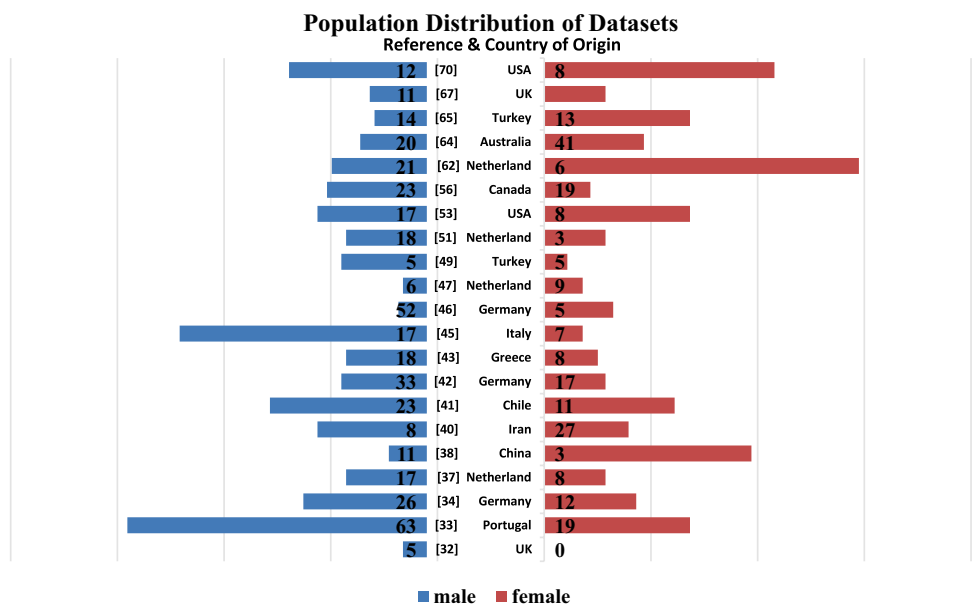
Fig. 5 The word cloud depicting retrieved keywords from the title of the articles

on multimodal data-gathering approaches for stress detection. The authors addressed several ML and RB approaches like SVM, LDA, LR, AdaBoost, KNN, fuzzy logic, NB, and convolutional neural networks (CNN). Smartphone sensors and wearable-based data were used only. They avoided research challenges and different preprocessing techniques for the data.

Bulagang et al. [14] looked into emotion categorisation based on ECG and EEG sensor data. They also used EDA, HR sensor, GSR, etc. and reviewed some ML and RB algorithms KNN, SVM, fuzzy logic, and random forest (RF) utilising data from numerous sensors. The authors considered the utilisation of multimodal physiological signals. They concentrated their research on several emotion categories rather than stress. The paper was also lacking in pre-processing methods and data fusion.

Pramanta et al. [15] studied stress identification methods to identify stress levels depending only on the HR data. The authors investigated various ways of extracting properties from heartbeat data collected using HRV, GSR, BP (blood pressure), and EEG sensors, as well as the performance of SVM, RF, NB, DT, and KNN approaches based on such data. They concentrated on classification methods and overlooked multimodal and fusion-based data processing. Furthermore, RB techniques were not included in their research. Both multimodal data and RB techniques can perform better when it comes to identification and classification tasks.

Katarya and Maan [16] proposed a review of stress detection using SVM, KNN, LR, RF by GSR, EDA, skin temperature (ST), blood volume pressure (BVP), HR, and HRV data collected from smartwatches. They explored various smartwatch-based data collection methods and compared

several ML techniques based on their stress detection abilities for different stress levels. The use of multimodal data or RB techniques ignored stress in the detection system. In addition, just a few related publications were examined for the purpose of comparing ML approaches.

Nath et al. [17] reviewed and discussed stress detection techniques which used ML algorithms like SVM, KNN, DT, LDA, NB, ANN, and RF. In their review study, the authors identified GSR, ANS, EDA, PPG, HR, HRV, EOG, EEG, ECG, EMG, EGG, and respiration-based physiological indicators for classifying stress based on subjective and objective assessments. They compared various ML algorithms' accuracy, classes, and acquisition windows. However, they did not mention data pre-processing strategies or the challenges encountered while doing research. RB procedures were not included in their assessment.

Smets et al. [18] compared SVM, LDA, Bayesian networks, DT, RF, and LR algorithms for the measurement of stress levels based on physiological responses. Their research includes data from ECG, GSR, HRV, ST, respiration, and EMG sensors. Rest detection rate, stress detection rate, and average detection rate were used to compare accuracy. The authors employed questionnaires and sensors for data collection, but data was used from one source. They tested six alternative ML algorithms but did not incorporate multimodal data or RB detection strategies in the detection system. Tonacci et al. [19] evaluated physiological data linked to ANS activity, along with ECG and GSR; ANS, ECG, HRV, HR, and cardiac sympathetic index (CSI) measures were used. The performances of SVM, KNN, DT, LDA, quadratic discriminant, and LR-based algorithms were compared for physiological stress-level detection. The authors talked about what the study could be used for in the future

and what problems other researchers might face. However, they ignored RB approaches for their study and only considered relaxation in place of stress detection.

In earlier review studies, several stress prediction approaches based on HRV were explored, which were done utilising a variety of ML techniques. The bulk of them were targeted at utilising ML to detect and classify stress. In most cases, the reviews were limited to ML methods. Only a tiny fraction of their research employed RB methods. Although pre-processing approaches prepare data for the core classification, they were mostly discarded in the bulk of the literature. For enhanced and more accurate data gathering, multimodal and fusion-based sensors are essential. Even so, the majority of the studies employed very specific types of sensors, and review papers ignored the use of multimodal sensors. A common framework for stress detection might be beneficial to add in review articles for possible future studies. However, none of the studies provided a unified framework for detecting stress.

There is no agreed standard for stress evaluation at present. This study intended to cover works that provide a basis for using HRV as a psychological stress indicator and to provide a comprehensive analysis of AI-based pre-processing and stress prediction models derived from HRV. Table 1 indicates the characteristics of the already available review articles in the field of stress prediction from HRV.

Stress Prediction and Heart Rate Variability

The field of stress research has a wide variety of applications, as it has the potential to boost learning and increase work productivity. The potential applications of stress research include the ability to enhance personal, government, and industrial operations and the resilience of military operations and life support systems [20]. As there may be discrepancies between numerical scales suggested by various researchers, stress detection systems rely on qualitative judgement. To evaluate stress levels, many researchers have utilised various sorts of phrases. Some class labels are determined only by the presence of stress; others are defined by stress and relaxation levels, which can be expressed as extremely stressed, mildly stressed, stressed, extremely relaxed, relaxed, and so on [11].

For identifying stress, HRV is a crucial feature and indicator for evaluating body and mind states. Therewith, while interpreting the relationship between HRV and stress (see Fig. 6 for the relationship between brain and HRV), it is critical to grasp the entire autonomic context and analyse a patient's medical and psychiatric history due to the diversity of possible stressors and individual stress responses [2].

Artificial Intelligence Algorithms

Recently, artificial intelligence (AI) has played a significant role in the methodological developments for diverse problem domains, including computational biology [21, 22], cyber security [23–26], disease detection [27–33] and management [34–39], elderly care [40, 41], epidemiological study [42], fighting pandemic [43–49], healthcare [50–54], healthcare service delivery [55–57], natural language processing [58–62], and social inclusion [63–65].

In this article, we categorised the algorithms found from the different review publications as RB approaches, shallow machine learning approaches, and deep machine learning approaches. This section discusses the basic principle of these three approaches and their pros and cons, along with all the algorithms we have found being used by different articles.

Rule-Based Approaches

Rule-based approach, often known as expert systems, makes judgements or solves issues by using logic and previously established rules. These rules are typically created by

Table 1 Characteristics of current review articles

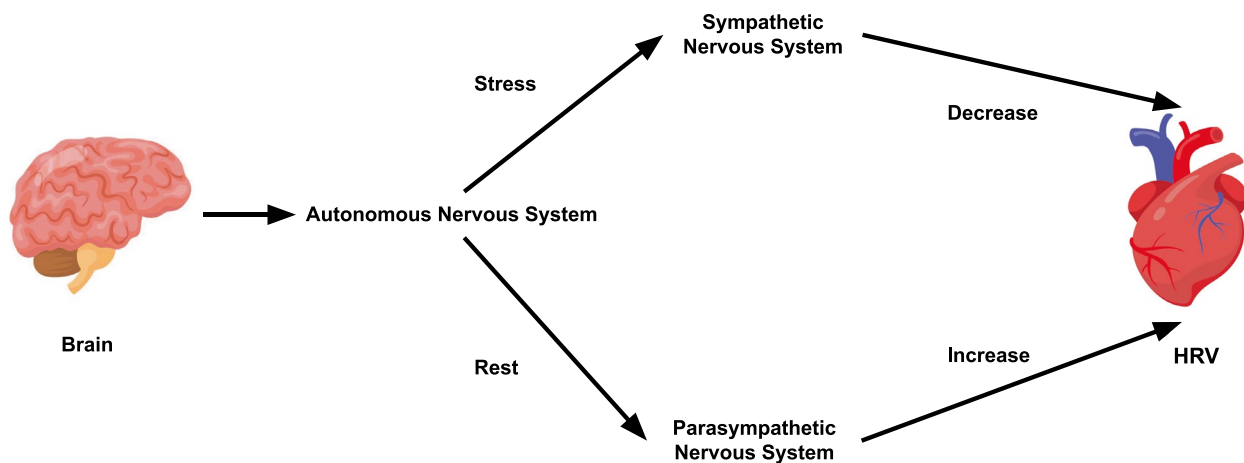| Ref	Datatype		Sensor	Data set	Data pre-processing	Classification		
	Multi-modal	Fusion				RB	ML	DL
[11]	✓	✓	✓	✓	✗	✓	✓	✓
[12]	✓	✗	✓	✗	✓	✗	✗	✗
[13]	✓	✓	✓	✓	✗	✓	✓	✓
[14]	✓	✗	✓	✓	✓	✗	✓	✓
[15]	✓	✗	✗	✗	✗	✗	✓	✗
[16]	✓	✗	✓	✗	✗	✗	✓	✗
[17]	✗	✗	✗	✗	✗	✗	✗	✗
[18]	✗	✗	✗	✗	✗	✗	✗	✗
This article	✓	✓	✓	✓	✓	✓	✓	✓

Fig. 6 The relation cycle of stress with HRV and relation with autonomous nervous system (ANS) is presented

experts in the field and are unique to their particular sectors. The system analyses incoming data and produces an output or recommendation by abiding by these rules. The decision-making process for RB approaches is transparent. In the majority of instances, the rationale and conditions for RB approaches are clearly stated and transparent. Additionally, it is useful for updating guidelines or rules of decision-making because they are clear and provide the approach flexibility and adaptability. On the other hand, RB approaches have limited compatibility in complex domains and require manual effort to design the rule bases.

Shallow Machine Learning Approaches

Shallow machine learning, sometimes called traditional machine learning or supervised learning, encompasses the process of training a model using labelled examples. Through this process, the model gains an understanding of patterns and relationships within the data, enabling it to predict outcomes or classify new, unseen data. The emphasis lies in extracting relevant features from the input data and utilising them to guide decision-making. In shallow machine learning, the model is required to be provided with labelled examples of inputs and their corresponding outputs. From these examples, the model is learned to make predictions on new, unseen data. However, shallow machine learning models often offer interpretability and take less training time than deep machine learning models. Shallow machine learning models are also easier to implement and debug.

Deep Machine Learning Approaches

Deep machine learning, often referred to as deep learning, is a subset of machine learning that uses neural networks with multiple layers to learn representations of data. It involves training a complex network of interconnected artificial neurons to automatically discover and learn hierarchical representations of the input data. Deep learning excels in tasks such as image and speech recognition, natural language processing, and other complex pattern recognition tasks. It can automatically extract features from raw data, eliminating the need for manual feature engineering. However, deep learning models often require substantial computational resources and often lack interpretability [66–70].

Summary of AI Algorithms

For stress prediction, AI algorithms have been extensively used in recent years. Table 2 described the basics of various RB and ML techniques extensively used to predict stress from HRV.

The AI algorithms which have been used throughout this reviewed article are conferred in Table 2. In this section, the algorithms, along with their pros and cons with graphical representation, have been presented. Figure 7 represents the AI algorithms in pictorial form.

AI for Stress Prediction

The widespread adoption of AI can be attributed to several factors, two of the most important of which are its remarkable accuracy and lightning-fast response times. Additionally, it does an excellent job of predicting stress, which is essential to living a healthy life.

Rule-Based Approach

Various types of RB systems, such as fuzzy logic, neuro-fuzzy systems or fuzzy neural networks [83], and fuzzy adaptive resonance theory (ART) [84], are used in clinical

Table 2 A summary of pros, cons, and descriptions of different AI algorithms used for stress prediction in recent years

Type	Pros	Cons	Description
FL	• FL models are flexible. • handle distorted, imprecise or vague data.	• The accuracy is not optimal on imprecise inputs.	Incorporates uncertainty in real-world problems. Boolean logic recognises only two states, whereas fuzzy logic can replicate human thinking considering more than two states, defined as membership degrees
FNN	• Combines the strengths of both NN and fuzzy Logic. • Allows integration of expert knowledge into the system.	• The structure of a fuzzy NN is not completely interpretable. • Not a concurrent system.	Blends fuzzy control systems' high-level, IF-THEN rule logic with neural networks' low-level learning and computing capabilities [71]. This algorithm resolves the deficit of flexibility in fuzzy number decision-making and the selection of fuzzy shapes which are able to represent expert experiences correctly [72]
NB	• Does not require that much data be trained. • Accepts both discrete and continuous-valued datasets.	• Assumes independent predictors, which is impossible in real-life data. • Has zero frequency issue in case of categorical variables.	Simply follows the basics of calculating frequency and the combination of values from a dataset [73]. A probability table serves as the model in this technique, and it is updated using training data [74]
LR	• Can classify new data from the continuous and discrete dataset. • Simple to implement and train.	• Cannot resolve nonlinear problems. • Only maps result in 0-1 Value.	Capable of using any actual value number and mapping it into the 0-1 value
DT	• Consider all possible outcomes. • Can solve both classification and regression-based problems.	• Small changes lead to structural changes. • Comparatively less accurate than other tree-based algorithms.	Constructs a model that predicts the target variable. The problem is solved using a tree representation, in which the leaf node corresponds to a class label and the core node of the tree expresses features [75]
KNN	• Can solve both regression and classification issues. • Does not require linear separability between classes.	• Suspectable to inaccurate feature variables. • Does not perform well in big data.	Used to calculate the average and maximum classification rates when assessing the baseline for test performance [76]. KNN is done by exploring for a group of K objects in the closest similar training data to objects in fresh data or data testing [75]
SVM	• Can conduct nonlinear classification. • Transfers input to high dimensional feature space.	• Has large computational cost. • Training time is high.	Seeks to identify an ideally separating hyperplane between classes using the structural risk minimisation principle. Determines the optimal decision boundary for categorising objects from the training data set [74]
KMC	• Can easily adapt to new data. • Better in solving classification tasks.	• Requires exact values of clusters. • Data grouping can be unstable.	An iterative clustering approach that classifies unlabelled datasets into different clusters by splitting from set data into cluster K. The initial value of the cluster's centre point influences cluster identification
RF	• Lower risk of overfitting. • Work on both regression and classification problems.	• Slower in computing. • Not for real-time problem-solving.	A DT-based ensemble classifier aggregates the findings of several decision trees to generate a single result [77]. Utilises the bagging method to generate several distinct sample sets [78]
RNN	• Works in a feedback loop. • Can perform speech recognition and video classification.	• Comparatively unstable. • Not possible to stack up. • Training time is slower.	Conventionally constructed by defining a transition and output function [79]. By assigning identical weights and biases to all of the layers, creates dependent activations

Table 2 (continued)

Type	Pros	Cons	Description
DNN	• Can do multi-level processing. • Can work with complicated feature representation.	• Slow learning process. • Requires a large amount of data.	DNN has multiple hidden layers in between the input and output layers. It has the ability to extract features from data without any feature extraction algorithms. DNNs mimic the multilevel processing mechanism of vision by Cortical areas of the brain [80]
CNN	• Little reliance on pre-processing. • Provide more accurate results.	• Slower in training than DNNs. • Training set needs to be much larger for successful training.	Backpropagation network consisting of multiple layers. It combines three architectural ideas: the receptive field, shared weights, and subsampling [81]. The architecture of CNN is based on the visual cortex of human beings [82]

FL fuzzy logic, *FNN* fuzzy neural network, *NB* naive Bayes, *LR* logistic regression, *DT* decision tree, *KNN* K-nearest neighbours, *SVM* support vector machine, *KMC* K-means clustering, *RF* random forest, *RNN* recurrent neural network, *DNN* deep neural network, *CNN* convolutional neural network

applications where the knowledge of different experts are converted into a set of "if-then" rules. Many researchers have utilised fuzzy logic to assess stress from HRV.

Kumar et al. [85] developed a fuzzy theoretic nonparametric deep model for predicting stress based on heartbeat analysis. In addition to the stress value, the authors created weights for subjective stress evaluation and empirical HRV analysis to illustrate the explainability of the proposed model.

El-Samahy et al. [83] proposed Mamdani fuzzy inference systems to find mental stress using heart rate and diameter of the pupil. The authors carried out a closed-loop experiment between two personal computers, one for imposing mental stress and the other for monitoring and managing the human mental state.

Ranganathan et al. [86] proposed a stress assessment approach that analyses heart rate signals using a wavelet transform and a neural fuzzy model. Techniques such as wavelet decomposition and reconstruction were employed to minimise noise and recover specific time-frequency features that were previously lost. It is necessary to apply neural fuzzy training in order to recognise spectral features, and fuzzy clustering techniques are used to evaluate mental stress. They kept track of the heart rate recordings and used the wavelet transform to evaluate the data (WT). Neuro-fuzzy evaluation approaches were used to improve the reliability of HRV analysis and to track the activity of the autonomic nervous system (ANS) under a variety of stress conditions.

Kumar et al. [87] developed a novel heart rate variability analysis technique for measuring mental stress based on fuzzy clustering. An accurate and dependable fuzzy identification technique was used to deal with the uncertainties created by individual differences in the assessment of mental stress levels. Their method requires the continuous monitoring of heart rate signals over the Internet. Later, the signals are processed by means of a continuous wavelet transform in order to recover the local features of HRV in the time-frequency domain.

Wang et al. [84] presented a pattern recognition system for learning complicated HRV-salivary stress correlations. In order to predict salivary response given a set of ECG measurements, the researchers used a fuzzy ARTMAP (FAM) classifier. They improved FAM utilising GA ensembles, which improved the training cycle order and ARTMAP parameters. They also devised a system for simultaneously collecting heart rate and salivary data under various stress induction strategies. A summary of used algorithms, pre-processing, sensors, and features by RB stress prediction approaches is presented in Table 3.
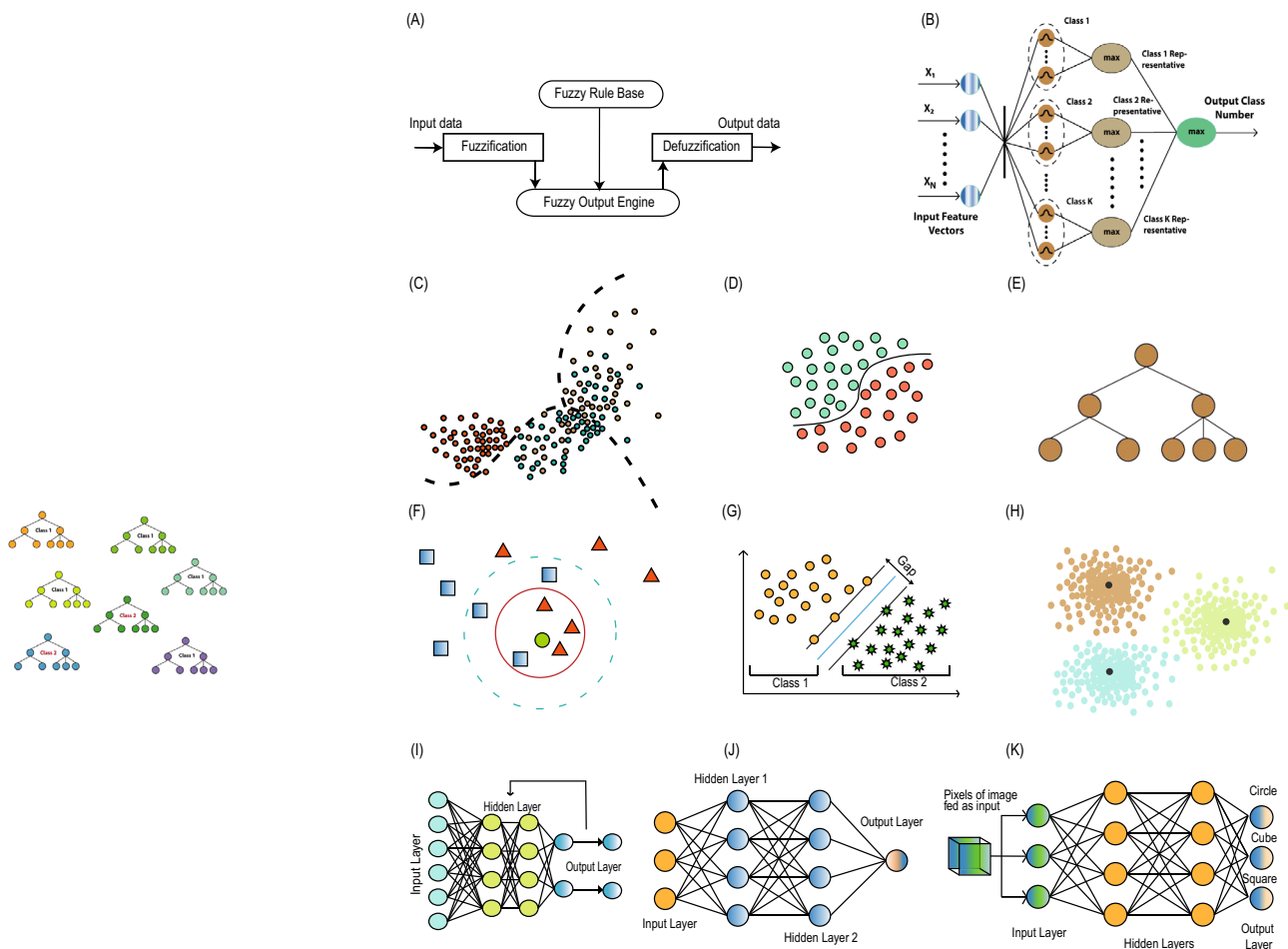
Fig. 7 AI/ML techniques used in recent research. **A** Fuzzy logic, **B** fuzzy neural network, **C** naive Bayes, **D** logistic regression, **E** decision tree, **F** random forest, **G** support vector machine, **H** *K*-means cluster, **I** RNN, **J** DNN, and **K** CNN

Shallow Machine Learning Approaches

In shallow ML, the training process is carried out using data with predefined features where it is necessary to perform feature extraction by hand, as the use of domain knowledge is essential. Shallow ML includes well-known algorithms such as RF, NB, DT, SVM, KNN, and LR. This section contains a list of studies which utilises shallow ML methods to predict stress.

Sriramprakash et al. [88] extracted the most important and overlapping characteristics from physiological sensors in order to identify stress in working individuals. The authors extracted time- and frequency-domain features as well as physiological features (HR, HRV, GSR, and so on) from the

Table 3 A summary of used algorithms, pre-processing, sensors, and features by rule-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[85]	Fuzzy theoretic nonparametric deep model	-	PPG	R-R Features
[83]	Mamdani fuzzy	Downsampling	Ohmeda 2300 Finapress, Gazepoint GP3 eye tracker	HRV2, mPD
[86]	Sugeno neuro fuzzy	Wavelet transformation, noise removal	ECG	Time-frequency features
[87]	Sugeno fuzzy clustering	Continuous wavelet transformation	Polar S810i	HR, Mw, 1/a, p1,p2,p3
[84]	Fuzzy ARTMAP	Dimensionality reduction, normalisation	ECG, microtiter plate spectrophotometer	Alpha amylase, cortisol, R-R intervals

physiological data. They employed SVM and KNN classifiers to detect stress and assess the validity of the retrieved features for stress detection.

Huang et al. [89] recruited 35 participants who wore wearable devices to collect ECG from the participants. In this experiment, the authors collected 8 HRV features, namely RMSSD, PNN50, TP, HF, LF, VLF, and the LF/HF ratio and transmitted these collected data to a smartphone via Bluetooth interface. SVM, KNN, NB, and LR were used to train the model that automatically detected the fatigue state.

Wu et al. [90] attempted to overcome the challenge of identifying physiological stress caused by engaging in physical activities. They used wristband sensors to capture biosignals. GSR, BVP, HR, ACC, and ST sensors were employed to acquire physiological data in this investigation. The authors utilised KNN, SVM, DT, NB, ensemble learning (EL), and DL models to categorise physical activities and acute physical stress.

Sevil et al. [75] reported models to detect stress and awareness levels in knowledge workers using biometric sensors. The authors used wristbands to collect biosignals like GSR, BVP, ST, and HR from knowledge workers. For the purpose of detecting stress levels and awareness, they used ML models, such as KNN, SVM, NB, DT, and DNN. The performance of these algorithms was compared with the state-of-the-art techniques.

Pourmohammadi and Maleki [91] compare the efficacy of the EMG signal and the ECG signal in detecting mental stress. This work examines the EMG signal of the right and left trapezius and the right and left erector spinal muscles in depth for multi-level stress recognition. To create stress in the laboratory, mental arithmetic, the Stroop colour word test, time constraints, and a stressful atmosphere were used. The effectiveness of EMG signals for stress detection was tested using an ECG signal.

Maldonado et al. [92] introduced an expert system that used an SVM-based features selection method to analyse the mental workload of individuals while performing daily tasks. The authors used multiple mobile devices to capture HR, blood oxygen saturation (SpO2), and temperature to construct a system for mental stress analysis.

Pluntke et al. [93] introduced a framework that uses HRV analysis to detect and classify physical and mental stress in real time without interfering with the person's activities. HRV data was labelled and gathered in controlled situations where subjects were subjected to physical, psychological, and combination stressors. They used SVM and C5 DT to segregate and identify distinct stress kinds and the relationship between HRV data and stress levels.

Giannakakis et al. [94] examines the effects of stress on HRV parameters and seeks to discover the best mix of HRV features for reliably detecting stress. In order to account for the individualised baseline of each phase in developing the stress model, the retrieved HRV features were converted correspondingly using the pairwise transformation.

Castaldo et al. [95] used linear and non-linear HRV characteristics extracted during an oral test (stress) and during rest after a holiday to detect mental stress. They showed that nonlinear ultrashort-term (3 min) HRV features might automatically predict mental stress in healthy participants. ECG sensor data was used to extract HRV features, which were then evaluated using Kubios software tools. Following that, the HRV properties were applied to statistical and data mining analysis.

Delmastro et al. [96] examine the impact of a specific training procedure on the cognitive function and stress response of a group of MCI-fragile older persons. They tested a stress detection system based on different ML algorithms to see how well they performed on a real-world dataset. They also proposed a mobile system architecture for online stress monitoring that can infer the amount of tension during a session.

Lima et al. [97] developed a model that can predict how people will react using HRV characteristics and EDA signals, which were extracted using a wearable device to provide continuous monitoring. Participants were placed through a mental arithmetic stress test to extract the HRV and EDA characteristics.

Yu et al. [98] propose a new way to track office workers' behaviour and HRV. They used ML techniques to create a classification model that could distinguish distinct work behaviours (moving the body, typing, talking, and reading) from sensor data. The system utilised a lightweight EMFi sensor for measuring the changes in pressure induced by human motions and heartbeat in office chairs.

Padmaja et al. [99] proposed a model based on four major well-being dimensions. The stress level of a person is determined by combining their HRV, sleeping pattern, social behaviour, and physical activity. They developed DetectStress, a cognitive stress-level detection system that uses smartphone daily activity data and data from a wireless physical activity tracker to evaluate an individual's stress levels in an unobtrusive manner (FITBIT).

Can et al. [100] developed an autonomous stress detection system that relies on physiological information collected from discreet smart wearable gadgets that people can take about with them. This system has modality-specific artefact removal and feature extraction techniques for real-world settings.

Chen et al. [101] investigated consumer-grade wrist-based PPG sensors, which are as cheap, convenient, and accurate as consumer ECG sensors. They created an individual stress prediction model to assess the performance of different PPG LED lights and the suitable window widths. To extract ten HRV characteristics, the authors utilised half-overlapping

moving windows (1/3/5 min). They find that a 3-min interval is adequate to distinguish between a stressful mental state and illustrate how to utilise ML methods to combine HRV features for reliable stress identification.

Koldijk et al. [102] discover that addressing individual variations is especially important when assessing mental states. The authors explored several ML techniques for inferring working circumstances and mental states from a multimodal set of sensor data, including computer logs, facial expressions, posture, and physiology. They discovered that sensor data can better predict the subjective variable "mental effort" than it can predict "felt stress".

Ciabattoni et al. [103] proposed a smart-watch-based system for collecting and analysing biosignal data in order to detect mental stress in the course of daily activities. Using data from a commercial wristwatch, they classified stress using GSR, RR interval, and body temperature (BT). Data from smartwatches is filtered and adjusted to smooth down noise and motion distortions.

Attaran et al. [104] presented a design for a multi-modal stress monitoring system. They extracted 17 different features from ECG, accelerometer, SpO2, EDA, and respiratory sensor to explore them for maximising the detection accuracy of SVM and KNN classifiers. Finally, they used the results to implement a low-power-consuming ASIC implementation of the SVM classifier in stress monitoring. Castaldo et al. [105] suggested a method using mental stress assessment to identify the extent of ultra-short HRV as a valid replacement for short HRV features. They extracted 23 ultra-short HRV features and used SVM and DT classifiers to identify their validity in the case of automatic stress assessment.

Hantono et al. [106] targeted to analyse the stress level of people while using smartphones. They used PPG heart rate sensing on mobile devices to record the heart rate of the subjects while they were doing different tasks. Finally, they compared NN, discriminant analysis, NB, and KNN algorithms while doing time- and frequency-domain analysis-based classifications.

Tiwari et al. [107] explored an SVM-based prediction model of mental stress and workload. The authors extracted HRV and breathing signals for computing ultra-short-term segments of the signals to use them as features. The system was developed to provide a fast prediction of stress and mental workload depending on frequency- and time-domain features from less than 5 min segments of the sensor readings.

Clark et al. [108] presented an RF classifier-based model for the prediction of people's stress levels at least one minute prior to the event. They extracted 42 features from GSR, respiration, and ECG sensors and expanded to 252 features. These features were used to identify whether the stress level of the subject would rise to a higher level in the coming scenarios.

Ahmad et al. [109] reported a study on stress-level assessment in virtual reality environments. They collected ECG signals from subjects under VR influence. They transformed the collected data into 1-D and 2-D forms to create a multimodal fusion of ECG data. Using this multimodal deep fusion model and RF, KNN, SVM, and XGBoost classifier (XGB) algorithms, they evaluated the performances for stress-level detection from 1-s windows.

Dalmeida and Masala [110] developed a comparative study that tests the compatibility of HRV features as physiological data to accurately classify the level of stress. This was achieved by extracting HRV parameters from ECG sensor data and selecting the more relevant features using Pearson's correlation, recursive feature elimination (RFE), and extra tree classifier. They used different ML methodologies such as KNN, SVM, MLP, RF, and gradient boosting (GB) to test and develop the best model for the purpose.

Sandulescu et al. [111] presented an SVM-based stress detection approach from data collected through wearable sensors on people. They collected the PPG value, PPG autocorrelation value, HRV value, and EDA value for each state to be determined. The model they proposed was demonstrated to detect real-time stress levels in people.

Munla et al. [112] investigated stress-level detection of drivers in a real-world driving situation. The authors extracted HRV features using domain analysis approaches such as time, frequency, time-frequency, or non-linear methods using wavelet and STFT. They built a feature vector out of the extracted parameters and tested KNN, RBF, and SVM ML approaches. A summary of used algorithms, pre-processing, sensors, and features by shallow ML-based stress prediction approaches is presented in Tables 4, 5, 6, and 7.

Deep Machine Learning Approaches

de Vries et al. [113] used learning vector quantisation (LVQ) to classify stress and relaxation from different physiological signals. To create the stress classifier, the authors collected features from ECG, GSR, and RSP data and observed cardiac activity. To train the LVQ classifier, the authors experimented with different very high-frequency band features in addition to common properties of these signals.

Son [114] created a model to forecast mood changes connected to LSTM, RNN, and LSTM-RNN in order to provide a framework that will estimate the mood based on a particular detail of people's qualitative ability to adapt. Variations in moods, such as his cognitive activity in response to his activities, surroundings, environment, HR, HRV, and other states, might be easily justified with this feature-rich wearable device in a consecutive time domain.

Rastgoo et al. [115] assessed a driver's critical situation, and the authors utilised CNN and LSTM. To construct this predictor, parameters were taken from ECG, vehicle

Table 4 A summary of used algorithms, pre-processing, sensors, and features by Shallow ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[88]	SVM, KNN	-	Kinect 3D sensor, ECG sensor	RMSSD, AVNN, SDANN, SDNN, NN50, PNN50, LF, HF, LF/HF
[89]	SVM, KNN, NB, LR	-	ECG	NN.Mean, PNN50, rMSSD, TP, LF, HF, VLF, LF/HF
[90]	k-NN, SVM, DT, NB	SG-filter, BWF, ANC algorithm	GSR, BVP, ST, 3-D ACC, HR	Time-domain, frequency-domain, and distribution features of BVP, ACC, GSR, and HR
[75]	KNN, SVM, NB, DT and DNN	-	Biometric Sensors	Statistical Features of RR
[91]	SVM	FIR filters, IIR filters, EMD and DWT	ECG v.12 devices, SX230 surface electrode, Skintact F-55 electrode	Statistical and time-domain features of RR
[92]	SVM	Canny's edge detection algorithm, LPFr	Pulse oximeter, ST, ECG, and eye tracker	HRV, HR, PS, temperature, SpO2
[93]	SVM, C4.5 DT	3-sigma rule	ECG, Polar H7 chest strap	Statistical features of RR; frequency-domain features of RR
[94]	RF, SVM	time series polynomial fit and bandpass filtered	ECG	Time-domain features of RR, HRV triangular index, ECG envelope, frequency-domain features of RR
[95]	NB, SVM, MLP, AB, C4.5 DT	QRS detector, PhysioNet's WAVE	ECG	Statistical features of RR, absolute power, frequency-domain features of RR, SampEn, D2, fa1, dfa2, ShanEn
[96]	BN, SVM, k-NN, C4.5 DT	Lomb-Scargle algorithm, LPF, CDA	Zephyr BioHarness34, Shimmer3 GSR + Development Kit5	Statistical features of RR, tot spectrum power, frequency-domain features of RR, Amps, ISCR, mean SCL, mean EDA, max EDA deflection

In pre-processing column: PPG-PD PPG peak detection, *HR* heart rate, *RR-ISF* RR-interval series filtering, *HRV and EDA* electrodermal activity features extraction, *BCG* ballistocardiography, *SG-Filter* Savitzky-Golay filter, *BWF* Butterworth filter, *ANC* adaptive noise cancellation, *EMD* empirical mode decomposition, *FIR* finite impulse response, *IIR* infinite impulse response, *LPF* low pass filter, *CDA* continuous deconvolution analysis

In sensor column: SCR skin conductance response, *SCL* skin conductance level, *RiseT* rise time, *ST* skin temperature, *ECG* ecocardiogram, *GSR* galvanic skin response, *ACC* accelerometer

characteristics, and relevant information and then input into separate CNNs as the driver's stress-level components were classified into low, medium, and high categories and then merged into a two-layer LSTM.

Akbulut et al. [116] provided a model to allow the simulation of stress as well as a variety of mood shifts based on physiological factors. The researchers used ECG, GSR, body temperature, blood pressure, glucose level, and SpO2 information to construct this framework, along with observing changes in behaviour and quantifying HRV according to stress levels. In addition to determining similar traits of these signals, the authors examined other often quite frequency band features as well as time-domain and variational analytic factors.

Coutts et al. [117] used an LSTM system to capture HRV signals from a wrist device that can monitor inter-beat intervals using mean, standard deviation, and root mean square successive difference. Physiological signals and characteristics were acquired to use this sensor reading device. The

spectrum properties were determined in a comprehensible fashion of frequency domain to construct the frequency-based ML technique.

He et al. [118] used CNN technique on various physiological signals to assess chronic perceptual anxiety and tranquillity. To construct this model, the researchers analysed characteristics from ECG, EEG, and EMG readings, along with observed heart activity. The scientists used several really quiet frequency band components as well as common aspects of these transmissions to develop the CNN-based analyser.

Qin et al. [119] assessed the BP feed-forward approach for the relaxed state, low stress, medium stress, high stress, and other metabolic variables. The authors had to obtain information from GSR or skin temperature and BVP to design an assessment technique to determine HRV characteristics. HRV features obtained from time- and frequency-domain evaluation of R-R intervals recorded during the enhanced practice session are the most efficient and precise

Table 5 A summary of used algorithms, pre-processing, sensors, and features by shallow ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[97]	SVM; LR; RF	PPG-PD; HR; RR-ISF; HRV; EDA	DA wrist, PPG, Spare, TVOC, CO2, TEMP	Statistical features of RR, SCR, SCL; RiseT
[98]	SVM; KNN; EnL	BCG processing	EMFi sensor	Statistical features
[99]	NB; DT	Normalization and transformation	Fitbit Tracker	Number of calls, duration, no. of SMS, the app usage information
[100]	PCA, SVM, KNN, LR, RF, MLP	Artefact-(interpolation/removal)	3D accelerometer (ACC), PPG, EDA	Statistical and frequency-domain features of RR
[101]	RF	Filtering using BWBPF	Wristband device, Polar H10	SD1, SD2, RMSSD, SDNN, MHR, MRRI, TP, VLP, LF, HF
[102]	NB, SVM, KNN, Bayes Net, RF; DT; MLP	-	ECG and skin conductance	Facial expressions, head orientation action units, emotion, body postures, joint angles, HR, HRV, skin conductance
[103]	KNN	Artefacts and noise removal	HR, GSR and BT	RR, GSR, BT

In pre-processing column: PPG-PD PPG peak detection, *HR* heart rate, *RR-ISF* RR-interval series filtering, *HRV and EDA* electrodermal activity features extraction, *BCG* ballistocardiography, *BWBPF* high-order Butterworth bandpass filter

In sensor column: SCR skin conductance response, *SCL* skin conductance level, *RiseT* rise time, *GSR* galvanic skin response, *PPG* photoplethysmogram, *EDA* electrodermal activity, *EMFi* electro-mechanical film

indicators of ANS at the time of constructing the artificial neural network algorithm.

Ding et al. [120] proposed a study that uses multimodal measurements to measure mental workload and validates the features for mental workload estimation. The authors created a backpropagation neural network (BPNN) classifier to evaluate the workload using physiological data (HR, HRV, EMG, ETA, and respiration), subjective ratings of mental exertion (NASA Task Load Index), and task performance metrics. They compared the BPNN's performance against KNN, SVM, medium tree, and LDA algorithms.

Kalatzis et al. [121] conducted a study that determines stress levels of older adults from ECG signals while performing a hand grip strength task. The author extracted time- and frequency-domain features of HR and HRV to perform the identification of stress and no-stress states. They proposed an optimised ANN model to identify the states and proposed the effects of this model for a better stress management system.

Dhaouadi and Ben Khelifa [122] utilised LSTM and deep neural networks (DNN) to assess legitimate anxiety levels as well as detect other lifestyle patterns in young gamers based on physiological measurements. To establish such models, researchers gathered the required characteristics from ECG, EEG, EDA, and EMG recordings and estimated emotional state variations. As a result, to construct the frameworks

Table 6 A summary of used algorithms, pre-processing, sensors, and features by shallow ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[104]	KNN, SVM		HR, accelerometer, EDA, Resp. Rate, SpO2	Statistical features of HR and RR, mean SpO2
[105]	SVM, DT	RR-SE and Corr	ECG	Statistical features of NN, dfa1, dfa2, RPlmean, RPlmax, REC, RPadet, ShanEn
[106]	NN, KNN, DA, NB	Corr	PPG	mHR, RR, SDHR, SDRR, CVRR, RMSSD, pRR20, and pRR50. ULF, VLF, LF, HF
[107]	SVM	BPF (5–25 Hz)	BioHarness 3, Zephyr	Time-domain and frequency-domain features of HRV
[108]	RF classifier	Normalising, BWF	GSR, ECG, Resp	Time-domain and frequency-domain features of GSR respiration: the mean and variance time-domain and frequency-domain features of HRV

In pre-processing column: HR heart rate, *RR-SE* RR-interval series extraction, *Corr* correlation, *BPF* bandpass filter, *BWF* Butterworth filter

In sensor column: SCR skin conductance response, *HR* heart rate, *ECG* ecocardiogram, *PPG* photoplethysmogram, *EDA* electrodermal activity, *Resp* respiration

Table 7 A summary of used algorithms, pre-processing, sensors, and features by shallow ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[109]	RF, KNN, SVM, and XGB	LS-method	ECG, EMG, PPG	Time-domain and frequency-domain features of HRV
[110]	KNN, SVM, MLP, RF, and GB	WQRS tool PhysioNet HRV toolkit, pyhrv, data normalization	Apple watch	HR, AVNN, SDNN, RMSSD, pNN50, TP, and VLF
[111]	SVM	Downsampling, noise removal	BioNomadix module from Biopac, model BN-PPGED	ppgt, ppgaut, HRV t, EDA t
[112]	SVM, KNN	Filtering, derivative, SW integration, QRS detection	ECG sensor	Time-domain and frequency-domain features of RR

In pre-processing column: SW squaring and window integration, *LS-method* Lomb-Scargle method

In sensor column: EMG electromyography, *HR* heart rate, *ECG* ecocardiogram

(LSTM and DNN), the scholars had to conduct several investigations and frequency variations to determine the frequent and unusual properties.

Stewart et al. [123] suggests neural processes as a technique for developing personalised models and addressing individual interactions with physiological processes. They used standard ML models (such as SVM, KNN) and neural processes to develop stress classifiers which were compared on two datasets using leave-one-participant cross-validation.

Silva et al. [124] compared baseline and stress situations to look at HR and HRV indicators. The authors used several statistical tests and ML models, both shallow (which includes SVM, KNN, and RF) and deep, to build a predictive model for stress monitoring, evaluation, and chronic stress prediction.

A summary of used algorithms, pre-processing, sensors, and features by deep ML-based stress prediction approaches is presented in Tables 8 and 9.

Performance Analysis and Discussion

Rule-Based Approaches

Kumar et al. [85] addressed the issue of explainability of fuzzy theoretic nonparametric deep model applications in biology and medicine. They used one previously studied dataset of 50 subjects and a new dataset of 100 subjects and obtained (Pearson's correlation coefficient (r): 0.8162 (old dataset) vs 0.6809 (new dataset), RMSE: 6.8382 (old dataset) vs 9.4872 (new dataset)).

El-Samahy et al. [83] found a close match between the measurement of the proposed system and the actual measurements acquired from human volunteers. The system was built and evaluated using heart rate and pupil diameter data collected from 5 people. To compare the achievements of subjects 1 and 2, an evaluation index (EI)

Table 8 A summary of used algorithms, pre-processing, sensors, and features by deep ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[113]	LVQ	R-PD, IBI outlier removal, and SCRG method	ECG, GSR, RSP	Time-domain, frequency-domain, and distribution features of ECG, GSR, RSP
[114]	LSTM-RNN	All features are normalised into a range of [-1,1]	A Tizen component on smart-watch	HR avg, HRV arg, HR min, HR max, SDHR, SDHRV, hr diff avg, hr diff var
[115]	CNN-LSTM	A Butterworth band-pass filter (5–15 Hz), R-peaks are extracted, Pan-Tompkins algorithm	ECG	Mean, standard deviation, mean of the first difference of HRV, average normal-to-normal (NN) and intervals, SDNN, RMSSD, PNN50
[116]	FFNN	DWT, Pan-Tompkins algorithm, down sampling	CVDiMo wearable sensor	Statistical and frequency-domain features of RR
[117]	LSTM	Noise filtering	The bio beam band	Statistical and frequency-domain features of RR
[118]	CNN	Bandpass filter	ECG	HR, LH, SDNN, SD2, pQ

In pre-processing column: SCRG SCRGauge method, *HR* heart rate, *RR-SE* RR-interval series extraction; *Corr* correlation, *BPF* bandpass filter, *BWF* Butterworth filter; *DWT* discrete wavelet transformation, *IBI* inter-beat interval

In sensor column: GSR galvanic skin response, *HR* heart rate, *ECG* ecocardiogram, *RSP* respiration

Table 9 A summary of used algorithms, pre-processing, sensors, and features by deep ML-based stress prediction approaches

Ref.	Model	Pre-processing	Sensors	Features
[119]	NN	-	Pulse oximeter	MEAN, SDNN, SDANN, RMSSD, TF, VLF, LF, HF, LF/HF
[120]	BPNN, SVM, KNN	WD and HP, LP, and RMS filtering	ECG, EDA, EMG, respiration sensors	AVHR, LF/HF, Yrm, MF, SC mean, respiration
[121]	ANN	WT	ECG probe, BIOPAC ECG100C	Time-domain and frequency-domain features of NN
[122]	LSTM, DNN	Transfer function, biosppy and pyhrv libraries	Polar H10, Actigraph wGT3X-BT	HR and HRV features
[123]	NP, SVC, KNN	HT algorithm, BWF	ECG, GSR	Time-domain and frequency-domain features of HR, time-domain features of GSR
[124]	LR, NB, NN, SVM, RF, KNN	-	PPG	Mean RR, Min RR, Max RR, Median RR, SDNN, RMSSD, pNN50

In pre-processing column: HR heart rate, *WD* wavelet denoising, *BPF* bandpass filter, *BWF* Butterworth filter, *WT* wavelet transformation, *HT* Hamilton-Tompkin, *HP* high pass, *LP* low pass;

In sensor column: GSR galvanic skin response, *HR* heart rate, *ECG* ecocardiogram, *EMG* electrmyogram, *EDA* electrodermal activity, *PPG* photoplethysmogram

was produced for each of them. During levels 1–3, subject 1 had a high EI of over 90%. On the other hand, subject 2 showed an EI between 60 and 90% throughout the whole experiment, which means the levels of mental stress will be unchanged.

Ranganath et al. [86], using their proposed wavelet transform and neuro-fuzzy inference system, evaluate stress using HRV. To investigate the activity of the ANS, the authors performed a time-frequency analysis (TFA) of HRV, which can be used to quantify mental stress. The authors studied 20 physically fit adults at two points in time: before and after they began smoking and acquired a spectral decomposition of HRV. These were used to build the proposed NF-based model.

Kumar et al. [87] proposed a fuzzy clustering method which helped to quantify mental stress and demonstrate a direct functional link between ANS activities and mental stress. The researchers used NASA Task Load Index to examine subjective ratings of mental workload in 38 physically fit volunteers in air traffic management task simulations.

Wang et al. [84] provided a way for utilising HRV to correlate the human body's salivary response to stress. They used 176 ECG recordings and 264 salivary samples from 22 people. They have generated six datasets (3-amylase, 3-cortisol) using alpha-amylase and cortisol measurements to label ECG feature vectors. The final classifier system correctly classified salivary cortisol based on ECG characteristics with an accuracy of 80%, compared to 75% for salivary alpha-amylase. A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of RB stress prediction research is presented in Table 10.

Shallow Machine Learning Approaches

Sriramprakash et al. [88] used ECG, skin conductance, and Kinect 3D sensor to collect data from selected individuals. The SWELL-KW dataset was used for classification (149 features and 2688 instances in total) and got accuracies: 66.52% (KNN) vs 72.83% (SVM-RBF kernel).

Huang et al. [89] demonstrated that the mental fatigue of the samples could be accurately identified with a wearable ECG device. They collected 58 samples of ECG signals and compared SVM, NB, KNN, and LR algorithms to obtain accuracy (57.08% 9(SVM) vs 48.84% (NB) vs 65.37% (KNN) vs 59.71% (LR)) and area under the curve (AUC) (0.68 (SVM) vs 0.64 (NB) vs 0.74 (KNN) vs 0.65 (LR)). Wu et al. [90] combined HRV sensors and accelerometers to develop a model for monitoring the perceived stress levels in daily life. They collected data from 8 participants for their daily life in about 2 weeks and compared the performances of NB, J48, RF, and bagging algorithms where accuracy 0.730 (NB) vs 0.819 (J48) vs 0.832 (RF) vs 0.8392 (bagging) were obtained.

Sevil et al. [75] addressed the problem of detecting psychological stress (APS) using data collected from wristbands. They collected data from 34 samples doing 166 clinical experiments and compared different classification algorithms: KNN, SVM, DT, NB, EL, LD, and DL, where SVM had the highest accuracy of 99.1%.

Pourmohammadi and Maleki [91] collected EMG and ECG signals concurrently from 34 healthy students (23 females and 11 males, ages 20 to 37). They used LIBSVM (a library for SVM) with RBF (radial basis function) kernel for training the model. Sequentially, stress identification

Table 10 A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of rule-based stress prediction research

Ref.	Models	Dataset	Evaluation metrics	Performance
[85]	Fuzzy theoretic nonpara-metric deep model	Private/100 subjects	RMSE	PCC: 0.8162 (old dataset) vs 0.6809 (new dataset), RMSE: 6.8382 (old dataset) vs 9.4872 (new dataset)
[83]	Mamdani fuzzy model	Private/3 subjects for training and 2 subjects for testing	EI	EI = 2.9412 (Subject 1), 1 (Subject 2)
[86]	Sugeno neuro-fuzzy model	Private/20 subjects	-	Established a direct functional relationship between heart rate variability and mental stress
[87]	Sugeno fuzzy clustering	Private/26 males, 12 females, aged 18–29 years	-	Minimised the worst-case influence of uncertainty on fuzzy parameter identification performance
[84]	Fuzzy ARTMAP	Private/22 subjects	Acc	Acc = 80% (ECG characteristics),75% (salivary alpha-amylase)

In evaluation metrics column: PCC Pearson's correlation coefficient, *EI* evaluation index

In performance column: PCC Pearson's correlation coefficient, *EI* evaluation index

accuracy was 100%, 97.6%, and 96.2 % for the two, three, and four levels. Maldonado et al. [92] collected data from 50 engineering students in Chile, with a total of 33 men and 17 women aged 22.4 ± 2.8 years. They took HR, SpO2, and temperature readings to utilise in their SVM model, which yielded an AUC of 0.994 with a variable collecting cost of 16.

Pluntke et al. [93] acquired HRV data from subjects in a laboratory setting, and SVM and DT were used to train the model. A set of labelled RR-interval signals was collected as a training set. They used an H7 chest strap sensor to collect data from 26 male and female participants ranging in age from 23 to 59. A precision, recalling, and F-score of almost 90% were shown in the best model based on a DT of C5.

Giannakakis et al. [94] evaluated 24 participants and 11 tasks, performing a research protocol for about 45 min. They used KNN, generalised linear model (GLM), NB, linear discriminant analysis (LDA), SVM, and RF classifiers, where RF excels with a classification accuracy of 75.1% above any other classification method. 84.4% classification accuracy in a 10-fold method is the best result in the proposal of stress recognition simply by using hRV characteristics.

Castaldo et al. [95] used a 3-lead electrocardiogram (ECG) to collect data from 42 students on two distinct days, including during an oral examination (stress) and during rest following a holiday. They employed five distinct algorithms (NB, SVM, MLP, AB, and C4.5 (DT)). With sensitivity, specificity, and accuracy rates of 78%, 80%, and 79%, correspondingly, the C4.5 tree algorithm was the best ML technique for distinguishing between stress and rest.

Delmastro et al. [96] collected data conducting a randomised cross-over observational study where Zephyr Bio-Harness34 device was used for ECG monitoring and Shimmer3 GSR+Development Kit5 for EDA. Some algorithms (BN, SVM, *k*-NN, C4.5 DT, AB) were used where RF and AB learning schemes outperform the other classifier learning methods (accuracy: 87% for RF and 88.2% for AB).

Lima et al. [97] gathered information using some sensors (such as PPG, Spare, TVOC) from a group of willing participants (15 participants, ranging in age from 21 to 55 years old (9 females and 6 males)). While under stress, the model had an accuracy of about 80% in terms of HRV features in baseline and about 77 % in terms of HRV and EDA simultaneous baseline characteristics.

Yu et al. [98] used the ensemble learning technique to create a classifier that incorporates three separate work activities: body movement, typing, and browsing. These can be identified with 94.2%, 93.2%, and 91.2% accuracy, correspondingly. They gathered information from ten office workers, all of whom were around 31 years old.

Padmaja et al. [99] collected data from a smartphone and a Fitbit and then preprocessed and normalised it. They used NB (accuracy: 72%) and DT (accuracy: 62%) for classification. DetectStress has a 72% accuracy rate in recognising perceived stress utilising data from both smartphones and wireless fitness trackers.

Can et al. [100] collected physiological signal and questionnaire data from the 21 participants by using Samsung Gear S and S2 and Empatica E4 sensors. From HR and ACC signals acquired using Empatica E4, the MLP algorithm produced the best results (92.19%), while the RF algorithm produced the best classification accuracy (88.26%) with HR and ACC data collected from all devices.

Chen et al. [101] collected data from PPG and Polar H10 sensors, used RF as a classifier, and compared it with the SVM, Naïve Bayes, and MLP model. In the PPG dataset, their approach obtains an overall leave-one-participant-out F1-score of 80%, while the ground truth ECG scores 79.7%. Koldijk et al. [102] used the SWELL-KW dataset (149 features and 2688 instances in

total) and compared SVM (accuracy: 90.0298%) with 7 other algorithms, which includes NB (64.7693%), K-star (65.8110%), Bayes net (69.0848%), J48 (78.1994%), IBk (nearest neighbour with euclidean distance (84.5238%)), RF(87.0908%), and MLP (88.5417%).

Ciabattoni et al. [103] utilised KNN to classify stress using uniform precedence probability and Euclidean distance metrics with one neighbour. An accuracy of 84.5% has been determined altogether. In recognition of stress, a 26% misclassification error was detected when the individual was calm.

Attaran et al. [104] utilised the ThreatFire belt for data collection and employed several physiological and behavioural factors with both SVM and KNN classifiers to increase the detection accuracy. The best classification accuracy to identify stress was observed for the heart rate (HR) and accelerometer characteristics. For hardware implementation, the SVM classification was utilised, and this system has an overall classification accuracy of 96%.

Castaldo et al. [105] collected 23 ultra-short HRV features from 42 healthy subjects. They found six out of 23 ultra-short HRV features (MeanNN, StdNN, MeanHR, StdHR, HF, and SD2) displaying consistency in the detection of stress. The authors employed 5 ML algorithms and found their accuracies: MLP (98%) vs SVM (88%) vs C4.5 DT(94%) vs IBK (94%) vs LDA (94%).

Hantono et al. [106] recorded heart rate data using PPG sensors in smartphones from 41 subjects. They analysed the data and extracted HRV features to detect mental stress. The authors employed NN, KNN, DA, and NB algorithms to find the accuracies: NN (73%) vs KNN (82%) vs DA (66%) vs NB (60%).

Tiwari et al. [107] collected ECG and breathing data from 27 police trainees over the course of 15 weeks. They extracted ultra-short-term HRV and breathing features from the data and predicted stress. Results suggested that ultra-short-term analysis for stress prediction results in performance losses lower than 7% when compared to short-term analysis. They used an SVM classifier with RBF kernel, resulting in 80% performance accuracy.

Clark et al. [108] proposed a model for driver stress prediction. They collected data from 17 subjects using ECG, GSR, and respiration sensors after they completed a 20-mile drive. The authors extracted 42 features from the data to use in an RF classifier which achieved an average accuracy of 94%. Ahmad et al. [109] collected the dataset named Ryerson Multimedia Research Laboratory (RML), which was recorded by physiological signals using 9 participants and measured ECG, GSR, and respiration signals. They used raw data, which is procured from the ECG signal. For the proposed fusion model, they got 66.6% and 72.7% in the RML and WESAD datasets, respectively.

Dalmeida et al. [110] investigated the role of HRV features stress predicted from ECG, EMG, GSR, and respiration sensor data. They used a dataset collected by MIT and available in Physionet. They tested different ML models such as KNN, SVM, MLP, RF, and GB. MLP was considered an appropriate stress classification method with an 80% sensitivity score. HRV features such as the AVNN, SDNN, and RMSSD were found to be relevant aspects for stress identification.

Sandulescu et al. [111] present an SVM-based approach for stress prediction by collecting PPG, HRV, and EDA sensor data from 5 participants. The results showed 82% accuracy on two participants and more than 80% precision level for all the participants.

Munla et al. [112] intended to study stress-level detection from HRV features extracted from 16 different subjects from the Stress Recognition in Automobile Driver database (DRIVEDB). They used three ML models and achieved accuracies: KNN (66.66%) vs SVM (83.33%) and SVM with RBF kernel (83.3%).

A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of shallow ML-based stress prediction research is presented in Tables 11, 12, and 13.

Deep Machine Learning Approaches

de Vries et al. [113] collected GSR, RSP, and ECG sensor data from 61 participants from the age of 18 to 28 years to perform stress and relaxation classification. They used learning vector quantisation to achieve an accuracy of 88% for the classification.

Rastgoo et al. [115] collected ECG, vehicle, and environmental data from 27 participants in a vehicle simulator. They proposed a CNN and LSTM-based multimodal fusion model, which showed an accuracy of 92.8%, sensitivity of 94.13%, specificity of 97.37%, and precision of 95.00%.

Akbulut et al. [116] developed a stress model that incorporates an algorithm for detecting affective states based on HRV analysis, emotion recognition, and other statistical data. They collected the dataset conducted with 30 volunteers and named it CVDiMo. In categorising the stress levels of all patients, their suggested method had a 90.5% accuracy rate. The average success rate of MES patients was found to be 92%, which is greater than the general performance of healthy people.

Coutts et al. [117] recorded HRV features from 652 participants using a wearable sensor. They employed an LSTM network for the detection of stress, anxiety, and depression levels, finding 85% classification accuracy.

He et al. [118] used ECG sensor data from 20 participants to extract six HRV features (HR, LH, pQ, SD2, SDNN, Comb). They used SVM, LDA, and CNN-based models to detect cognitive stress from these models, where CNN

Table 11 A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of Shallow ML-based stress prediction research

Ref.	Models	Dataset	Evaluation metrics	Performance
[88]	SVM and KNN	SWELL-KW	Acc	Acc of 0.9275
[89]	SVM, KNN, NB, and LR	Private/35 participant (mean age of 23 ± 4 years and a male-to-female ratio of 1:1.3)	AUC	Acc = 0.755 AUC = 0.74
[90]	NB, J48, RF and bagging	Private/8 participants	Acc	prediction Acc = 0.857
[75]	KNN, SVM, DT, NB	Private/34 participants	Acc	Acc = 0.991
[91]	SVM	Private/34 students (23 females and 11 males, aged 20–37 years)	Acc	The accuracies two level = 1.0, three level = 0.976, and four levels were 0.962
[92]	SVM	Private/50 participants (33 men & 17 women)	AUC, variable collection cost	AUC = 0.994, Vcost = 16
[93]	SVM, C5 DT	Private	Pre, Rec f1, Acc	88% Acc, all
[94]	RF, SVM	Private/24 participants ageing 47.3±9.3 years	Acc	Acc = 0.844
[95]	NB, SVM, MLP, AB, Dt C4.5	Private/42 participants	Sen, Spe and Acc	Sen=0.78, Spe=0.80 and Acc rate=0.79
[96]	BN, SVM, k-NN, C4.5 DT, AB	Private/9 older adults	Acc; Pre; Rec; AUPRC	RF (Acc =87.0%; Pre =92.4%; Rec=88.2%; AUPRC =0.97) and AB (Acc=88.2%; Pre =92.3%; Rec=92.0%; AUPRC =0.92)

In evaluation metrics column: Acc accuracy, AUC area under the ROC curve, Vcost variable collection cost, Pre precision, Rec recall, Sen sensitivity, Spe specificity

In performance column: Acc accuracy, AUC area under the ROC curve, Vcost variable collection cost, Pre precision, Rec recall, Sen sensitivity, Spe specificity

Table 12 A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of Shallow ML-based stress prediction research

Ref.	Models	Dataset	Evaluation metrics	Performance
[97]	SVM, LR, RF	Private	Acc	80% accuracy for HRV features in baseline and about 77% for HRV and EDA simultaneous features
[98]	SVM, KNN, and EnL	15 office workers (five female, five males, age: 31 ± 5.3)	Acc	Accuracies of up to 91%
[99]	NB, DT	35 young adults	Sen, Spe, Acc, Pearson's correlation	NB classifier has 72% accuracy
[100]	PCA, SVM, KNN, LR, RF, MLP	21 participants (18 males and 3 females with an average age of 20)	Acc, f-Measure, Pre, Rec	Obtained 92.15% accuracy maximum three-level classification
[101]	RF	6 healthy participants ages 21–40 years old	Acc	10-fold accuracy of stress state is 98%, and F1-score reaches 80%
[102]	NB, SVM, KNN, BN, RF, DT, MLP	25 participants (8 female, average age 25, stdv 3.25)	Acc	Best results were obtained with an SVM (RBF kernel): 90.0298%
[103]	KNN	Private/10 young subjects (mean age 24; 5 female)	Acc, Error	Acc = 0.845, misclassification Error = 0.26
[104]	KNN, SVM	Private	Acc	The overall classification accuracy of this system is 96%

In evaluation metrics column: Acc accuracy, AUC area under the ROC curve, Vcost variable collection cost, Pre precision, Rec recall, Sen sensitivity, Spe specificity

In performance column: Acc accuracy, AUC area under the ROC curve, Vcost variable collection cost, Pre precision, Rec recall, Sen sensitivity, Spe specificity

Table 13 A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of Shallow ML-based stress prediction research

Ref.	Models	Dataset	Evaluation metrics	Performance
[105]	SVM, DT	ECG from 42 healthy subjects (19 female, 23 male)	AUC, Sen, Spe, Acc	Achieved good performance accuracy above 88%
[106]	NN, KNN, DT, NB	41 students	AUC, ROC, Acc, confusion matrix	-
[107]	SVM	Data was collected from 27 (6 females) police	BAC	In HRV segments BAC =0.579 for 300 s window duration
[108]	RFr	Private	Acc	Acc = 94%
[109]	RF, KNN, SVM, XGB	RML, WESAD	Acc., precision, recall, F1-Score	Acc. = 66.6 (RML dataset), Acc. = 72.7 (WESAD dataset)
[110]	KNN, SVM, MLP, RF, GB	PhysioNet	AUROC	MLP, RF and GB yielded an AUROC of 83%, 85%, and 85%, respectively
[111]	SVM	Private/5 participants aged 18 to 39	Acc, Pre	Best Acc and Pre for P3:83.08(Acc) & 83.87(Pre)
[112]	SVM-RBF, KNN	DRIVEDB	Acc	Acc = 83%

In evaluation metrics column: Acc accuracy, *AUROC* area under the ROC curve, *Vcost* variable collection cost, *Pre* precision, *Rec* recall, *Sen* sensitivity, *Spe* specificity, *BAC* balanced accuracy

(17.3%) outperformed LDA (25.1 ± 14.2%) and SVM (24.5 ± 13.2%) according to detection error rate.

Qin et al. [119] used 10 HRV features extracted from 56 samples of R-R intervals recorded during the modified Stroop test. They used 40 samples as training data and 16 as testing for a stress evaluation system based on the BP neural network, which could detect different levels of stress with an accuracy rate of 93.75%.

Ding et al. [120] recruited 18 healthy individuals to collect heart rate, heart rate variability, electromyography, electrodermal activity, and respiration physiological data to measure changes in physiological activity with varied levels of tasks. While combining physiological signals and task performance, their classification models could achieve accuracy at 96.4% but 78.3% when taking physiological features only.

Kalatzis et al. [121] recruited 57 participants to extract time- and frequency-domain features of HR and HRV using ECG sensors. They used an ANN-based model to classify stress and no-stress states, achieving a 90.83% accuracy level.

Qin et al. [119] used 10 HRV features extracted from 56 samples of R-R intervals recorded during the modified Stroop test. They used 40 samples as training data and 16 as testing for a stress evaluation system based on the BP neural network, which could detect different levels of stress with an accuracy rate of 93.75%.

Ding et al. [120] recruited 18 healthy individuals to collect heart rate, heart rate variability, electromyography, electrodermal activity, and respiration physiological data to measure changes in physiological activity with varied levels of tasks. While combining physiological signals and task performance, their classification models could achieve

accuracy at 96.4% but 78.3% when taking physiological features only.

Kalatzis et al. [121] recruited 57 participants to extract time- and frequency-domain features of HR and HRV using ECG sensors. They used an ANN-based model to classify stress and no-stress states, achieving a 90.83% accuracy level.

Dhaouadi and Ben Khelifa [122] used ECG, EDA, and EMG measures taken by wearable devices from 15 young gamers in order to stress monitoring in real time. They explored LSTM and DNN networks where the DNN model obtained the best accuracy of 65% at 15 and 30 epochs, but LSTm achieved the best accuracy of 95% at 30 epochs.

Stewart et al. [123] used two publicly available datasets, which include drivedb and WESAD. Data was collected from both datasets using multiple sensor recordings, including ECG and GSR. They used shallow ML models (such as KNN, SVM, and LR). Neural processes models outperformed those models (WESAD: 0.957 (average precision), drivedb: 0.804 (average precision)) and had the best performance when using periods of stress and baseline as context.

Silva et al. [124] monitored the stress of 83 medical students by comparing stress levels during academic exams and a regular week. Data was collected from wearable sensors such as Microsoft Smart band 2 and PPG. The neural network revealed better performance (model-1: sensitivity, 75.2%; specificity, 77.9%. Model-2: sensitivity, 74.2%; specificity, 78.1%.) where two models were established to predict stress comparing shallow ML algorithms (such as SVM, KNN, LR, RF). A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of deep ML-based stress prediction research is presented in Table 14.

Table 14 A summary of used algorithms, datasets, evaluation metrics, and obtained outcomes of deep ML-based stress prediction research

Ref.	Models	Dataset	Evaluation metrics	Performance
[113]	LVQ	Private/61 (20 male, 41 female)	Acc	RSP = 0.71, ECG = 0.834, Acc = 0.877
[115]	LSTM-RNN	Private/a group of Vietnamese students	Framework proposed	-
[114]	CNN-LSTM	Private/27 participants aged 21–40 years (55% male)	Acc Sen Spe: Pre	Acc: 0.928, Sen: 0.9413, Spe: 0.9737 and Pre: 0.95
[116]	FFNN	CVDiMo/conducted with 30 volunteers	AUC, Acc	AUC = 0.978, Acc= 0.92
[117]	LSTM	Private/for trial-1: 91 participants (62% female, 38% male); for trial-2: 600 (72% female, 28% male)	ACC	Acc = 0.85
[118]	CNN	Private/20 healthy subjects, aged from 18 to 35	ER, FAR, and FRR	CNN ER=17.3 FAR= 0.01 FRR = 32.1
[119]	NN	56 samples	Acc	93.75% accuracy
[120]	BPNN, SVM, KNN	18 right-handed, healthy individuals, 20.1 ± 0.94 years	Acc, Rec, Pre	Accuracy can reach 96.4% and 78.3%
[121]	ANN	Private/57 participants, and all are above 65 years	acc	Acc = 90.83%
[122]	LSTM, DNN	Private/15 gamers Age of 10 to 22	acc	Acc = 64% (DNN) vs 92% (LSTM)
[123]	LR, SVM, KNN	Drivedb, WESAD	Avg precision, AUC	WESAD: 0.957 average precision, same-participant vs 0.780 other-participant, drivedb: 0.804 vs 0.757
[124]	LR, NN, NB, SVM, RF, and KNN	Private/63 (76.8%) were female, and 19 (23.2%) were male aged 17 to 38 years	Sen, Spe	For Model 1, Sen = 0.752 and Spe = 0.779. For Model 2, Sen = 0.742 and Spe = 0.781

In evaluation metrics column: Acc accuracy, *AUC* area under the ROC curve, *Vcost* variable collection cost, *Pre* precision, *Rec* recall, *Sen* sensitivity, *Spe* specificity, *FAR* false acceptance rate, *FRR* false rejection rate

Discussion

Stress can lead to a variety of psychological issues. Many disorders are more likely to develop in a stressful environment, particularly if the stress is intense and long-lasting [125]. Therefore, being able to predict stress in an effective manner is a crucial fact. In this research, we observed HRV characteristics as physiological indicators for stress detection based on a review of 43 studies published between 2016 and 2021. RMSSD, SDNN, pNN50, and AVNN are determined to be the most often utilised HRV features in our tables. ECG, PPG, and GSR are the most deployed sensors for data collection.

In AI, accuracy is one of the most important performance indicators. The present research has been examined in this article in order to provide a full understanding of the field of stress prediction via HRV.

According to Fig. 8 displaying the performance comparison of the papers based on accuracy level, only one article by Wang et al. [84] employed accuracy as a performance measure for RB techniques. Using the fuzzy ARTMAP classifier, they explored the stress association between HRV and salivary, achieving an overall accuracy of 80% for ECG records.

In the case of shallow ML approaches, Sevil et al. [75] achieved the highest accuracy among the 21 studies utilising accuracy as a performance measure. They used wristband data

to quantify psychological stress and attained 99.1% accuracy using the SVM classifier, which is also the highest among all the publications reviewed in this review article. For deep ML techniques, Ding et al. [120] used a BPNN classifier to assess stress based on physiological activity with varying levels of tasks and achieved high accuracy. Their classification models have a 96.4% accuracy rate.

Another performance metric for assessing classification errors is the AUC. This review article contained 5 studies that employed the AUC measure, a two-dimensional area beneath the ROC curve. The highest AUC value for deep ML techniques was attained by Akbulut et al. [116], as shown in Fig. 9. They created a stress model based on HRV analysis, emotion recognition, and other statistical data from the CVDiMo dataset, which includes an algorithm for recognising affective states. Using FFNN, they were able to attain an AUC of 0.97. Maldonado et al. [92] used shallow ML to get the best AUC value of 0.99 for stress detection, which is significantly higher than other models that use AUC as a performance indicator.

Challanges and Future Scope

Due to a lack of quality data, data collection procedures, detection methodology selection, and other factors, research for predicting and detecting mental stress confront numerous
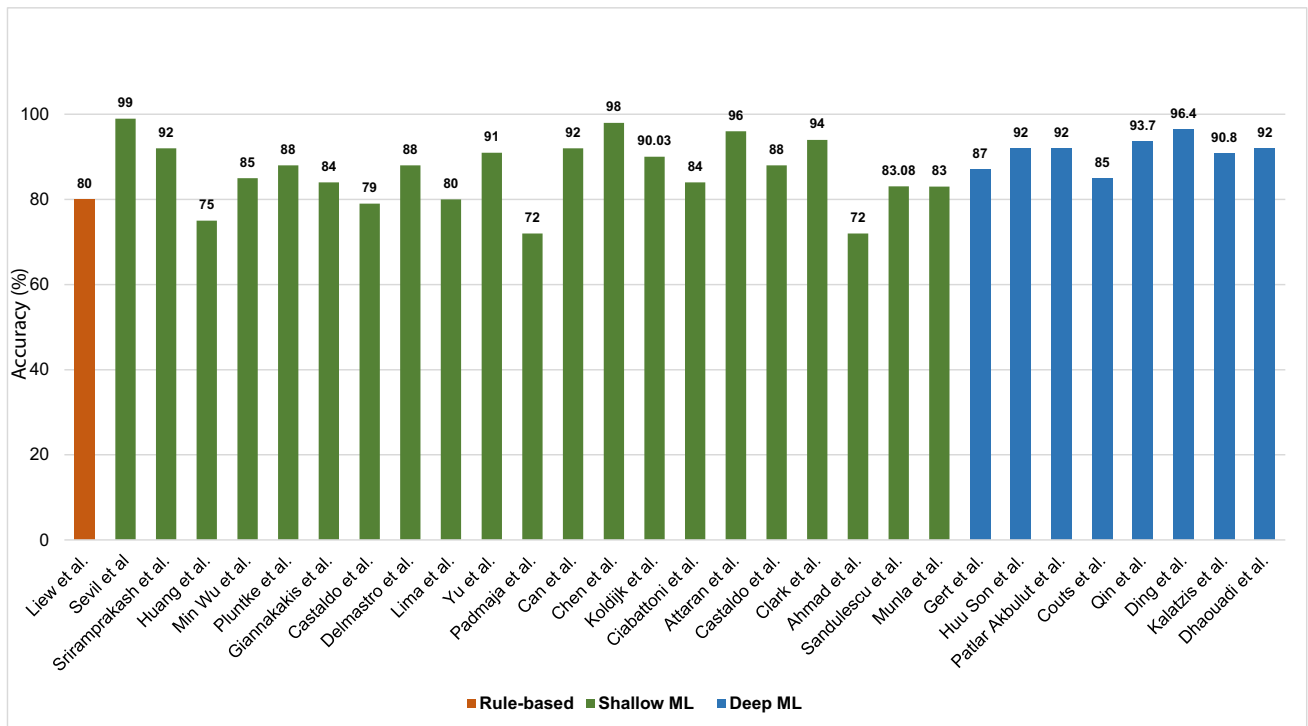
Fig. 8 Performance comparison of the articles based on accuracy level. The different algorithm types are presented using different colours

challenges. In this section, we will discuss the difficulties that stress researchers face and how to overcome them, which might be very useful for future researchers.

1. **Effect of individual moods and health**

   HRV is very much dependent on the change in ANS activity. In fact, HRV is controlled by ANS, a primitive part of the nervous system. As a result, individuals' native mood and health issues like blood sugar, hor-

mones, and blood pressure largely affect the measure of HRV. So, the consideration of baseline mood and health issues of the individual under observation needs to be considered during data collection.

2. **Controlled environment and biased dataset**

   Most of the datasets for stress prediction from HRV are collected inside a controlled environment inside the laboratory setup, and as a result, the effect of real-life scenarios is missing, which can be overcome by collecting data from

Fig. 9 Performance comparison of the articles based on AUC level. The different algorithm types are presented using different colours
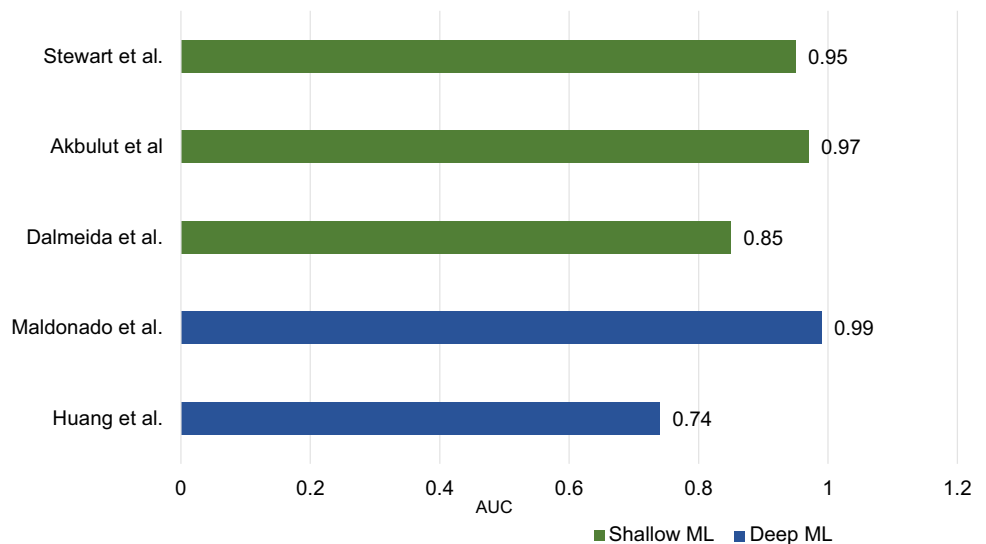
the real-world office or driving situations. Moreover, the dataset largely comprises male participants. As a result, the data are more biased towards male participants and can show poor performance in female-centric data.

3. **Lack of large benchmark dataset**

    Much research included in this article has used its own dataset, but most of which is not publicly available. But datasets that were publicly available were collected from a small number of participants. As a result, the data is not that generalised. So, there is not really a benchmark dataset that can be used for all AI approaches to make a performance comparison among them.

4. **Sensor quality and multimodal sensors**

    Data collection is the most important part of any research process. For HRV-based stress prediction, ECG, EMG, GSR, etc., sensors are used in different articles reviewed earlier, but the quality of sensors used and fusion of the right sensors are very important in this case. A multimodal dataset with data collected from high-quality and suitable sensors can produce a better and more fitting dataset for future research.

5. **Real-time stress monitoring**

    In real life, stress has been described in various ways, but it has been established that any stress leads to an unbalanced bodily and mental situation. This can lead to productivity loss, diminished work abilities, and a slew of other health issues. However, a real-time stress monitoring system is rarely investigated. As a result, a real-time stress monitoring system could be a promising future study topic.

6. **Fusion of hybrid architectures**

    Many ML and DL approaches have been used in the reviewed research in this article, but there have not been cases where hybrid architecture has been used to develop the stress detection or prediction model. Even though hybrid architectures can be a promising future prospect for accurate results.

7. **Exploration of HRV features**

    The majority of datasets utilised in recent studies have employed the same HRV features to identify stress in individuals, more or less. However, more noteworthy statistical, frequency-domain, and time-domain features could be investigated to provide effective stress prediction datasets.

8. **Less use of rule-based approaches**

    Throughout stress-related research, very rarely fuzzy and researchers have used other RB approaches to develop stress prediction systems. But due to human-like inference ability and understandability, RB approaches can be applied to develop a more suitable decision support system. So, fuzzy-based stress management and decision support systems can be a possible future research topic.

9. **Differences in evaluation metrics**

    Researchers have utilised a variety of metrics to demonstrate the performance of their stress prediction or detection system in various studies. As a result, new researchers and explorers in this sector are finding it increasingly difficult to compare these approaches to find one more appropriate. Setting a very acceptable and benchmark evaluation metric could be a solution to this issue.

Conclusion

Stress has become an inevitable element of our daily routines. It has resulted in an alarming scenario for adolescent and juvenile mental health throughout the world. Controlling stress has become a critical issue since it directly impacts physical and mental health. It has a negative influence on a country's socioeconomic condition. The growing AI discipline can provide effective solutions for stress prediction. In this study, we have conducted a comprehensive survey of the sensors employed for acquiring HRV data and their features, AI models applied on those data and their performance assessed using avialable evaluation metrics, pre-processing methods applied on multi-modal data, and existing datasets. The identified approaches have been summarised in tables and explored and their results were compared in depth. Stated outcomes of the methodologies, used datasets, and applied evaluation criteria were also presented.

Declarations

References

1. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Compu. 2006;44(12):1031–51.

2. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig. 2018;15(3):235.

3. Sztajzel J, et al. Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. Swiss Med Wkly. 2004;134(35–36):514–22.

4. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017:258.

5. Mohan PM, Nagarajan V, Das SR. Stress measurement from wearable photoplethysmographic sensor using heart rate variability data. In: 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE; 2016. p. 1141–4.

6. Oskooei A, Chau SM, Weiss J, Sridhar A, Martínez MR, Michel B. Destress: deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (HRV) data. In: Explainable AI in Healthcare and Medicine. Springer; 2021. p. 93–105.

7. Maxhuni A, Hernandez-Leal P, Sucar LE, Osmani V, Morales EF, Mayora O. Stress modelling and prediction in presence of scarce data. J Biomed Inform. 2016;63:344–56.

8. Bauer G, Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops. IEEE; 2012. p. 423–6.

9. Carneiro D, Castillo JC, Novais P, Fernández-Caballero A, Neves J. Multimodal behavioral analysis for non-invasive stress detection. Expert Syst Appl. 2012;39(18):13376–89.

10. Gaurav AR, Kumar V. EEG-metric based mental stress detection. Netw Biol. 2018;8(1):25–34.

11. Panicker SS, Gayathri P. A survey of machine learning techniques in physiology based mental stress detection systems. Biocybern Biomed Eng. 2019;39(2):444–69.

12. Piotrowski Z, Szypulska M. Classification of falling asleep states using HRV analysis. Biocybern Biomed Eng. 2017;37(2):290–301.

13. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform. 2019;92.

14. Bulagang AF, Weng NG, Mountstephens J, Teo J. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. Inform Med Unlocked. 2020;20:100363.

15. Pramanta SA, Prihatmanto AS, Park MG. A study on the stress identification using observed heart beat data. In: 2016 6th International Conference on System Engineering and Technology (ICSET). IEEE; 2016. p. 149–52.

16. Katarya R, Maan S. Stress detection using smartwatches with machine learning: a survey. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE; 2020. p. 306–10.

17. Nath RK, Thapliyal H, Caban-Holt A, Mohanty SP. Machine learning based solutions for real-time stress monitoring. IEEE Consum Electron Mag. 2020;9(5):34–41.

18. Smets E, Casale P, Großekathöfer U, Lamichhane B, De Raedt W, Bogaerts K, et al. Comparison of machine learning techniques for psychophysiological stress detection. In: International Symposium on Pervasive Computing Paradigms for Mental Health. Springer; 2015. p. 13–22.

19. Tonacci A, Dellabate A, Dieni A, Bachi L, Sansone F, Conte R, et al. Can machine learning predict stress reduction based on wearable sensors' data following relaxation at workplace? A pilot study. Processes. 2020;8(4):448.

20. Sharma N, Gedeon T. Objective measures, sensors and computational techniques for stress recognition and classification: a survey. Comput Methods Programs Biomed. 2012;108(3):1287–301.

21. Rahman MA. Gaussian process in computational biology: covariance functions for transcriptomics [phd]. University of Sheffield; 2018. Available from: https://etheses.whiterose.ac.uk/19460/.

22. Rakib AB, Rumky EA, Ashraf AJ, Hillas MM, Rahman MA. Mental healthcare chatbot using sequence-to-sequence learning and BiLSTM. In: Mahmud M, Kaiser MS, Vassanelli S, Dai Q, Zhong N, editors. Brain Informatics. Cham: Springer International Publishing; 2021. p. 378–87.

23. Islam N, et al. Towards machine learning based intrusion detection in IoT networks. Comput Mater Contin. 2021;69(2):1801–21.

24. Farhin F, Kaiser MS, Mahmud M. Secured smart healthcare system: blockchain and Bayesian inference based approach. In: Proc. TCCE. 2021. p. 455–65.

25. Ahmed S, et al. Artificial intelligence and machine learning for ensuring security in smart cities. In: Data-driven mining, learning and analytics for secured smart cities. Springer; 2021. p. 23–47.

26. Zaman S, et al. Security threats and artificial intelligence based countermeasures for Internet of Things networks: a comprehensive survey. IEEE Access. 2021;9:94668–90.

27. Noor MBT, Zenia NZ, Kaiser MS, Mamun SA, Mahmud M. Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. Brain Inform. 2020;7(1):1–21.

28. Ghosh T, Al Banna MH, Rahman MS, Kaiser MS, Mahmud M, Hosen AS, et al. Artificial intelligence and Internet of Things in screening and management of autism spectrum disorder. Sustain Cities Soc. 2021;74.

29. Biswas M, Kaiser MS, Mahmud M, Al Mamun S, Hossain M, Rahman MA, et al. An XAI based autism detection: The context behind the detection. In: Proc. Brain Informatics. 2021. p. 448–59.

30. Wadhera T, Mahmud M. Computing hierarchical complexity of the brain from electroencephalogram signals: a graph convolutional network-based approach. In: Proc. IJCNN. 2022. p. 1–6.

31. Wadhera T, Mahmud M. Influences of social learning in individual perception and decision making in people with autism: a computational approach. In: Proc. Brain Inform. 2022. p. 50–61.

32. Wadhera T, Mahmud M. Brain networks in autism spectrum disorder, epilepsy and their relationship: a machine learning approach. In: Artificial Intelligence in Healthcare: Recent Applications and Developments. Springer; 2022. p. 125–42.

33. Wadhera T, Mahmud M. Brain functional network topology in autism spectrum disorder: a novel weighted hierarchical complexity metric for electroencephalogram. IEEE J Biomed Health Inform. 2023:1–8.

34. Sumi AI, et al. fASSERT: a fuzzy assistive system for children with autism using Internet of Things. In: Proc. Brain Inform. 2018. p. 403–12.

35. Akhund NU, et al. ADEPTNESS: Alzheimer's disease patient management system using pervasive sensors-early prototype and preliminary results. In: Proc. Brain Inform. 2018. p. 413–22.

36. Al Banna M, Ghosh T, Taher KA, Kaiser MS, Mahmud M, et al. A monitoring system for patients of autism spectrum disorder using artificial intelligence. In: Proc. Brain Informatics; 2020. p. 251–62.

37. Jesmin S, Kaiser MS, Mahmud M. Artificial and Internet of Healthcare Things based Alzheimer care during COVID 19. In: Proc. Brain Inform.; 2020. p. 263–74.

38. Ahmed S, Hossain M, Nur SB, Shamim Kaiser M, Mahmud M, et al. Toward machine learning-based psychological assessment of autism spectrum disorders in school and community. In: Proc. TEHI; 2022. p. 139–49.

39. Mahmud M, Kaiser MS, Rahman MA, Wadhera T, Brown DJ, Shopland N, et al. Towards explainable and privacy-preserving artificial intelligence for personalisation in autism spectrum disorder. In: Universal Access in Human-Computer Interaction. User and Context Diversity: 16th International Conference, UAHCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II. Springer; 2022. p. 356–70.

40. Nahiduzzaman M, et al. Machine learning based early fall detection for elderly people with neurological disorder using multimodal data fusion. In: Proc. Brain Inform.; 2020. p. 204–14.

41. Biswas M, et al. Indoor navigation support system for patients with neurodegenerative diseases. In: Proc. Brain Inform.; 2021. p. 411–22.

42. Sadik R, Reza ML, Al Noman A, Al Mamun S, Kaiser MS, Rahman MA. COVID-19 pandemic: a comparative prediction using machine learning. International Journal of Automation, Artificial Intelligence and Machine Learning. 2020;1(1):1–16.

43. Mahmud M, Kaiser MS. Machine learning in fighting pandemics: a COVID-19 case study. In: COVID-19: prediction, decision-making, and its impacts. Springer; 2021. p. 77–81.

44. Kumar S, et al. Forecasting major impacts of COVID-19 pandemic on country-driven sectors: challenges, lessons, and future roadmap. Pers Ubiquitous Comput. 2021:1–24.

45. Bhapkar HR, et al. Rough sets in COVID-19 to predict symptomatic cases. In: COVID-19: Prediction, Decision-Making, and its Impacts. Springer; 2021. p. 57–68.

46. Satu MS, et al. Short-term prediction of COVID-19 cases using machine learning models. Appl Sci. 2021;11(9):4266.

47. Prakash N, et al. Deep transfer learning for COVID-19 detection and infection localization with superpixel based segmentation. Sustain Cities Soc. 2021;75: 103252.

48. AlArjani A, et al. Application of mathematical modeling in prediction of COVID-19 transmission dynamics. Arab J Sci Eng. 2022:1–24.

49. Paul A, et al. Inverted bell-curve-based ensemble of deep learning models for detection of COVID-19 from chest X-rays. Neural Comput Appl. 2022:1–15.

50. Mahmud M, Kaiser MS, Rahman MM, Rahman MA, Shabut A, Al-Mamun S, et al. A brain-inspired trust management model to assure security in a cloud based IoT framework for neuroscience applications. Cogn Comput. 2018;10(5):864–73.

51. Mahmud M, Kaiser MS, Hussain A, Vassanelli S. Applications of deep learning and reinforcement learning to biological data. IEEE Trans Neural Netw Learn Syst. 2018;29(6):2063–79.

52. Mahmud M, Kaiser MS, McGinnity TM, Hussain A. Deep learning in mining biological data. Cogn Comput. 2021;13(1):1–33.

53. Nasrin F, Ahmed NI, Rahman MA. Auditory attention state decoding for the quiet and hypothetical environment: a comparison between bLSTM and SVM. In: Kaiser MS, Bandyopadhyay A, Mahmud M, Ray K, editors. Proceedings of TCCE. Advances in Intelligent Systems and Computing. Singapore: Springer; 2021. p. 291–301.

54. Rahman MA, Brown DJ, Mahmud M, Shopland N, Haym N, Sumich A, et al. Biofeedback towards machine learning driven self-guided virtual reality exposure therapy based on arousal state detection from multimodal data. In: Proc. BI2022; 2022. p. 1–12.

55. Farhin F, Kaiser MS, Mahmud M. Towards secured service provisioning for the Internet of Healthcare Things. In: Proc. AICT; 2020. p. 1–6.

56. Kaiser MS, et al. 6G access network for intelligent Internet of Healthcare Things: opportunity, challenges, and research directions. In: Proc. TCCE; 2021. p. 317–28.

57. Biswas M, et al. ACCU3RATE: a mobile health application rating scale based on user reviews. PLoS ONE. 2021;16(12).

58. Rabby G, et al. A flexible keyphrase extraction technique for academic literature. Procedia Comput Sci. 2018;135:553–63.

59. Rabby G, Azad S, Mahmud M, Zamli KZ, Rahman MM. TeKET: a tree-based unsupervised keyphrase extraction technique. Cogn Comput. 2020;12(4):811–33.

60. Adiba FI, Islam T, Kaiser MS, Mahmud M, Rahman MA. Effect of corpora on classification of fake news using naive Bayes classifier. International Journal of Automation, Artificial Intelligence and Machine Learning. 2020 Oct;1(1):80-92. Number: 1. Available from: https://researchlakejournals.com/index.php/AAIML/article/view/45.

61. Das S, Yasmin MR, Arefin M, Taher KA, Uddin MN, Rahman MA. Mixed Bangla-English spoken digit classification using convolutional neural network. In: Kaiser MS, Kasabov N, Iftekharuddin K, Zhong N, editors. Mahmud M. Applied Intelligence and Informatics. Communications in Computer and Information Science. Cham: Springer International Publishing; 2021. p. 371–83.

62. Nawar A, Toma NT, Al Mamun S, Kaiser MS, Mahmud M, Rahman MA. Cross-content recommendation between movie and book using machine learning. In: 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT); 2021. p. 1–6.

63. Rahman MA, Brown DJ, Shopland N, Burton A, Mahmud M. Explainable multimodal machine learning for engagement analysis by continuous performance test. In: Stephanidis C, Antona M, editors. Universal Access in Human-Computer Interaction. User and Context Diversity. Lecture Notes in Computer Science. Cham: Springer International Publishing; 2022. p. 386–99.

64. Rahman MA, Brown DJ, Shopland N, Harris MC, Turabee ZB, Heym N, et al. Towards machine learning driven self-guided virtual reality exposure therapy based on arousal state detection from multimodal data. In: Mahmud M, He J, Vassanelli S, van Zundert A, Zhong N, et al., editors. Brain Informatics. Cham: Springer International Publishing; 2022. p. 195–209.

65. Mahmud M, Kaiser MS, Rahman MA. Towards explainable and privacy-preserving artificial intelligence for personalisation in autism spectrum disorder. In: Stephanidis C, Antona M, editors. Universal Access in Human-Computer Interaction. User and Context Diversity. Lecture Notes in Computer Science. Cham: Springer International Publishing; 2022. p. 356–70.

66. Tasnim N, Al Mamun S, Shahidul Islam M, Kaiser MS, Mahmud M. Explainable mortality prediction model for congestive heart failure with nature-based feature selection method. Appl Sci. 2023;13(10):6138.

67. Banerjee JS, Mahmud M, Brown D. Heart rate variability-based mental stress detection: an explainable machine learning approach. SN Comput Sci. 2023;4(2):176.

68. Vimbi V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen F. Application of explainable artificial intelligence in Alzheimer's

disease classification: a systematic review. Cogn. Comput. 2023:1–27. [ePub Ahead of Print].

69. Banerjee JS, Chakraborty A, Mahmud M, Kar U, Lahby M, Saha G. Explainable artificial intelligence (XAI) based analysis of stress among tech workers amidst COVID-19 pandemic. In: Advanced AI and Internet of Health Things for Combating Pandemics. Springer; 2023. p. 151–74.

70. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cogn. Comput. 2023:1–30. [ePub Ahead of Print.].

71. Lohani A, Kumar R, Singh R. Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. J Hydrol. 2012;442:23–35.

72. Kuo R, Hong S, Huang Y. Integration of particle swarm optimization-based fuzzy neural network and artificial neural network for supplier selection. Appl Math Model. 2010;34(12):3976–90.

73. Dimitoglou G, Adams JA, Jim CM. Comparison of the C4. 5 and a Naïve Bayes classifier for the prediction of lung cancer survivability. arXiv:1206.1121 [Preprint]. 2012.

74. Ray S. A quick review of machine learning algorithms. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE; 2019. p. 35–9.

75. Sevil M, Rashid M, Hajizadeh I, Askari MR, Hobbs N, Brandt R, et al. Discrimination of simultaneous psychological and physical stressors using wristband biosignals. Comput Methods Programs Biomed. 2021;199: 105898.

76. Tang C, Chen F, Li X. Perceptron implementation of triple-valued logic operations. IEEE Trans Circuits Syst II Express Briefs. 2011;58(9):590–4.

77. Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. In: Journal of Physics: Conference Series, vol. 1142. IOP Publishing; 2018. p. 012012.

78. Liu Y. Random forest algorithm in big data environment. Computer Modelling & New Technologies. 2014;18(12A):147–51.

79. Pascanu R, Gulcehre C, Cho K, Bengio Y. How to construct deep recurrent neural networks. arXiv:1312.6026 [Preprint]. 2013.

80. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.

81. LeCun Y, Bengio Y, et al. Convolutional networks for images, speech, and time series. The Handbook of Brain Theory and Neural Networks. 1995;3361(10):1995.

82. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE Access. 2019;7:53040–65.

83. El-Samahy E, Mahfouf M, Torres-Salomao L, Anzurez-Marin J. A new computer control system for mental stress management using fuzzy logic. In: 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE; 2015. p. 1–7.

84. Wang J, Belatreche A, Maguire LP, McGinnity TM. SpikeTemp: an enhanced rank-order-based learning approach for spiking neural networks with adaptive structure. IEEE Trans Neural Netw Learn Syst. 2015;28(1):30–43.

85. Kumar M, Zhang W, Weippert M, Freudenthaler B. An explainable fuzzy theoretic nonparametric deep model for stress assessment using heartbeat intervals analysis. IEEE Trans Fuzzy Syst. 2020.

86. Ranganathan G, Rangarajan R, Bindhu V. Estimation of heart rate signals for mental stress assessment using neuro fuzzy technique. Appl Soft Comput. 2012;12(8):1978–84.

87. Kumar M, Weippert M, Vilbrandt R, Kreuzfeld S, Stoll R. Fuzzy evaluation of heart rate signals for mental stress assessment. IEEE Trans Fuzzy Syst. 2007;15(5):791–808.

88. Sriramprakash S, Prasanna VD, Murthy OR. Stress detection in working people. Procedia Comput Sci. 2017;115:359–66.

89. Huang S, Li J, Zhang P, Zhang W. Detection of mental fatigue state with wearable ECG devices. Int J Med Informatics. 2018;119:39–46.

90. Wu M, Cao H, Nguyen HL, Surmacz K, Hargrove C. Modeling perceived stress via HRV and accelerometer sensor streams. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015. p. 1625–8.

91. Pourmohammadi S, Maleki A. Stress detection using ECG and EMG signals: a comprehensive study. Comput Methods Programs Biomed. 2020;193.

92. Maldonado S, López J, Jimenez-Molina A, Lira H. Simultaneous feature selection and heterogeneity control for SVM classification: an application to mental workload assessment. Expert Syst Appl. 2020;143.

93. Pluntke U, Gerke S, Sridhar A, Weiss J, Michel B. Evaluation and classification of physical and psychological stress in firefighters using heart rate variability. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019. p. 2207–12.

94. Giannakakis G, Marias K, Tsiknakis M. A stress recognition system using HRV parameters and machine learning techniques. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). IEEE; 2019. p. 269–72.

95. Castaldo R, Xu W, Melillo P, Pecchia L, Santamaria L, James C. Detection of mental stress due to oral academic examination via ultra-short-term HRV analysis. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2016. p. 3805–8.

96. Delmastro F, Di Martino F, Dolciotti C. Cognitive training and stress detection in MCI frail older people through wearable sensors and machine learning. IEEE Access. 2020;8:65573–90.

97. Lima R, de Noronha Osório DF, Gamboa H. Heart rate variability and electrodermal activity in mental stress aloud: predicting the outcome. In: Biosignals. 2019. p. 42–51.

98. Yu B, Zhang B, An P, Xu L, Xue M, Hu J. An unobtrusive stress recognition system for the smart office. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019. p. 1326–9.

99. Padmaja B, Prasad VR, Sunitha K, Reddy NCS, Anil C. Detect-Stress: a novel stress detection system based on smartphone and wireless physical activity tracker. In: First international conference on artificial intelligence and cognitive computing. Springer; 2019. p. 67–80.

100. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors. 2019;19(8):1849.

101. Chen C, Li C, Tsai CW, Deng X. Evaluation of mental stress and heart rate variability derived from wrist-based photoplethysmography. In: 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). IEEE; 2019. p. 65–8.

102. Koldijk S, Neerincx MA, Kraaij W. Detecting work stress in offices by combining unobtrusive sensors. IEEE Trans Affect Comput. 2016;9(2):227–39.

103. Ciabattoni L, Ferracuti F, Longhi S, Pepa L, Romeo L, Verdini F. Real-time mental stress detection based on smartwatch. In: 2017 IEEE International Conference on Consumer Electronics (ICCE). IEEE; 2017. p. 110–1.

104. Attaran N, Brooks J, Mohsenin T. A low-power multi-physiological monitoring processor for stress detection. In: 2016 IEEE Sensors. IEEE; 2016. p. 1–3.

105. Castaldo R, Montesinos L, Melillo P, James C, Pecchia L. Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life. BMC Med Inform Decis Mak. 2019;19(1):1–13.

106. Hantono BS, Nugroho LE, Santosa PI. Mental stress detection via heart rate variability using machine learning. Int J Electr Eng Inform. 2020;12(3):431–44.

107. Tiwari A, Cassani R, Gagnon JF, Lafond D, Tremblay S, Falk TH. Prediction of stress and mental workload during police academy training using ultra-short-term heart rate variability and breathing analysis. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020. p. 4530–3.

108. Clark J, Nath RK, Thapliyal H. Machine learning based prediction of future stress events in a driving scenario. arXiv:2106.07542 [Preprint]. 2021.

109. Ahmad Z, Rabbani S, Zafar MR, Ishaque S, Krishnan S, Khan N. Multi-level stress assessment from ECG in a virtual reality environment using multimodal fusion. arXiv:2107.04566 [Preprint]. 2021.

110. Dalmeida KM, Masala GL. HRV features as viable physiological markers for stress detection using wearable devices. Sensors. 2021;21(8):2873.

111. Sandulescu V, Andrews S, Ellis D, Bellotto N, Mozos OM. Stress detection using wearable physiological sensors. In: International work-conference on the interplay between natural and artificial computation. Springer; 2015. p. 526–32.

112. Munla N, Khalil M, Shahin A, Mourad A. Driver stress level detection using HRV analysis. In: 2015 International Conference on Advances in Biomedical Engineering (ICABME). IEEE; 2015. p. 61–4.

113. de Vries JJG, Pauws SC, Biehl M. Insightful stress detection from physiology modalities using learning vector quantization. Neurocomputing. 2015;151:873–82.

114. Son HH. Toward a proposed framework for mood recognition using LSTM recurrent neuron network. Procedia Computer Science. 2017;109:1028–34.

115. Rastgoo MN, Nakisa B, Maire F, Rakotonirainy A, Chandran V. Automatic driver stress level classification using multimodal deep learning. Expert Syst Appl. 2019;138: 112793.

116. Akbulut FP, Ikitimur B, Akan A. Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome. Artif Intell Med. 2020;104: 101824.

117. Coutts LV, Plans D, Brown AW, Collomosse J. Deep learning with wearable based heart rate variability for prediction of mental and general health. J Biomed Inform. 2020;112: 103610.

118. He J, Li K, Liao X, Zhang P, Jiang N. Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal. IEEE Access. 2019;7:42710–7.

119. Qin Z, Li M, Huang L, Zhao Y. Stress level evaluation using BP neural network based on time-frequency analysis of HRV. In: 2017 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE; 2017. p. 1798–803.

120. Ding Y, Cao Y, Duffy VG, Wang Y, Zhang X. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. Ergonomics. 2020;63(7):896–908.

121. Kalatzis A, Stanley L, Karthikeyan R, Mehta RK. Mental stress classification during a motor task in older adults using an artificial neural network. In: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers; 2020. p. 244–8.

122. Dhaouadi S, Ben Khelifa MM. A multimodal physiological-based stress recognition: deep Learning models' evaluation in gamers' monitoring application. In: 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE; 2020. p. 1–6.

123. Stewart CL, Folarin A, Dobson R. Personalized acute stress classification from physiological signals with neural processes. arXiv:2002.04176 [Preprint]. 2020.

124. Silva E, Aguiar J, Reis LP, e Sá JO, Gonçalves J, Carvalho V. Stress among Portuguese medical students: the EuStress solution. J Med Syst. 2020;44(2):1–6.

125. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. EXCLI J. 2017;16:1057.