

Received April 20, 2021, accepted May 16, 2021, date of publication May 20, 2021, date of current version July 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082423

A Sensitivity Analysis of Biophysiological Responses of Stress for Wearable Sensors in Connected Health

TALHA IQBAL¹, PAU REDON-LURBE^{1,2}, ANDREW J. SIMPKIN³, ADNAN ELAHI⁴, SANDRA GANLY¹, WILLIAM WIJNS¹, AND ATIF SHAHZAD^{1,5}

¹Smart Sensor Laboratory, School of Medicine, National University of Ireland Galway, Galway, H91 TK33 Ireland

²CÚRAM Center for Research in Medical Devices, Galway, H91 W2TY Ireland

³School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, H91 TK33 Ireland

⁴Electrical and Electronic Engineering, National University of Ireland Galway, Galway, H91 TK33 Ireland

⁵Centre for Systems Modelling and Quantitative Biomedicine, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham B15 2TT, U.K.

Corresponding author: Talha Iqbal (t.iqbal1@nuigalway.ie)

This work was supported in part by the Science Foundation Ireland Research Professorship Award under Grant 15/RP/2765, and in part by the Irish Research Council Postdoctoral Fellowship under Grant GOIPD/2018/644.

This work did not involve human subjects or animals in its research.

ABSTRACT Stress is known as a silent killer that contributes to several life-threatening health conditions such as high blood pressure, heart disease, and diabetes. The current standard for stress evaluation is based on self-reported questionnaires and standardized stress scores. There is no gold standard to independently evaluate stress levels despite the availability of numerous biophysiological stress indicators. With an increasing interest in wearable health monitoring in recent years, several studies have explored the potential of various biophysiological indicators of stress for this purpose. However, there is no clear understanding of the relative sensitivity and specificity of these stress-related biophysiological indicators of stress in the literature. Hence this study aims to perform statistical analysis and classification modelling of biophysiological data gathered from healthy individuals, undergoing various induced emotional states, and to assess the relative sensitivity and specificity of common biophysiological indicators of stress. In this paper, several frequently used key indicators of stress, such as heart rate, respiratory rate, skin conductance, RR interval, heart rate variability in the electrocardiogram, and muscle activation measured by electromyography, are evaluated based on a detailed statistical analysis of the data gathered from an already existing, publicly available WESAD (Wearable Stress and Affect Detection) dataset. Respiratory rate and heart rate were the two best features for distinguishing between stressed and unstressed states.

INDEX TERMS Stress monitoring, biophysiological stress response, sensitivity analysis, heart rate, respiratory rate, skin conduction, electrocardiogram, electromyograph.

I. INTRODUCTION

It is well understood that every human being is exposed to some level of stress more than once in their lifetime. Stress can be defined as a non-specific response of our body to meet a certain demand in extreme conditions [1]. According to the British Health and Safety Executive (HSE), 44% of all work-related illnesses in 2017/18 was due to stress [2]. It has been seen that stress generally has negative effects on the mental health and well-being of a person [3]. Acute stressors (stimuli that cause stress) may not impose any health burden on young and healthy people having an adaptive and good

coping response, but if the stressors are too persistent or too strong, these stressors may lead to depression and anxiety [4]. Chronic stress is known to contribute to life-threatening conditions such as heart disease, high blood pressure, diabetes, and obesity, and an acute episode of stress can trigger a heart attack or stroke by causing arterial inflammation [5].

The current standard for clinical evaluation of stress is based on self-reported questionnaires or standardized stress scores, such as the Perceived Stress Scale (PSS) [1]. However, with the recent development in wearable biosensor technologies, a huge interest has been seen in measuring biophysiological responses of stress for the evaluation and monitoring of stress. To develop a reliable device for stress monitoring, it is important to understand how stress affects the human

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi¹.

body from a physiological and biochemical point of view. Under the influence of a stressor, the stress triggers the sympathetic nervous system, causing the release of various hormones such as adrenaline or cortisol [6], [7]. The release of these hormones leads to changes in heart rate, respiratory rate, and causes muscle tension among other physiological responses. These changes in the body prepare the individual for a physical fight or flight reaction. The changes caused in both the biochemical and physiological state of the human body in response to stress can be observed and used as an indicator of stress. Physiological indicators are of particular interest due to the possibility of measuring these indicators non-invasively.

The wearable sensor technology has progressed to the level that several physiological parameters can be measured continuously as well as wirelessly. Some real-time stress detecting models have been described [8]–[10]. Even with all these advances, no clinically validated stress model can be considered suitable for clinical use to monitor stress in real-time (in the natural environment). The lack of a reliable stress model can be explained by several key challenges that need to be addressed before the development and clinical validation of a model. Firstly, there is no universally accepted definition of stress. Secondly, there is a lack of corresponding gold standard ground truth values or real data which can be somewhat associated with the unavailability of a standard definition of stress. For example, cortisol is considered as a stress hormone in some studies and a self-reported questionnaire is used as ground truth to assess the stress level in the field. However, the correlation between cortisol (a stress indicator) and self-reporting questionnaires is reported in the range of 0.26 to 0.36 [9], [10]. This poor correlation can be associated with several factors including poor reporting in the questionnaires, which makes it difficult to assess the prediction accuracy of cortisol.

The third challenge is the collection of physiological data. The collection of stress parameters in the natural environment is very difficult, especially in the presence of various sources of error and noise [11]. For example, to collect day-long electrocardiogram (ECG) data, ECG sensor electrodes are connected to the subject's body for a day. The adhesion of the ECG electrodes may degrade over time as the day goes by, thus producing a lot of noisy readings. The physical movement of the subject may also cause noisy spikes in the data due to changes in the contact of the electrodes with the body. Moreover, data could also be lost during wireless transmission.

The fourth challenge is dealing with the confounding variables. Physiological stimulation that is indicative of the subject's stress can easily be obfuscated by changes in posture, movement of limbs, or any other physical activities. Separation of good quality signal for analysis of stress is therefore significantly challenging.

The fifth challenge is the identification and calculation of discriminative features that are specific and can be easily

distinguished as a stress response from other comparable physiological stimuli.

The final challenge is the development of classifiers using computed features, and training and validation of the model for field usage. This is the most difficult challenge to tackle due to the lack of a gold standard dataset that can be used to train and validate the model. If self-reported questionnaires are used as a label for stress, then for consistency analysis the threshold is set to be 0.7 for declaration of a concordance [12], which also reflects inherent biases and variabilities in the self-reported data.

Many sensor-based stress monitoring devices and research studies exploited the relationship between stress and resulting physiological variations [13]–[17]. These include machine learning techniques to detect stress from physiological and activity data collected from respiration (RESP), electrocardiogram (ECG) and accelerometer (ACC) sensors [18], [19]. Other less frequently used indicators, blood volume pulse (BVP), skin temperature (TEMP), electromyography (EMG), photoplethysmogram (PPG) and electrodermal activity (EDA), have also been recorded and used for stress monitoring [20], [21]. These physiological indicators are not specific to stress response, therefore, the stress prediction based on the physiological indicators may have varying accuracy, be it for any individual indicator or a combination of these indicators.

Besides the physiological indicators, several biochemical indicators are also used for stress detection. In humans, these indicators include the level of cortisol, adrenaline, alpha-amylase, copeptin and prolactin [4], [22], [23]. Among these indicators, cortisol is considered the primary stress hormone [24]. Several techniques have been proposed to measure cortisol level in saliva, sweat and hair [25]–[27]. This paper is intended to provide a sensitivity analysis of biophysiological indicators of stress rather than biochemical ones. A comprehensive review on biochemical stress indicators can be found elsewhere [28].

A. RELATED WORK

Han *et al.* [21] proposed a stress detection technique that detects three levels of stress i.e. no stress, moderate stress, and high perceived stress using ECG and PPG signals. The authors collected data from 39 subjects and reported classification accuracy of 84% using a random forest and support vector machine (SVM) classifier for three-stage classification of stress. For binary classification i.e. rest and stress, an accuracy of 94% was achieved. Choi *et al.* [29] proposed a wearable device to measure the stress, drowsiness, and fatigue of the vehicle drivers. Stress indicators they measured were (Galvanic Skin Response) GSR, activity data from accelerometer, skin temperature, and PPG signals of 28 drivers. The authors reported an accuracy of 68.3% for four classes i.e. normal, stressed, drowsiness, and fatigue and an accuracy of 84.5% for three classes i.e. normal, stressed, drowsiness or fatigue classification. Mohino-Herranz *et al.* [30] assessed the mental fitness of different subjects. They used ECG and Thoracic Electrical

Bioimpedance (TEB) signals to monitor the stress of 40 subjects. The authors achieved error rates of 21.2%, 32.3%, and 4.8% for activity identification, mental activity, and emotional state, respectively. Liu *et al.* [31] determined the feasibility of the EDA signal parameter and developed a stress monitoring device. The authors used only EDA signals for the detection of the stress of 11 drivers. After computing Fisher projection and Linear Discriminant Analysis (LDA), the authors reported a classification accuracy of 81.8% by only using EDA.

Lee *et al.* [32] and Healey *et al.* [33] wanted to develop a wearable glove that could detect the stress of drivers and collected data from 28 and 10 drivers, respectively. The authors of both studies recorded PPG signals for analysis. Lee *et al.* achieved a classification accuracy of 95% while Healey *et al.* reported an accuracy of 62.2% using only PPG signal (respiratory rate) recorded through sensors on driver's gloves. Similarly, Wijsman *et al.* [34] and Healey *et al.* [33] developed an algorithm for detecting mental stress using physiological signals. These signals included ECG, PPG, EMG, and EDA from 18 and 10 subjects, respectively. Wijsman *et al.* claimed to reach 80% accuracy for two-class (i.e. rest and stress) classification and concluded that the accuracy indicates that these features are suitable to be used for an individual's stress detection while Healey *et al.* reported classification accuracy of 86.6% using the same physiological signals.

Chen *et al.* [35], Shi *et al.* [36] and Kim *et al.* [37] developed a stress detection system based on multimodal features and kernel-based classifiers using ECG, EDA and PPG signals. The studies collected data from 14, 22, and 175 subjects, respectively. Chen *et al.* analysed the data in terms of precision, sensitivity and specificity. While using a full feature set, SVM with a linear kernel gave the highest inter-drive classification precision. For the cross-driver analysis, SVM with radial basis function (RBF) kernel gave a precision score of 89.7%. Shi *et al.* concluded that the SVM based model detected stress with high precision and recall rate and classification accuracy of 68%. Kim *et al.* reported that they achieved a classification accuracy of 78.4% for three emotional states classification problems and an accuracy of 61.8% for four-state classification problems. Sun *et al.* [38] determined mental as well as the physical stress of 20 subjects during different physical activities. The authors used ECG, EDA, and accelerometer signals. They reported a classification accuracy of 92.4% using accelerometer data along with ECG and EDA physiological signals. The inter-subject classification accuracy was reported to be 80.9%.

Mozos *et al.* [39] and Sandulescu *et al.* [40] presented a stress detection methodology for people who suffer from stress in social situations. Both studies used EDA and PPG signals for stress detection collected from 5 and 18 subjects, respectively. After experimentation, Mozos *et al.* reported an accuracy of 92 % with the SVM (RBF kernel) classifier, in comparison to Linear kernel SVM (80%), AdaBoost (67%) and k-nearest neighbours (KNN) (62%),

when using a selected set of features. Sandulescu *et al.* were successful in classifying the stress of each participant with an average accuracy of 79%. The authors concluded that their approach is a good starting point for the detection of a subject's stress state in real-time. Such detection alongside some intervention in real-time may improve quality of life.

Muaremi *et al.* [42] presented a stress detection system using a smartphone and wearable chest belt. The authors evaluated their system in a real-world environment with 35 test subjects studied for 4 months. The prediction accuracy was calculated using the leave-one-out-cross-validation (LOOCV) method. The system achieved a 55% accuracy using mobile phone features only (accelerometer) while a 59% prediction accuracy was obtained using the heart rate variability (HRV) feature. The combination of both features gave a prediction accuracy of 61%. Lai *et al.* [42] described an intelligent stress monitoring assistant (SMA) prototype and used a deep learning-based method for stress detection using the WESAD dataset. The authors used Residual-Temporal Convolutional Network (Res-TCN) to recognise and detect stress states with an accuracy of 86% and 96%, respectively. Smets *et al.* [43] used a data-driven approach for stress detection, using real-life data obtained from 1002 subjects in five consecutive, free-living days. The authors found a significant difference between the ECG, skin conductance and skin temperature for different stress levels. They compared their self-reported data with the standard digital phenotypes-based wearable device and achieved the F1-score (a measure of test accuracy using precision and recall) of 0.43, which suggests that the physiological stress response varies greatly between individuals. Thus, the stress detection systems should apply personalised models for accurate stress detection.

From a review of the related literature, it can be noticed that the best features for stress detection and monitoring are still unclear. Several studies have used the same physiological parameters and have implemented the same classifier, yet reported different accuracies. It is important to note that no study has previously tried to find what parameter is the best predictor for stress. The first step should be to establish a statistically significant difference between baseline and stress state before developing any machine learning model. Most studies have reported machine learning models for stress identification using various features available from the sensor that was used to collect the data. The results reported in these studies may only apply to those particular features or sensors. The main objective of this study is to analyze the relative importance of the most common and clinically relevant biophysiological stress indicators and identify the most useful specific indicators for a wearable sensor-based stress monitoring solution. Most previous studies have focused on the determination of sensor/signal ratios. Our study ranks biophysiological stress indicators in order of diagnostic performance using single and multivariable (deviance) analysis. This is a commonly used approach to assess the predictive model, in this case, the stress state response.

The only other study closest to this work is by Zhen *et al.* [44]. According to the authors, the improper imposition of workload on pilots is the most critical cause of the human error. Thus, the authors studied different physiological responses of pilots during flight. These parameters included eye blinking, saccade, pupil diameter, fixation, respiratory rate, and heart rate. They performed statistical analysis to check the sensitivity and diagnostic ability of the aforementioned physiological parameters. They collected data from 12 healthy student pilots and applied a one-way ANOVA test to the collected data. After the experiment, they concluded that from all the physiological parameters, pupil diameter and respiratory rate turned out to be the most sensitive parameters in distinguishing different stages. The diagnostic capability of the parameters was different. Respiratory rate and eye blinking were directly related to the difficulty of the task (stress) while other parameters were affected by external factors, for example, fatigue and attention.

The advantages of our study over the Zhen *et al.* study are three-fold. First, the set of stress measuring features analysed in their paper is different from ours. Secondly, we have performed descriptive and regression analysis. Thirdly, we developed a classification task to evaluate the sensitivity and specificity of selected biophysiological parameters for stress detection. The analysis was performed on a publicly available dataset collected in the Wearable Stress and Affect Detection (WESAD) project [1]. The main objectives of this paper are:

- Descriptive analysis of the commonly used stress monitoring features.
- Regression analysis for the selection of the most important features that can be used for stress monitoring devices in the future.
- Implementation of a uni-variable and multi-variable classification model (using logistic regression) to classify stress state from the non-stress state of an individual.

Figure 1 summarizes the pipeline of the proposed work. The rest of the paper is organized as follows: Section II provides an overview of the WESAD dataset and the methods for preprocessing, normalization and statistical tests to obtain the statistical importance of the features (stress indicators); Section III presents the results and discussion, and conclusions are provided in Section IV.

II. METHODOLOGY

A. STUDY PARTICIPANT

The data were collected using two multimodal devices: a chest-worn device (BioSignalPlux RespiBAN Professional); and a wrist-worn device (Empatica E4). Some recent studies that have used WESAD datasets are Reiss *et al.* [45], Jiang *et al.* [46], Aridas *et al.* [47] and Taufeeq *et al.* [48]. The data included a high-resolution measurement of BVP, EMG, EDA, ECG, RESP, TEMP, and movement from ACC. All the participants were healthy graduate students of the University of Siegen, Germany [1]. Study participants with

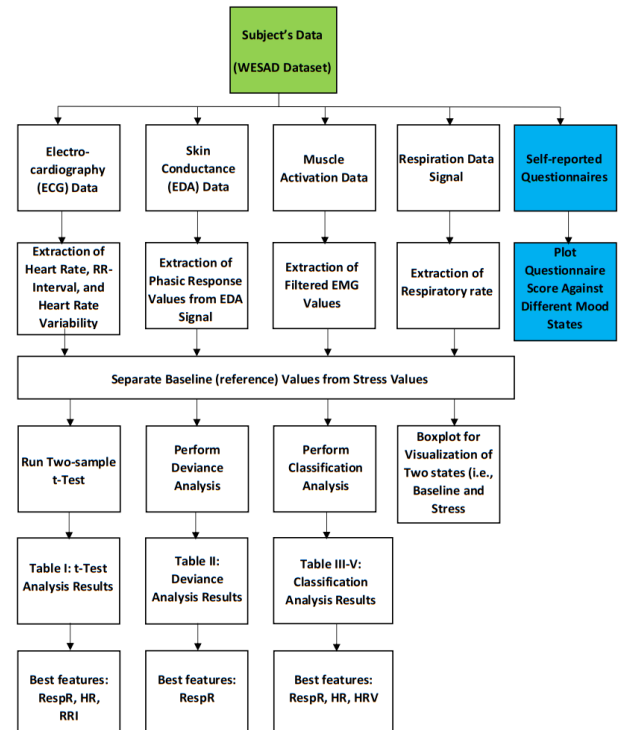


FIGURE 1. Block diagram to summarize the pipeline of the proposed work.

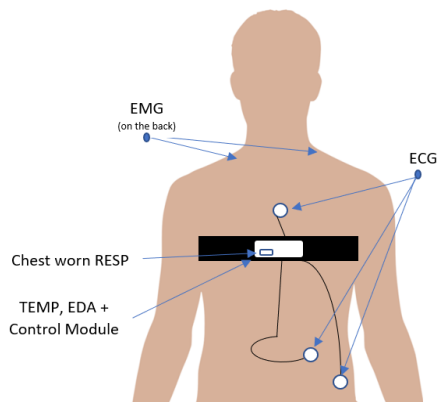
mental disorders, heavy smoking, pregnancy, or those suffering from any cardiovascular and other chronic diseases were excluded from the study. A total of 17 individuals participated in the study but the data of two participants were incomplete due to the malfunctioning of sensors and were therefore removed from the dataset. There were 12 males and 3 females in the remaining 15 subjects with a mean age of 27.5 ± 2.4 (SD) years. Some of the variables were missing in the data from subject no. 11. Thus, all analyses for this paper were completed using 14 subjects. In the dataset, there are 11,500,000 baseline (non-stress) samples and 6,400,000 stress samples.

B. FEATURES RELATED TO STRESS

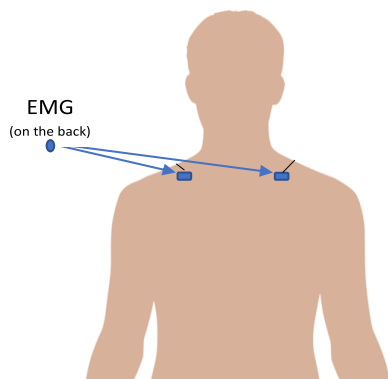
During stress, heart rate usually increased, thus causing more blood to flow within the body. This change in blood flow can be measured through BVP, which is derived from a PPG signal. Change in heart rate and heart rate variability can also be monitored using ECG signals [49], [50]. Stress also causes the release of sweat, thus changing skin conductance properties. This change is measured by the EDA device. There is vast literature available that demonstrates the association of muscle tension with stress. Muscle tension changes are measured using EMG signals [51], [52]. In some people, chronic stress causes a low-grade fever (between 99° to 100° F) and may also cause anxiety as well as restlessness. Thus, Temperature (TEMP) sensors and accelerometer (ACC) readings can also be used to monitor stress [53]–[55].

C. SETUP AND PLACEMENT OF SENSORS

The chest-worn device, RespiBAN Professional, was used to record ECG, EMG, EDA, TEMP, and RESP along with additional ACC data. The placement of the device control unit and sensors is shown in Figure 2.



(a) Front View



(b) Back View

FIGURE 2. Placement of RespiBAN Professional Device (a) shows the placement of different sensors on the front of the human body (b) shows the placement of EMG sensors on the back of the body.

The data from RespiBAN Professional was sampled at 700Hz. The ECG signal was recorded using a standard 3-lead approach (as shown in Figure 2) and an inductive respiration sensor was used to record the RESP signals.

The EDA signals were recorded from the abdomen and EMG was recorded from the muscles of the upper trapezius on both sides of the spine. In addition, Empatica E4 was worn on the dominant hand by all subjects, and BVP, EDA, TEMP, and ACC signals were recorded at the sampling rate of 64Hz, 4Hz, 4Hz, and 32Hz, respectively. All the participant data were recorded on the devices and then transferred to a computer through a wired connection. On the day of study, upon arrival, participants were equipped with chest and wrist-worn sensors. A functionality test was performed to test the working of the sensors. After that, both the devices were synchronised using a double-tap gesture, manually.

D. STUDY PROTOCOL

The study protocol was designed to record readings of 3 different states of the participants, i.e. baseline, amusement, and stress. Participants were also asked to complete a self-reporting questionnaire after each session and undergo a guided meditation session to get de-excited after amusement and stress conditions. Participants could not intake tobacco or caffeine one hour before the study commenced. Moreover, the participants were asked to avoid strenuous exercise on the day of the study. All study participants signed informed consent before commencing. A short sensor test was conducted while equipping the participants. Finally, both the devices (RespiBAN Professional and Empatica E4) were manually synchronized.

For baseline readings, participants were asked to stand or sit at a table and read a magazine. Baseline readings were recorded for 20 minutes and were labelled as a baseline state. Amusement state was induced by showing eleven different funny clips with a gap of 5 seconds between them. The total length of the amusement state was 392 seconds for each participant.

The stress condition was induced using the Trier Social Stress Test (TSST) [56]. TSST consists of mental arithmetic and a public speaking task. Both tasks are considered reliable to evoke stress [19] as they inflict a high mental load and are categorized as a social-evaluative threat in subjects. The participants had to deliver a speech for five minutes on their strengths and weaknesses in front of a panel. Participants were told that the judging panel is from the human resource department and impressing them will increase their hiring chances. After the speech, the panel asked each participant to count backwards, with the gap to 17, from 2023 to 0. If the participant makes any mistake while counting, they had to start over. This exercise following the speech also lasted for five minutes. So, TSST was conducted for a total of 10 minutes. After TSST, participants were given a rest period of 10 minutes. After the amusement and stress period, participants were asked to perform some predefined meditation steps to de-excite and bring them back to a neutral state. Meditation included controlled breathing instructed through an audio track. After removing the sensors, participants were told that the panel was of normal researchers so that they can recover from the test induced stress.

As humans are naturally good at adapting to different situations quickly, two study protocols were designed for this study to keep the randomness and collect the true feelings of the subjects. The two protocols are shown in Figure 3. Half of the subjects followed version 1 protocol while the other half followed version 2 protocol.

E. SIGNAL PROCESSING AND FEATURE EXTRACTION

Raw data from all the sensors (ECG, EDA, EMG, and RESP) were collected using a 0.2-second non-overlapping sliding window, and all physiological features, except EMG, were computed using a 60-second non-overlapping

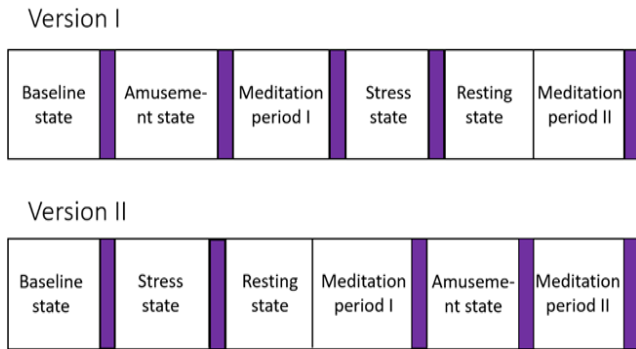


FIGURE 3. The two distinct protocol versions for the proposed study. The dark boxes indicate filling out self-reporting questionnaires.

sliding window. The window sizes were chosen following the recommendations of Koelstra *et al.* [57].

From raw signals of ECG, the heart rate was calculated using the Hamilton peak detection algorithm [58]. Moreover, heart rate variability (HRV) was derived from the locations of the peaks in ECG. Figure 4 shows the block diagram of the Hamilton peak detection algorithm. The algorithm works on the detection of the QRS complex in the ECG signal. The preprocessing steps involve rectification of the signal rather than squaring the signal as in [59], averaging sliding window, low and high pass filtering followed by some QRS detection rules. Rectification of the signal gives us better sensitivity of detection algorithm, which is also indicated in [60]. The QRS complex detection rules are as follows:

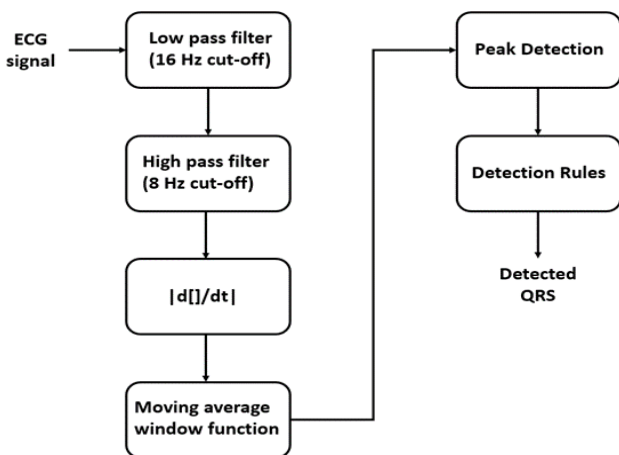


FIGURE 4. Steps to detect QRS complex using Hamilton peak detection algorithm.

- Ignore all the detected peaks preceding or following larger peaks by less than 200 milliseconds.
- If the peak is detected, check whether the signal contains both positive and negative peaks. If not, the detected peak represents a baseline shift.
- If a peak is detected within 360ms of the previously detected peak and had a maximum slope less than 50% of the maximum slope of the previous peak then assume it as T-wave.

- If the detected peak is larger than the detection threshold then consider it as a QRS complex otherwise consider it as noise.

The detection threshold is calculated using estimates of QRS peaks and noise peaks heights and is mathematically represented as:

$$\text{Detection_threshold} = \text{average_noise_peak} + \text{TH} \times (\text{average_QRS_peak} - \text{average_noise_peak}) \quad (1)$$

In equation 1, TH denotes the threshold coefficient between 0.3124 and 0.475. Each time the QRS complex is detected, it is stored in a buffer with previously eight most recent peaks while every non-QRS complex is stored in a buffer that contains previous eight non-QRS peaks also called noise peaks. Through equation 1, we set the detection threshold between the mean or median of QRS and noise peaks. The noise detection is done similarly to [61]. The algorithm characterizes low-frequency noise by the interval between the end T-wave and the start of P-wave while high-frequency noise by bandpass filtered beats outside the QRS complex. In this study, we have used the heart rate and RR interval extracted from the ECG signal using the above-mentioned algorithm.

The sympathetic nervous system controls the EDA response that provides high arousal states with high sensitivity. EDA signals were first passed through a low-pass filter with a critical frequency of 5 Hz, similar to work reported in [62], [63] and phasic (skin conductance response) and tonic (skin conductance level) components were extracted. The phasic component is a short-term response due to some stimulus while the tonic component shows a slow variation in baseline conductance. EDA features can be found in [64], [65]. In this study, we used the phasic components.

The raw EMG signal was processed in two steps. In the first step, the DC component was removed using a high-pass filter and the peak frequency was calculated from the filtered signal by applying a 5-second window. In the second step, a raw EMG signal was passed through a low-pass filter with a cut-off frequency of 50 Hz, to suppress the power line noise, and features were extracted using the method described in [66]. A normalized root means squared (RMS) value of EMG voltage amplitude is used as a feature in this study.

The RESP signal was used to extract the respiratory rate (RspR). Before computing the features of respiration, the raw signal was filtered using a band-pass filter with critical frequencies of 0.1 and 0.35 Hz. A peak detection algorithm was used to identify minima and maxima in the signal and inspiration volume, respiration duration, respiration rate, and inhalation and exhalation ratio were derived as in [19].

F. STATISTICAL FEATURE

We normalized the data from the phasic component of skin conductance, muscle activation, heart rate (HR), RR-interval (RRI), heart rate variability (HRV) and the respiratory rate (RspR) using min-max normalization to eliminate initial

variation in the readings. Data for each are summarized using the mean and standard deviation separately for stress and baseline scenarios. All features along with their mathematical representation are listed in the next subsections.

Let us suppose the above physiological signals are x and x_i is an i -th sample of the signal within the sliding window, where $i = 1, \dots, n$. Then:

1) MEAN

Mean is denoted by \bar{x} and represents the mean value of a raw signal within a sliding window. Mean is calculated by the following equation:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{2}$$

2) STANDARD DEVIATION

Standard deviation is denoted by S and represents the deviation of raw signal around the mean of the signal within the sliding window. Standard deviation is calculated using the following equation:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \tag{3}$$

3) MEDIAN

Median corresponds to the cumulative percentage of 50% i.e. middle reading in a dataset. It is calculated using the equation as:

$$median = \left(\frac{n+1}{2}\right)^{th} \text{ value} \tag{4}$$

Here n is the total number of entries in a dataset.

G. STRESS EVALUATION METHODOLOGY: QUESTIONNAIRES

To validate the protocol, five different self-reports were filled by each participant after every session. First, participants filled out a Positive and Negative Affect Schedule, also known as PANAS. In the second place, six items were picked from the State-Trait Anxiety Inventory (STAI) to measure the anxiety level of each participant. Thirdly, a Self-Assessment Manikins questionnaire (SAM) was used to generate labels in valence arousal space. Finally, nine items were included in a questionnaire from the Short Stress State Questionnaire (SSSQ) to identify the type of stress that prevailed [1]. The outcome of these questionnaires can be considered as subjective reports showing how the participants felt during the test and can be used to train any personalized model. However, for the defined dataset, the study protocol was used to differentiate between the three states and therefore contributing to label different readings. The options of answers were given for each questionnaire. PANAS questionnaire was answered using 5 points scale (1 = not at all and 5 = extremely). The questionnaire asked the subjects about their emotional state i.e. stressed, happy, sad, or frustrated. STAI questionnaire was answered on 4 points

scale (1 = not at all, 4 = very much so) and included questions about the subject’s feelings i.e. were they feeling nervous, relaxed, worried, pleasant, jittery, or ease. Valence and arousal were scored on the scale from 1 = low to 9 = high. The SSSQ questionnaire included questions about what the subject’s mindset was while answering the questionnaire. Subjects answered on 5 points scale where 1 = not at all and 5 = extremely.

The self-reports were also analysed to make sure that the designed experiment was suitable for inducing stress and manipulating the subject’s affective states. Authors in [1] calculated the mean and standard deviation of the anticipated self-reports of three states i.e. baseline, amusement, and stress states along with their subscales. The result of the analysis is shown in Figure 5.

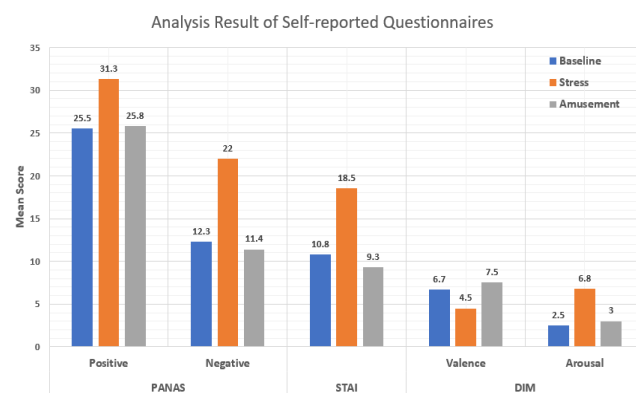


FIGURE 5. Analysis result of self-reported questionnaires.

After baseline and amusement states, the comparison of self-reports revealed that amusement state had the desired effect on the subject i.e. the subjects reported score was high in valence and arousal (dimensional approach, DIM) and less in STAI (anxiety).

The impact of induced stress was noticeably pronounced across all the questionnaires. Analysis of the SSSQ score revealed that the subjects felt more worried and engaged as compare to distressed during the Trier Social Stress Test (TSST) tasks. The score calculated are: Worrying = 10.6, engaged = 11.7 and stressed = 6. The higher value of positive affect (PA) score shows that subject felt energetic and concentrating during the TSST tasks that also resulted in a higher engagement score in SSSQ. The elevated score of negative affect (NA) indicated an increased level of the subject’s stress. The dimensional approach (DIM) score also supports these observations by indicating an increase in arousal score and a decrease in valence score. We had a higher STAI score after TSST, as expected for a subject in a stressful state.

Overall, the analysis of self-reported questionnaires revealed that the designed experimental protocol was suitable to induce desired effective stress in the subjects, especially with respect to stress conditions.

H. STATISTICAL ANALYSIS

For the statistical analysis, only two-state data (Baseline and Stressed states) were used to evaluate the relative importance of each physiological indicator of stress in stress prediction. Three types of analysis were performed: 1) an independent analysis for each biophysiological indicator via a two-sample t-test under the null hypothesis that the mean biophysiological indicator is equal during the Baseline and the Stressed States; 2) a multivariable (deviance) analysis to rank the contribution of each biophysiological indicator in a logistic regression model, defined as follows:

$$\log \left(\frac{p(\text{Stress})}{p(\text{Baseline})} \right) = c_0 + c_1\text{EDA} + c_2\text{EMG} + c_3\text{RRI} + c_4\text{HR} + c_5\text{RspR} + c_6\text{HRV} \quad (5)$$

The logit link function $\log \left(\frac{p(\text{Stress})}{p(\text{Baseline})} \right)$ is used (p is the probability) to relate the log odds of being stressed to the linear predictor where ($c_0, c_1, c_2, c_3, c_4, c_5$ and c_6 are the coefficients showing the direction of the relationship); 3) logistic regression classification analysis to determine the mean absolute error, root mean square error, classification accuracy, sensitivity and specificity of the model.

1) A TWO-SAMPLE T-TEST

The data statistics and the results of the t-test are provided in Table 1. The units of each feature are RspR (breaths per min), HR (beats per min), RRI (milli-sec), Phasic EDA (micro-siemens), EMG (micro-volts) and HRV (milli-sec). The p -value > 0.05 shows the relevant feature has the non-significant mean difference between stress and baseline state values while p -value < 0.05 shows a significant difference in the mean values of stress and baseline condition.

TABLE 1. Statistical analysis result of physiological parameters.

Features	Mean		Standard Deviation		Median		t-test (p-values)
	Baseline	Stress	Baseline	Stress	Baseline	Stress	
RspR	15.51	12.35	2.48	2.43	15.65	12.22	6.02e-08
HR	66.9	95.35	20.01	20.45	68.8	90.66	0.0003
RRI	871	710	157	132	863	702	0.003
Phasic EDA	3.75	5.24	2.92	4.27	2.58	3.89	0.310
EMG	9.5e-5	1.1e-4	5.3e-5	10.8e-5	8.1e-5	8.8e-5	0.660
HRV	51.13	46.97	25.79	22.92	44.33	42.95	0.668

Units: RspR (breaths/min), HR (beats/min), RRI (milli-seconds), Phasic EDA (micro-siemens), EMG (micro-volts) and HRV (milli-seconds)

2) DEVIANCE ANALYSIS

In logistic regression, deviance can be used to assess how good the model is to predict the response (which in this case is stress state) – the lower the deviance, the better the fit to the sample data. To analyze the independent effect of the variables in determining stress, separate regression

models were constructed for combinations of indicators, and the deviance is then used to measure the strength of the relationship between the response and independent variables. Deviance analysis using logistic regression was performed using MATLAB’s statistics toolbox while the classification model was developed using Python code.

3) CLASSIFICATION METHODOLOGY

For the regression model, logistic regression was selected instead of linear regression since our dependent variable in this study is binary i.e. Stress vs No-stress/Baseline. Logistic regression uses the maximum likelihood method to arrive at the solution. Also, the logistic loss function causes large errors to be penalized to an asymptotically constant. The dataset has baseline (11,500,000 per subject) and stress (6,400,000 per subject) samples of the 14 subjects. Given that the number of subjects in the dataset is small, we used stratified k-fold cross-validation with $k = 14$, to ensure that the results achieved are generalizable. Stratified k-fold cross validation ensures the selection of the same proportion of samples of each class in each fold. The time complexity of k-fold is measured by $O(Kn)$, where n is the number of samples. The $O(Kn)$ means that the experiment is repeated K time. When K approaches n , the time complexity becomes $O(n^2)$. So, it can be concluded that as the value of k increases, the systems becomes complex and computationally expensive. We evaluated the classification model using leave-one-out cross-validation (LOOCV) to have an unbiased estimate of the model performance. Stratified k-fold cross validation differs from simple k-fold cross validation by splitting the dataset in such a way that the mean values of all the splits are almost equal.

III. RESULTS AND DISCUSSION

The data from 14 participants were used in the analysis, as some of the variables were missing in the data from subject no. 11. Figure 6 shows the distribution of the data in two states for each variable on boxplots. The data statistics and the results of the t-test are provided in Table 1. Logistic regression is used whenever the outcomes of the analysis are limited, which in this case is stress and baseline (unstress). Thus, logistic regression is used to perform a deviance analysis. Similarly, for the classification task, the response variables (classes) are categorical (yes/no or true/false), so the logistic regression classifier fits best for such type of classification problem and is used as stress versus non-stress classifier.

A. A TWO-SAMPLE T-TEST

Based on the analysis of the p-values, the magnitudes of the coefficients in the logistic regression, and the effect of each variable on the deviance, it can be concluded that respiratory rate (RspR) is the best predictor of stress among these six variables. This result re-enforces the outcome of Zhen et al. [45] that respiratory rate is the most specific and sensitive parameter out of all the other physiological

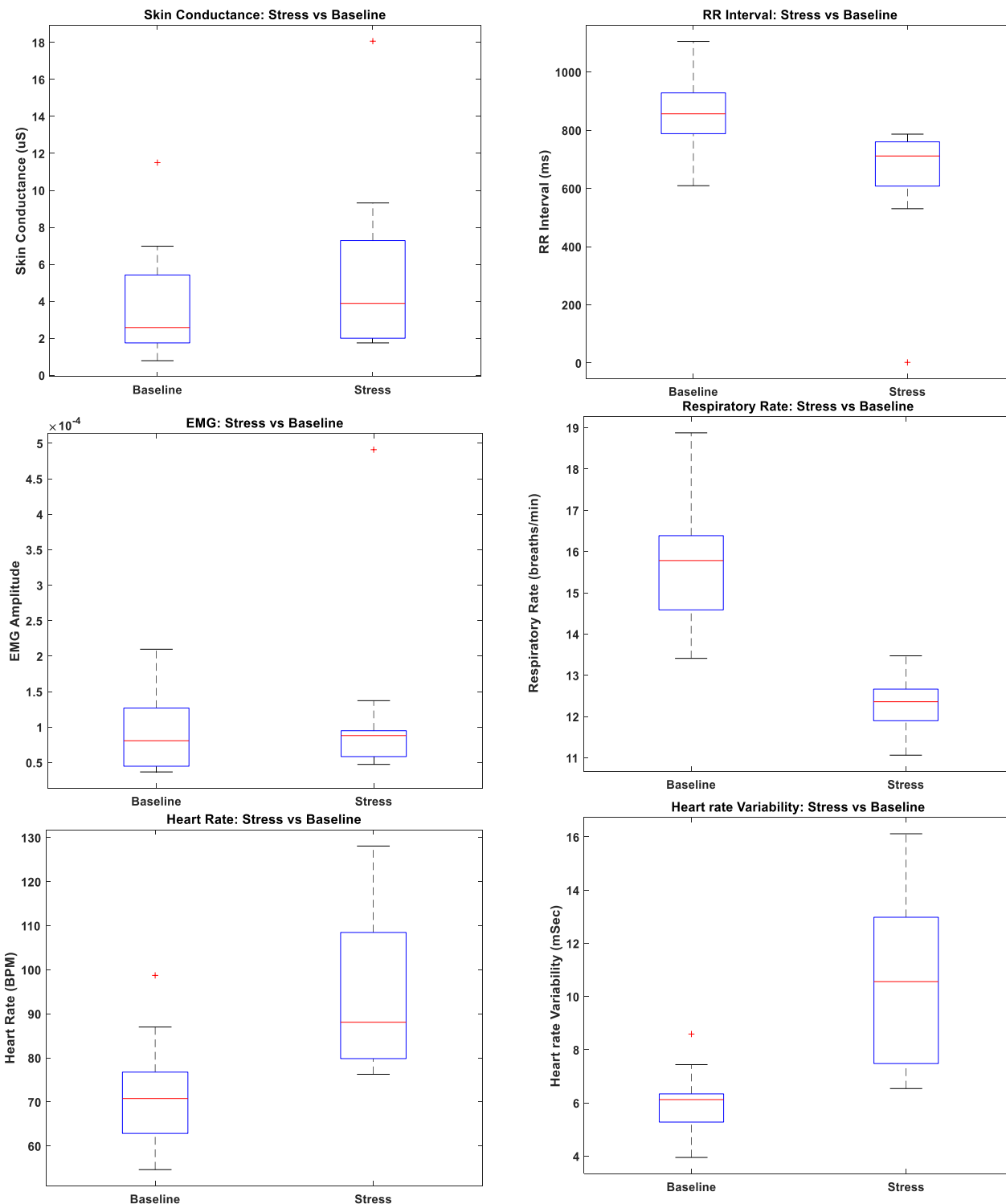


FIGURE 6. Boxplots of the data from two states (Baseline and Stress) for EDA, EMG, HR, RRI, RspR, and HRV, respectively.

parameters and could be used as a stand-alone parameter to detect stress in the lab as well as in the natural environment. Heart rate (HR) combined with respiratory rate can provide a slight improvement in the evaluation or monitoring of stress using wearable sensors. On the other hand, electrodermal activity and electromyogram are poor predictors of stress

and may not add value to the wearable stress monitoring system.

B. DEVIANCE ANALYSIS

Table 2 shows the results of deviance analysis of the fit for single and multi-variant logistic regression models.

The values are sorted in decreasing deviance order. The lowest value model is the best. It is evident that the deviance decreases when the model includes RspR compared to those without RspR. Interestingly, a single variable model, comprising only the RspR, fits better than the multivariable model using EDA, EMG, HR and RRI together. Using any other feature in combination with RspR achieves deviance of close to 0, suggesting a perfect fit for these 14 individuals. Without further samples, it is unclear which combination is optimal.

The box plot of six variables (EMG, EDA, RspR, RRI, HRV and HR) shows that there was evidence of a difference in mean values of RspR, RRI, HRV and HR between baseline and stress states. On the other hand, there was a little difference in mean EMG and EDA for stress and baseline states (see Figure 6), which is also evident by the results of the t-test (see Table 1).

From the above-mentioned results, it can be concluded that the data of EDA and EMG cannot be separated easily, which reinforce our p-value (i.e., the p-value is greater than 0.05) and deviance analysis (Table 2) results. The data of respiratory rate can easily be separated using any logistic fitting curve and thus, qualifies as a most distinctive feature to distinguish baseline state from stress state.

TABLE 2. Logistic regression model fitting (largest to smallest).

Features	Deviance
HRV	39.00
EMG	38.60
EDA	37.65
EDA + EMG	37.62
RRI	25.00
HR	23.73
EDA+EMG+HR	21.49
EDA+EMG+HR+RRI	21.06
RspR	4.70
All	0.00001

C. CLASSIFICATION METHODOLOGY

The result of the logistic regression classification model is shown in Table 3a and Table 3b. The table shows the test-train split for classification, classification accuracy, sensitivity, specificity, Receiver operating characteristic Area Under the Curve (ROC AUC) score, 95% confidence interval of sensitivity and specificity along with likelihood ratios of the developed model. Since the number of subjects in the dataset is very small (n = 14), therefore leave-one-out-cross-validation is also performed to obtain a more robust estimate of the model performance. From the tables, we can see that among the analysed physiological parameters, respiration rate and heart rate give better accuracy than RR interval, skin conductance, muscle activation and heart rate variation. The combination of respiratory rate, heart rate and heart rate variability gives us almost the same accuracy as of combination of all the six parameters. So, we can conclude that the combination of respiratory rate, heart rate and heart rate variability, which can be calculated using a single PPG sensor, is the best predictor of stress.

TABLE 3. (a). Logistic regression classifier results (using 14-fold/LOOC validation): Accuracy, Sensitivity, Specificity and Confidence Intervals. (b). Logistic regression classifier results (using 14-fold/LOOC validation): Likelihood ratio, Variance, Standard Deviance and ROC AUC score.

Features	Test-Train Split	Accuracy (%)	Sensitivity	Specificity	95% Confidence Intervals of sensitivity and specificity	
					Lower	Upper
EMG	Leave One Out Cross Validation	64.96	1.00	0.00	[1.00, 0.00]	[1.00, 0.00]
EDA		63.07	0.93	0.07	[0.93, 0.072]	[0.93, 0.07]
RspR		76.96	0.87	0.59	[0.87, 0.57]	[0.87, 0.57]
RRI		75.35	0.89	0.50	[0.89, 0.50]	[0.89, 0.50]
HR		76.38	0.88	0.55	[0.88, 0.55]	[0.88, 0.55]
HR +RspR		82.81	0.89	0.72	[0.89, 0.71]	[0.89, 0.72]
HR + RspR + HRV		85.70	0.86	0.86	[0.86, 0.67]	[1.04, 1.04]
All combined		85.71	0.86	0.86	[0.86, 0.67]	[1.04, 1.04]

(a)

Features	Test-Train Split	LR+	LR-	Variance	Standard deviance	ROC AUC
EMG	Leave One Out Cross Validation	1.00	inf	0.00	0.00	0.50
EDA		1.00	0.94	0.01	0.11	0.71
RspR		2.12	0.22	0.00	0.06	0.83
RRI		1.78	0.22	0.01	0.11	0.86
HR		1.95	0.22	0.01	0.11	0.88
HR +RspR		3.12	0.16	0.00	0.06	0.93
HR + RspR +HRV		5.97	0.16	0.51	0.22	0.93
All combined		6.00	0.16	0.52	0.23	0.93

(b)

D. GENERAL DISCUSSION

While these classification results indicate the potential of the logistic regression (machine learning) technique to predict stress using the above features, there is still a question of generalizability due to the very small size of the dataset, despite rigorous cross-validation. Therefore, these results need to be validated using a larger dataset. The main conclusion from all three analysis results is that RspR is the best singular feature for detecting stress (from table 1) while the combination of RspR and HR (RRI) are key multi-features of stress, with HRV emerging as the next best.

IV. CONCLUSION

Several human biophysiological variables have been explored to evaluate and monitor both physical and mental stress levels in recent literature. Many of these variables have been independently used in wearable sensor-based devices. This paper is particularly focused on a comparative analysis of these variables in terms of sensitivity and prediction specificity

for stress monitoring. The comparative analysis has been performed by applying a t-test to validate the hypothesis that the physiological data for each variable for the stress and non-stress (baseline) states is statistically differentiable, and logistic regression was applied to identify the strongest predictor of stress.

A logistic regression-based classifier was also trained and validated during this study to determine the classification accuracy of the model. The results of two types of statistical analysis and classification model suggest that respiratory rate is the strongest (stand-alone) predictor of stress compared to other commonly used physiological variables that include heart rate, RR interval, heart rate variability in the ECG/PPG, skin conductance (electrodermal activity) and muscle activation (electromyogram). Heart rate (RRI) emerged as the second-best predictor of stress. The prediction model, consisting of the combination of respiratory rate, heart rate and heart rate variation, derived from a single sensor, gives accurate classification results as a combination of EDA, EMG, RspR, HR (RRI), and HRV. The latter is a more complex sensory system, prone to motion artefacts.

It is important to note that all efforts were focused to provide a fair comparison by using data from the same device and participants. However, there may be other excitation sources (of similar responses) that these experiments failed to capture. Therefore, including context to the data, for example, physical activity, will be key to effective monitoring of stress on a daily basis.

REFERENCES

- [1] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Boulder, CO, USA, Oct. 2018, pp. 400–408, doi: [10.1145/3242969.3242985](https://doi.org/10.1145/3242969.3242985).
- [2] Health and Safety Executive. *Work-Related Ill Health and Occupational Disease in Great Britain*. Accessed: Oct. 31, 2020. [Online]. Available: <https://www.hse.gov.uk/statistics/causdis/>
- [3] E. S. Epel, A. D. Crosswell, S. E. Mayer, A. A. Prather, G. M. Slavich, E. Puterman, and W. B. Mendes, "More than a feeling: A unified view of stress measurement for population science," *Frontiers Neuroendocrinol.*, vol. 49, pp. 146–169, Apr. 2018, doi: [10.1016/j.ynrne.2018.03.001](https://doi.org/10.1016/j.ynrne.2018.03.001).
- [4] A. Arza, J. M. Garzón-Rey, J. Lázaro, E. Gil, R. Lopez-Anton, C. de la Camara, P. Laguna, R. Bailon, and J. Aguiló, "Measuring acute stress response through physiological signals: Towards a quantitative assessment of stress," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 271–287, Jan. 2019, doi: [10.1007/s11517-018-1879-z](https://doi.org/10.1007/s11517-018-1879-z).
- [5] A. Tawakol, A. Ishai, R. A. P. Takx, A. L. Figueroa, A. Ali, Y. Kaiser, Q. A. Truong, C. J. E. Solomon, C. Calcagno, V. Mani, C. Y. Tang, W. J. M. Mulder, J. W. Murrough, U. Hoffmann, M. Nahrendorf, L. M. Shin, Z. A. Fayad, and R. K. Pitman, "Relation between resting amygdala activity and cardiovascular events: A longitudinal and cohort study," *Lancet*, vol. 389, no. 10071, pp. 834–845, Feb. 2017, doi: [10.1016/S0140-6736\(16\)31714-7](https://doi.org/10.1016/S0140-6736(16)31714-7).
- [6] E. G. Brown, A.-M. Creaven, and S. Gallagher, "Loneliness and cardiovascular reactivity to acute stress in younger adults," *Int. J. Psychophysiol.*, vol. 135, pp. 121–125, Jan. 2019, doi: [10.1016/j.ijpsycho.2018.07.471](https://doi.org/10.1016/j.ijpsycho.2018.07.471).
- [7] M. al'Absi, D. Hatsukami, G. L. Davis, and L. E. Wittmers, "Prospective examination of effects of smoking abstinence on cortisol and withdrawal symptoms as predictors of early smoking relapse," *Drug Alcohol Dependence*, vol. 73, no. 3, pp. 267–278, Mar. 2004, doi: [10.1016/j.drugalcdep.2003.10.014](https://doi.org/10.1016/j.drugalcdep.2003.10.014).
- [8] M. Kusserow, O. Amft, and G. Troster, "Monitoring stress arousal in the wild," *IEEE Pervas. Comput.*, vol. 12, no. 2, pp. 28–37, Apr. 2013, doi: [10.1109/MPRV.2012.56](https://doi.org/10.1109/MPRV.2012.56).
- [9] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smaligic, and L. E. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, Chicago, IL, USA, Apr. 2011, pp. 97–108.
- [10] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Informat.*, vol. 92, Apr. 2019, Art. no. 103139, doi: [10.1016/j.jbi.2019.103139](https://doi.org/10.1016/j.jbi.2019.103139).
- [11] S. Gradl, M. Wirth, R. Richer, N. Rohleder, and B. M. Eskofier, "An overview of the feasibility of permanent, real-time, unobtrusive stress measurement with current wearables," in *Proc. 13th EAI Int. Conf. Pervas. Comput. Technol. Healthcare*, New York, NY, USA, May 2019, pp. 360–365, doi: [10.1145/3329189.3329233](https://doi.org/10.1145/3329189.3329233).
- [12] J. C. Nunnally and I. H. Bernstein, "The role of university in the development of entrepreneurial vocations: A Spanish study," in *Psychometric theory*. New York, NY, USA: McGraw-Hill, May 1978, pp. 387–405.
- [13] J. S. Sandhu, M. Paul, and H. Agnihotri, "Biofeedback approach in the treatment of generalized anxiety disorder," *Iranian J. Psychiatry*, vol. 2, no. 3, pp. 90–95, Jan. 2007.
- [14] M. J. H. Begemann, E. J. R. Florisse, R. van Lutterveld, M. Kooyman, and I. E. Sommer, "Efficacy of EEG neurofeedback in psychiatry: A comprehensive overview and meta-analysis," *Transl. Brain Rhythmicity*, vol. 1, no. 1, pp. 19–29, 2016, doi: [10.15761/TBR.1000105](https://doi.org/10.15761/TBR.1000105).
- [15] J. Sánchez-Molina, J. J. Robles-Pérez, and V. J. Clemente-Suárez, "Assessment of psychophysiological response and specific fine motor skills in combat units," *J. Med. Syst.*, vol. 42, no. 4, p. 67, Mar. 2018, doi: [10.1007/s10916-018-0922-9](https://doi.org/10.1007/s10916-018-0922-9).
- [16] R. Delgado-Moreno, J. J. Robles-Pérez, and V. J. Clemente-Suárez, "Combat stress decreases memory of warfighters in action," *J. Med. Syst.*, vol. 41, no. 8, p. 124, Aug. 2017, doi: [10.1007/s10916-017-0772-x](https://doi.org/10.1007/s10916-017-0772-x).
- [17] V. J. Clemente-Suárez, J. J. Robles-Pérez, and J. Fernández-Lucas, "Psychophysiological response in parachute jumps, the effect of experience and type of jump," *Physiol. Behav.*, vol. 179, pp. 178–183, Oct. 2017, doi: [10.1016/j.physbeh.2017.06.006](https://doi.org/10.1016/j.physbeh.2017.06.006).
- [18] K. Hovsepian, M. Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "CStress: Towards a gold standard for continuous stress assessment in the mobile environment," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. (UbiComp)*, Osaka, Japan, 2015, pp. 493–504, doi: [10.1145/2750858.2807526](https://doi.org/10.1145/2750858.2807526).
- [19] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, Jul. 2020, Art. no. 105408, doi: [10.1016/j.cmpb.2020.105408](https://doi.org/10.1016/j.cmpb.2020.105408).
- [20] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: In laboratory and real life," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput., Adjunct*, Heidelberg, Germany, Sep. 2016, pp. 1185–1193, doi: [10.1145/2968219.2968306](https://doi.org/10.1145/2968219.2968306).
- [21] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao, "Detecting work-related stress with a wearable device," *Comput. Ind.*, vol. 90, pp. 42–49, Sep. 2017, doi: [10.1016/j.compind.2017.05.004](https://doi.org/10.1016/j.compind.2017.05.004).
- [22] E. Ullmann, A. Barthel, K. Petrowski, T. Stalder, C. Kirschbaum, and S. R. Bornstein, "Pilot study of adrenal steroid hormones in hair as an indicator of chronic mental and physical stress," *Sci. Rep.*, vol. 6, no. 1, pp. 1–7, May 2016, doi: [10.1038/srep25842](https://doi.org/10.1038/srep25842).
- [23] S. Klangphukhiew, R. Srichana, and R. Patramanon, "Cortisol stress biosensor based on molecular imprinted polymer," *Multidisciplinary Digit. Publishing Institute*, Paris, France, Tech. Rep. 4, Sep. 2017, p. 538, vol. 1, no. 4, doi: [10.3390/proceedings1040538](https://doi.org/10.3390/proceedings1040538).
- [24] B. Valentin, O. Grotte, M. Skorning, S. Bergrath, H. Fischermann, D. Rörtgen, M.-T. Mennig, C. Fitzner, M. P. Müller, C. Kirschbaum, R. Rossaint, and S. K. Beckers, "Cortisol and alpha-amylase as stress response indicators during pre-hospital emergency medicine training with repetitive high-fidelity simulation and scenarios with standardized patients," *Scandin. J. Trauma, Resuscitation Emergency Med.*, vol. 23, no. 1, pp. 1–8, Apr. 2015, doi: [10.1186/s13049-015-0110-6](https://doi.org/10.1186/s13049-015-0110-6).
- [25] J. Botelho, V. Machado, P. Mascarenhas, J. Rua, R. Alves, M. A. Cavacas, A. Delgado, and J. J. Mendes, "Stress, salivary cortisol and periodontitis: A systematic review and meta-analysis of observational studies," *Arch. Oral Biol.*, vol. 96, pp. 58–65, Dec. 2018, doi: [10.1016/j.archoralbio.2018.08.016](https://doi.org/10.1016/j.archoralbio.2018.08.016).

- [26] A. Zamkah, T. Hui, S. Andrews, N. Dey, F. Shi, and R. S. Sherratt, "Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors," *Biosensors*, vol. 10, no. 4, p. 40, Apr. 2020, doi: [10.3390/bios10040040](https://doi.org/10.3390/bios10040040).
- [27] E. Iob and A. Steptoe, "Cardiovascular disease and hair cortisol: A novel biomarker of chronic stress," *Current Cardiol. Rep.*, vol. 21, no. 10, pp. 1–11, Aug. 2019, doi: [10.1007/s11886-019-1208-7](https://doi.org/10.1007/s11886-019-1208-7).
- [28] T. Iqbal, A. Elahi, P. Redon, P. Vazquez, W. Wijns, and A. Shahzad, "A review of biophysiological and biochemical indicators of stress for connected and preventive healthcare," *Diagnostics*, vol. 11, no. 3, p. 556, Mar. 2021, doi: [10.3390/diagnostics11030556](https://doi.org/10.3390/diagnostics11030556).
- [29] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 3, pp. 634–645, Mar. 2018, doi: [10.1109/TIM.2017.2779329](https://doi.org/10.1109/TIM.2017.2779329).
- [30] I. Mohino-Herranz, R. Gil-Pita, J. Ferreira, M. Rosa-Zurera, and F. Seoane, "Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones," *Sensors*, vol. 15, no. 10, pp. 25607–25627, Oct. 2015, doi: [10.3390/s151025607](https://doi.org/10.3390/s151025607).
- [31] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, pp. 50–53, Apr. 2018, doi: [10.1016/j.bbr.2017.12.021](https://doi.org/10.1016/j.bbr.2017.12.021).
- [32] D. S. Lee, T. W. Chong, and B. G. Lee, "Stress events detection of driver by wearable glove system," *IEEE Sensors J.*, vol. 17, no. 1, pp. 194–204, Jan. 2016, doi: [10.1109/JSEN.2016.2625323](https://doi.org/10.1109/JSEN.2016.2625323).
- [33] J. Healey and R. Picard, "SmartCar: Detecting driver stress," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, Barcelona, Spain, Sep. 2000, pp. 218–221, doi: [10.1109/ICPR.2000.902898](https://doi.org/10.1109/ICPR.2000.902898).
- [34] J. Wijnsman, B. Grundleher, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Boston, MA, USA, Aug./Sep. 2011, pp. 1798–1801, doi: [10.1109/IEMBS.2011.6090512](https://doi.org/10.1109/IEMBS.2011.6090512).
- [35] L.-L. Chen, Y. Zhao, P.-F. Ye, J. Zhang, and J.-Z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Syst. Appl.*, vol. 85, pp. 279–291, Nov. 2017, doi: [10.1016/j.eswa.2017.01.040](https://doi.org/10.1016/j.eswa.2017.01.040).
- [36] Y. Shi, M. H. Nguyen, P. Blitz, B. French, S. Fisk, F. De la Torre, A. Smailagic, D. P. Siewiorek, M. Absi, E. Ertin, T. Kamarck, and S. Kumar, "Personalized stress detection from physiological measurements," in *Proc. Int. Symp. Qual. Life Technol.*, Pittsburgh, PA, USA, Jun. 2010, pp. 28–29. Accessed: Apr. 20, 2021. [Online]. Available: https://www.google.com.hk/url?sa=t&rt=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi31Y-Atf3uAhXTzlsBHab1BRQQFjABegQIARAD&url=http%3A%2F%2Fwww.humansensing.cs.cmu.edu%2Fsites%2Fdefault%2Ffiles%2F8stress_detect.pdf&usg=AOvVaw0wS9vviQE4GWxjDptTX9hy
- [37] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, May 2004, doi: [10.1007/BF02344719](https://doi.org/10.1007/BF02344719).
- [38] F. T. Sun, C. Kuo, H. T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *Proc. Int. Conf. Mobile Comput., Appl., Services*, vol. 76, Seattle, WA, USA: Springer, Oct. 2012, pp. 211–230, doi: [10.1007/978-3-642-29336-8_12](https://doi.org/10.1007/978-3-642-29336-8_12).
- [39] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *Int. J. Neural Syst.*, vol. 27, no. 2, Mar. 2017, Art. no. 1650041, doi: [10.1142/S0129065716500416](https://doi.org/10.1142/S0129065716500416).
- [40] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress detection using wearable physiological sensors," in *Proc. Int. Work-Confer. Interplay Between Natural Artif. Comput.* Elche, Spain: Springer, Jun. 2015, pp. 526–532, doi: [10.1007/978-3-319-18914-7_55](https://doi.org/10.1007/978-3-319-18914-7_55).
- [41] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *Bio-NanoScience*, vol. 3, no. 2, pp. 172–183, May 2013, doi: [10.1007/s12668-013-0089-2](https://doi.org/10.1007/s12668-013-0089-2).
- [42] K. Lai, S. N. Yanushkevich, and V. P. Shmerko, "Intelligent stress monitoring assistant for first responders," *IEEE Access*, vol. 9, pp. 25314–25329, Feb. 2021, doi: [10.1109/ACCESS.2021.3057578](https://doi.org/10.1109/ACCESS.2021.3057578).
- [43] E. Smets, E. R. Velazquez, G. Schiavone, I. Chakroun, E. D'Hondt, W. De Raedt, J. Cornelis, O. Janssens, S. Van Hoecke, S. Claes, I. Van Diest, and C. Van Hoof, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–10, Dec. 2018, doi: [10.1038/s41746-018-0074-9](https://doi.org/10.1038/s41746-018-0074-9).
- [44] Z. Wang and S. Fu, "An analysis of pilot's physiological reactions in different flight phases," in *Proc. Int. Conf. Eng. Psychol. Cognit. Ergonom.* Heraklion, Greece: Springer, Jun. 2014, pp. 94–103, doi: [10.1007/978-3-319-07515-0_10](https://doi.org/10.1007/978-3-319-07515-0_10).
- [45] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, Jul. 2019, doi: [10.3390/s19143079](https://doi.org/10.3390/s19143079).
- [46] Y. Jiang, W. Li, M. S. Hossain, M. Chen, A. Alelawi, and M. Al-Hammadi, "A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition," *Inf. Fusion*, vol. 53, pp. 209–221, Jan. 2020, doi: [10.1016/j.inffus.2019.06.019](https://doi.org/10.1016/j.inffus.2019.06.019).
- [47] C. K. Aridas, S. Karlos, V. G. Kanas, N. Fazakis, and S. B. Kotsiantis, "Uncertainty based under-sampling for learning naive Bayes classifiers under imbalanced data sets," *IEEE Access*, vol. 8, pp. 2122–2133, Dec. 2020, doi: [10.1109/ACCESS.2019.2961784](https://doi.org/10.1109/ACCESS.2019.2961784).
- [48] M. T. Uddin and S. Canavan, "Synthesizing physiological and motion data for stress and meditation detection," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Cambridge, U.K., Sep. 2019, pp. 244–247, doi: [10.1109/ACIIW.2019.8925245](https://doi.org/10.1109/ACIIW.2019.8925245).
- [49] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, A Coruna, Spain, Sep. 2019, pp. 1–5, doi: [10.23919/EUSIPCO.2019.8902631](https://doi.org/10.23919/EUSIPCO.2019.8902631).
- [50] A. Affanni, "Wireless sensors system for stress detection by means of ECG and EDA acquisition," *Sensors*, vol. 20, no. 7, p. 2026, Apr. 2020, doi: [10.3390/s20072026](https://doi.org/10.3390/s20072026).
- [51] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, early access, Jul. 9, 2019, doi: [10.1109/TAFFC.2019.2927337](https://doi.org/10.1109/TAFFC.2019.2927337).
- [52] S. Pourmohammadi and A. Maleki, "Stress detection using ECG and EMG signals: A comprehensive study," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105482, doi: [10.1016/j.cmpb.2020.105482](https://doi.org/10.1016/j.cmpb.2020.105482).
- [53] T. Oka, "Psychogenic fever: How psychological stress affects body temperature in the clinical population," *Temperature*, vol. 2, no. 3, pp. 368–378, Jul. 2015, doi: [10.1080/23328940.2015.1056907](https://doi.org/10.1080/23328940.2015.1056907).
- [54] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. F. McKeegan, "Skin temperature reveals the intensity of acute stress," *Physiol. Behav.*, vol. 152, pp. 225–230, Dec. 2015, doi: [10.1016/j.physbeh.2015.09.032](https://doi.org/10.1016/j.physbeh.2015.09.032).
- [55] F. de Arriba-Pérez, J. M. Santos-Gago, M. Caeiro-Rodríguez, and M. Ramos-Merino, "Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 12, pp. 4925–4945, Jan. 2019, doi: [10.1007/s12652-019-01188-3](https://doi.org/10.1007/s12652-019-01188-3).
- [56] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'—A tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, nos. 1–2, pp. 76–81, 1993, doi: [10.1159/000119004](https://doi.org/10.1159/000119004).
- [57] S. D. Kreibitz, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, Jul. 2010, doi: [10.1016/j.biopsycho.2010.03.010](https://doi.org/10.1016/j.biopsycho.2010.03.010).
- [58] P. Hamilton, "Open source ECG analysis," in *Proc. Comput. Cardiol.*, Memphis, TN, USA, Sep. 2002, pp. 101–104, doi: [10.1109/CIC.2002.1166717](https://doi.org/10.1109/CIC.2002.1166717).
- [59] X. Tang, Q. Hu, and W. Tang, "A real-time QRS detection system with PR/RT interval and ST segment measurements for wearable ECG sensors using parallel delta modulators," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 4, pp. 751–761, Aug. 2018, doi: [10.1109/TBCAS.2018.2823275](https://doi.org/10.1109/TBCAS.2018.2823275).
- [60] C. A. Ledezma and M. Altuve, "Optimal data fusion for the improvement of QRS complex detection in multi-channel ECG recordings," *Med. Biol. Eng. Comput.*, vol. 57, no. 8, pp. 1673–1681, Aug. 2019, doi: [10.1007/s11517-019-01990-3](https://doi.org/10.1007/s11517-019-01990-3).
- [61] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, May 2018, doi: [10.1016/j.bspc.2018.03.003](https://doi.org/10.1016/j.bspc.2018.03.003).
- [62] G. Pope, V. Mishra, S. Lewia, B. Lowens, D. Kotz, S. Lord, and R. Halter, "An ultra-low resource wearable EDA sensor using wavelet compression," in *Proc. IEEE 15th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Las Vegas, NV, USA, Mar. 2018, pp. 193–196, doi: [10.1109/BSN.2018.8329691](https://doi.org/10.1109/BSN.2018.8329691).

- [63] B. Lamichhane, U. Grobekahtofer, G. Schiavone, and P. Casale, "Towards stress detection in real-life scenarios using wearable sensors: Normalization factor to reduce variability in stress physiology," in *eHealth 360°*. Cham, Switzerland: Springer, Dec. 2017, pp. 259–270, doi: [10.1007/978-3-319-49655-9_34](https://doi.org/10.1007/978-3-319-49655-9_34).
- [64] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 2, pp. 279–286, Mar. 2012, doi: [10.1109/TITB.2011.2169804](https://doi.org/10.1109/TITB.2011.2169804).
- [65] S. Kim, W. Rhee, D. Choi, Y. J. Jang, and Y. Yoon, "Characterizing driver stress using physiological and operational data from real-world electric vehicle driving experiment," *Int. J. Automot. Technol.*, vol. 19, no. 5, pp. 895–906, Sep. 2018, doi: [10.1007/s12239-018-0086-0](https://doi.org/10.1007/s12239-018-0086-0).
- [66] J. Wijsman, B. Grundlehner, J. Penders, and H. Hermens, "Trapezius muscle EMG as predictor of mental stress," *ACM Trans. Embedded Comput. Syst.*, vol. 12, no. 4, Jul. 2013, Art. no. 99, doi: [10.1145/2485984.2485987](https://doi.org/10.1145/2485984.2485987).



TALHA IQBAL received the B.Sc. and M.Sc. degrees in electrical engineering from COMSATS University Islamabad, Pakistan, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Medicine, National University of Ireland, Galway.

His research spans the disciplines of engineering and medicine, with a focus on AI/machine learning, data science, biosensors, wearable devices, and signal processing. His Ph.D.

is focused on the investigation and development of a novel wearable device to detect and predict stress using bio-physiological biomarkers.

Mr. Iqbal was a recipient of the COMSATS full-time Scholarship (A+ category) (2017–2018), the QCRI/Boeing Travel Grant, Qatar, in 2018, the Top-10 APPRC Award (2017–2018), Australia, and the Scholarship under SFI Professorship Award, Ireland, in 2019.



PAU REDON-LURBE received the B.S. degree in chemical engineering from Alfonso X El Sabio University, Madrid, Spain, in 2002, and the M.S. degree in mechanical engineering and project management and the Ph.D. degree in mechanical and chemical engineering from the Polytechnic University of Valencia, in 2007 and 2013, respectively.

From 2013 to 2018, he was a Postdoctoral Fellow with the Carlos III Health Institute, Madrid,

in the development of new clinical procedures, patient-friendly and using wearable technology, oriented to fight childhood obesity, cardiovascular disease, and the neonatal population. He has published his research results in 30 articles, as the first author, and has more than 20 oral presentations in prestigious peer-review journals and international congresses, respectively. His research interests include wearable technology, biosensors, cardiac biomarkers, machine learning and artificial intelligence, and commercialization of research-based developed technology.

Dr. Redon-Lurbe was awarded the prestigious Marie Skłodowska-Curie Fellowship to continue his research at CÚRAM SFI Research Centre for Medical Devices and the Smart Sensor Laboratory, Galway, Ireland, in 2018.



ANDREW J. SIMPKIN received the B.A. degree (Hons.) in mathematics from Trinity College Dublin, in 2007, and the Ph.D. degree in statistics from the National University of Ireland, Galway, in 2011.

From 2011 to 2017, he was a Research Fellow with the School of Social and Community Medicine, University of Bristol. Since 2017, he has been a Lecturer (above the bar) in statistics with the School of Mathematics, Statistics and Applied

Mathematics, NUI. His research focuses on longitudinal data analysis, data science, and functional data analysis, with applications to sensor technology.



ADNAN ELAHI received the B.S. degree in computer engineering from COMSATS University, Lahore, Pakistan, in 2006, the M.Sc. degree in embedded digital systems from the University of Sussex, in 2009, and the Ph.D. degree in electrical and electronic engineering and the M.Sc. degree in data analytics from the National University of Ireland (NUI) Galway, Ireland, in 2018 and 2020, respectively.

He is currently a Lecturer in electrical and electronic engineering with the NUI Galway. He has previously worked as a Lecturer at COMSATS University, from 2007 to 2012, and as a Postdoctoral Researcher at Translational Medical Device Laboratory, NUI Galway, from 2016 to 2019. His research spans the disciplines of engineering and medicine, with a particular focus on smart devices for remote patient monitoring; novel and personalised therapeutics using electroporation and neurostimulation; and AI/machine learning for biomedical signals.



SANDRA GANLY received the B.E. and Ph.D. degrees in biomedical engineering from the National University of Ireland, Galway, in 2002 and 2016, respectively. From 2002 to 2005, she was an engineer in the industry. She started her Ph.D. research in the area of gene-eluting coronary stents, in 2005, and has been held successively senior roles managing multi-disciplinary projects important to the Irish Medtech Sector, since 2008. She currently holds

a research fellow position at NUI Galway's cardiovascular research group focused on preventive cardiology and risk-factor research.



WILLIAM WIJNS received the M.D. degree in cardiology and the Ph.D. degree from the University of Louvain Medical School, Belgium.

After working at Erasmus University, The Netherlands, UCLA, University of Louvain, and the Cardiovascular Research Center Aalst, he joined the National University of Ireland Galway as a Science Foundation Ireland Research Professor Awardee, in 2016, and co-directing the Smart Sensors Laboratory, NUI Galway. His

research focuses on the development and evaluation of novel device-based therapies for cardiovascular diseases. Ongoing projects aim to prevent high-risk subjects from suffering cardiac disability by modification of vulnerable plaques and modulation of trigger mechanisms that precipitate acute events.

Dr. Wijns is currently the Chairman of PCR, an educational platform that connects the interventional cardiology community across the globe.



ATIF SHAHZAD received the B.S. degree in computer engineering from COMSATS University, Lahore, Pakistan, in 2006, the M.S. degree in electronic and electrical engineering from the University of Leeds, U.K., in 2009, and the Ph.D. degree in electrical engineering from the National University of Ireland (NUI) Galway, Ireland, in 2017.

He is an Honorary Lecturer with the School of Medicine, NUI Galway, and a Research Fellow with the Institute of Metabolism and Systems Research, University of Birmingham. He is also the Joint Director of the Smart Sensors Laboratory, School of Medicine, NUI Galway. His research interests include biosensing and medical technologies, medical signal and image processing, applied electromagnetics, and computational modeling. He is also a Topic Editor of *BIOSENSORS* journal.