



Toward soft real-time stress detection using wrist-worn devices for human workspaces

Sunder Ali Khowaja¹ · Aria Ghora Prabono¹ · Feri Setiawan¹ · Bernardo Nugroho Yahya¹ · Seok-Lyong Lee¹

Published online: 28 September 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Continuous exposure to stress leads to many health problems and substantial economic loss in companies. A lot of attention has been given to the development of wearable systems for stress monitoring to tackle its long-term effects such as confusion, high blood pressure, insomnia, depression, headache and inability to take decisions. Accurate detection of stress from physiological measurements embedded in wearable devices has been the primary goal in the healthcare industry. Advanced sensor devices with a high sampling rate have been proven to achieve high accuracy in many earlier works. However, there has been a little attempt to employ consumer-based devices with a low sampling rate, which potentially degrades the performance of detection systems. In this paper, we propose a set of new features, local maxima and minima (LMM), from heart rate variability and galvanic skin response sensors along with the voting and similarity-based fusion (VSBF) method, to improve the detection performance. The proposed feature set and fusion method are first tested on the acquired dataset which is collected using the wrist-worn devices with a low sampling rate in workplace environments and validated on a publicly available dataset, driveDB from PhysioNet. The experimental results from both datasets prove that the LMM features can improve the detection accuracy for different classifiers in general. The proposed VSBF method further boosts the recognition accuracy by 5.69% and 2.90% in comparison with the AdaBoost, which achieves the highest accuracy as a single classifier on the acquired, and the DriveDB dataset, respectively. Our analyses show that the stress detection system using the acquired dataset yields an accuracy of 92.05% and an F_1 score of 0.9041. Based on the analyses, a soft real-time system is implemented and validated to prove the applicability of the proposed work for stress detection in a real environment.

Keywords Stress detection · Classification · HRV · GSR · Wearable sensors · Soft real-time system

1 Introduction

The use of wearable technologies is increasing at a fast pace in the field of smart homes, clinical perspectives and healthcare environments (Patel et al. 2012). The emergence of the technologies allows us to monitor the mental and

physical health status of an individual across the range of contexts. Among them, stress is considered to be a major issue that affects both personal and professional lives (Cox et al. 2000; World Health Organisation 2013; Mental Health Foundation 2017). Continuous exposure to stressful conditions may lead to some mental health problems, such as anxiety and depression. Therefore, early recognition of stress conditions not only helps an individual to cope with the negative mental state but also helps to improve the

Communicated by V. Loia.

✉ Bernardo Nugroho Yahya
bernardo@hufs.ac.kr

✉ Seok-Lyong Lee
sllee@hufs.ac.kr

Sunder Ali Khowaja
sandar.ali@usindh.edu.pk

Aria Ghora Prabono
ghora@hufs.ac.kr

Feri Setiawan
ferisetiawan.mail@gmail.com

¹ Department of Industrial and Management Engineering, Hankuk University of Foreign Studies, 89, Wangsan-ri, Mohyun-myon, Yongin-si, Gyeonggi-do 449-791, Republic of Korea

quality of life. From the research point of view, stress recognition allows the researchers to have an insight into the trigger points or key activities that stimulate negative cognition and behavior.

Stress is an inherent response of the human body caused by external disturbances. These physiological responses can be recorded using biological sensors, such as HRV, GSR, pupil dilation, skin temperature and blood volume pulse (BVP) (Renaud and Blondin 1997; Wikgren et al. 2012). Though the stress at a very small scale may have some constructive influence on the human body, but in general its impact is negative which causes a gradual loss of memory, lack of decision-making ability and reduced focus (Stawski et al. 2006; Sandi 2013). On the other hand, large-scale stress can affect the quality of life as well as may lead to some mental health problems, like impetuous aging, depression, and anxiety (Dickerson and Kemeny 2004; Rai et al. 2012). Stress recognition systems have been extensively studied for elder care homes and hospitals, but very little attention has been drawn toward workplaces. At workspaces, psychosocial stress has an adverse effect not only on health care but also on the economy, which makes it one of the major problems for society. Mental Health Foundation in the UK (Mental Health Foundation 2017) stated that approximately 12 million adults suffer from stress-related problems. Similarly, a study conducted by the World Health Organization (2016) reveals that UK enterprises have to bear an amount of £8.4 million annually due to stress-related illness. Hence, stress is now globally considered as a significant problem at office workspaces which include white collar jobs and blue collar jobs. In this regard, numerous international organizations are working on a priority basis to reduce their impact (Cox et al. 2000; World Health Organisation 2013). Therefore, the design of recognition systems with pervasive devices that can timely detect the stress condition is in need, so that its effect can be reduced or lessened, accordingly.

In workplaces, stress can be caused by many reasons. Some of the notables include shorter deadlines, high workload, work-life imbalance and so forth (Sano and Picard 2013; Saleem et al. 2015). Psychosocial stress recognition can be performed either by using computer vision techniques or by employing physiological sensors. Vision-based methods exploit visual data (photos and videos) to recognize stress. These methods utilize cameras to detect and recognize facial expressions using several different techniques such as image analysis, feature extraction and depth imaging (Khowaja et al. 2015). Although the vision-based method has been recognized as an approach with high accuracy, there are still unavoidable limitations, like poor lighting conditions or inhomogeneity issues (Hernández et al. 2014). Other issues related to this

method are the complexity of arrangement, i.e., steady cameras with fixed angles to determine the point of interest, and the extraction of descriptive features from image or videos that takes much time for computational processing, restricting the applicability of such systems for real-time applications. On the other hand, sensor-based methods allow us to collect and store multiple sensor data, such as GSR, HRV and BVP over protracted periods of time (Habib et al. 2014). As our application is for the human workspaces, i.e., white collar and blue collar jobs, the complex set of arrangements of computer vision methods can be a hindrance, and thus, wearable sensors might be more suitable for this particular context.

Existing studies mainly took into account the sensor readings from high-sampling-rate (HSR) devices, such as Biopac systems (Systems 2017) and FlexComp (Technology 2016). These systems can record the data with HSR but are complex in terms of arrangement and usability. In recent times, wearable technologies, such as smart pendants, smart glasses (Google glass) and tracking devices (including PillCam), have experienced an era of drastic growth (MarketsandMarkets 2016). It has also been predicted that the consumer market for the pervasive wearable sensors will continue to grow in healthcare and medical sectors and will have a market share of over \$31.96 billion by the end of 2025 at a projected compound annual growth rate of 19.15% with the largest market being North America and the fastest growing market being Asia Pacific (Intelligence 2019). A survey report by (Zimmermann) at CNN Business shows that the sales of smartwatches have been increased from 5.0 million in 2014 to 79.1 million in 2019. The trending use of smartwatch and wearable devices in the healthcare industry has also been reported by various studies and surveys (Hayward 2018; Future 2019; Insights 2019; Research 2019). Some of the studies have started focusing on the use of low sampling rate (LSR) wrist-worn devices for affective computing research to prove the real-life applicability of such systems (Gjoreski et al. 2016; Zenonos et al. 2016; Setiawan et al. 2018). The market growth and research trend suggest that people prioritize to wear the sensors having simplicity and easiness rather than the complex arrangement, and such simple devices are commonly LSR. In particular, we use a smartwatch that provides real-time biological responses from HRV and GSR sensors. HSR wearable devices allow researchers to derive complex features from the sensor measurements with shorter time windows, i.e., 1–10 s, which subsequently results in high detection accuracy such as average beat detection (ABD) (Keshan et al. 2015), but the same features cannot be derived accurately using LSR wrist-worn devices, leading to intrinsic low detection rates.

In this regard, we propose a new feature set for stress detection which is designed for achieving a considerable

level of detection accuracy from LSR wrist-worn devices using longer time windows (60 s). The proposed feature set is based on local maxima and minima (LMM) derived from different probability distributions. LMM features have binary characteristics that allow us to further use our proposed decision-level method, i.e., voting and similarity-based fusion (VSBF) to improve the detection accuracy. It has been proven by existing studies that the combination of multiple classifiers tends to improve the performance of classification systems (Khowaja et al. 2017, 2018a; Khowaja and Lee 2020); therefore, we assume that proposing a new decision-level fusion method with LMM features might help in improving the detection accuracy. The main motivation of our proposed work using wrist-worn devices is to make the stress detection system applicable to real-life environments. Our method needs to process the data from longer time windows continuously; therefore, it is hard to implement the system with real-time characteristics due to the strict time constraints, i.e., deadlines, and processing overhead. However, the soft real-time systems can compromise over the deadlines for optimizing specific application criteria (Laplante 2004). In this regard, we implement a soft real-time system for stress detection to prove the applicability of our proposed work in real-time environments.

We perform two kinds of analyses which cover (1) validating the performance of our method in terms of accuracy using the dataset acquired from LSR wrist-worn devices and (2) showing the strength of the LMM features and VSBF method using a publicly available dataset. For the former analysis, we collect the data using the international affective picture system (IAPS) (Lang et al. 1999) which is widely used for inducing stress. The latter analysis shows the effectiveness of our method using a dataset, “driveDB” from PhysioNet (Healey 2000), by comparing the obtained results with existing works. Our experimental analyses reveal that the LMM features and fusion method can improve the accuracy of both datasets. In summary, the contributions of this study are as follows:

- A set of new features are introduced to improve the accuracy of stress detection and evaluated on the acquired and publicly available datasets.
- We introduce the consideration of longer time windows for LSR devices to improve the performance of the detection system.
- A new decision-level fusion method is proposed based on the voting and similarity measure from binary features.
- In-depth analyses for stress detection are carried out using different classification algorithms.

- A soft real-time system for stress detection is implemented to prove the applicability of the detection system in real-world environments.

The paper is structured as follows: Sect. 2 describes the related work. Section 3 explains the methodology for our stress detection system. Section 4 presents the quantitative analyses for validating LMM features and the VSBF method on the acquired and the driveDB dataset. Section 5 elaborates on the details regarding the soft real-time implementation of our system. Section 6 presents a discussion with quantitative analyses, merits and limitations of our work with future directions. Finally, Sect. 7 concludes our work.

2 Related works

Stress detection systems have emerged greatly as less comfortable sensors in a constrained environment were changed to more comfortable sensors in the less constrained environment. Healey and Picard (2005) were the pioneers in detecting stress using physiological sensors by using intrusive wires and electrodes to acquire the data. With the emergence of sophisticated devices, these wires and electrodes are replaced with more comfortable sensors, such as a smartwatch or smart pendants which can acquire physiological data quite effectively. Since 2005, a lot of focus was given to detect stress using signal processing and machine learning techniques with a complex arrangement of sensors.

Most of the studies use sensors, such as GSR (Healey and Picard 2005; Sano and Picard 2013), electrocardiogram (ECG) (Healey and Picard 2005; Sierra et al. 2011; Muaremi et al. 2014), BVP (Handouzi et al. 2014), respiration (RESP) (Healey and Picard 2005; Muaremi et al. 2014; Hovsepian et al. 2015), electromyogram (EMG) (Healey and Picard 2005; Wijnsman et al. 2013) and heart rate (HR) (Sierra et al. 2011). Work illustrated in (Setz et al. 2010) uses Montreal imaging stress task (MIST) (Dedovic et al. 2005) to induce the stress state on participants and use GSR measurements to classify stress and normal state. The study reported a cross-validation accuracy of 82.8%. Similarly, in Salahuddin et al. (2007), the Stroop test (Stroop 1935) was used to tempt the emotion and an ECG sensor was employed for recording the stress state. The study mentioned above only documented the details for short-term HRV features from the ECG sensor, and the classifiers were not employed to discriminate between stress and normal states. The work in Zhai and Barreto (2006) integrated GSR, BVP, skin temperature and pupil diameter measurements to detect stress using the Stroop test as an induction method. The results reported

90.1% classification accuracy; however, it is difficult to embed the proposed set of measurements in a wearable sensor due to the pupil diameter.

Some of the existing constrained environments on which the stress recognition system has been applied are a car (while driving the car) (Healey and Picard 2005), a laboratory (Sierra et al. 2011), a call center (Hernandez et al. 2011), virtual environments (Crescentini et al. 2016) and a bed (while sleeping) (Muaremi et al. 2014). Our intended environment for the detection system is the workplace. (Hovsepian et al. 2015) proposed a continuous stress assessment method using RESP (a chest-belt) and ECG sensor data for stress recognition and suggested the use of smartwatches to detect the stress. The affective and mental health monitor (AMMON) (Cheng et al. 2011) used mobile phones and speech analysis libraries to detect the stress and mental health state. However, the libraries only work on speech and are tested using an emotion corpus (Steidl 2009). “StressSense” (Lu et al. 2012) also acquires a human voice to recognize stress in real-life conversational situations. “MoodSense” (LiKamWa et al. 2011) uses Web browsing, mobile applications, phone call, e-mail, SMS and location data to infer the mood of a user. MoodSense was available only for iOS systems and utilizes “LiveLab” library (Shepard et al. 2011), but the library does not work well with default iOS factory settings. Moreover, these applications do not consider biosensor data. We employed the IAPS dataset for visual elicitation of stress for our subjects. Some of the studies have proved that pictures showing mutilations, blood and injuries can evoke stress emotion (Bradley et al. 1993, 2001; Palomba and Stegagno 1993; Palomba et al. 2000; Herbert et al. 2010). Thus, unpleasant pictures, such as scenes showing mutilated people and animals, injuries, and faces covered with blood, were chosen from the IAPS dataset.

HRV is considered to be one of the most widely used physiological sensors for stress detection (Taelman et al. 2009). Mariani et al. (2012) characterized the phases of bipolar patients using the HRV embedded sensorized t-shirt. Kim et al. (2008) presented a classification method to distinguish between low and high stress with an accuracy of 66.1%. Valenza et al. (2012) and Melillo et al. (2011) used HRV features for recognizing stress conditions using visual elicitation and situation, i.e., examination for students to induce stress. Lawanont et al. (2019) used activity trackers including the heart rate data to develop the stress recognition system based on Internet of Things (IoT) architecture. Montesinos et al. (2019) used wearable devices such as shimmer and Empatica E4 to recognize stress from users. They aimed to detect episodes of acute stress at early stages to recommend a befitting remedy. Similarly, GSR is also used as one of the physiological traits that can be employed for stress recognition (Boucsein

2012). Hernandez et al. (2011) used GSR features in the call center environment to classify stress and non-stress conditions. Setz et al. (2010) used the cognitive load to measure the stress from GSR sensors, and they achieved slightly higher than 80% detection accuracy. Arnrich et al. (2010) also used GSR sensors along with the seating pressure to measure the stress and were able to achieve over 70% accuracy. Mokhayeri et al. (2011) used multi-modal physiological signals, i.e., ECG, photoplethysmogram (PPG) and pupil diameter to classify relax and stress states. Han et al. (2017) used ECG and RSP signals to classify work-related stress and achieved 94% accuracy for the binary classification. Egilmez et al. (2017) employed multiple body and wrist-worn sensors to predict stress. They reported the prediction accuracy of 59.1% for intended stress using four wrist-worn and chest-mounted sensors.

Some of the highly related works with our proposed system that use wearable sensors are compared in Table 1 with reference to the sensors, induction method, sampling rate, and accuracy. This comparison will help in providing an insight into the contribution of the proposed work in contrast with the existing works. The studies Han et al. (2017) and Jebelli et al. (2019) support our assumption that the use of complex wearable sensors and high sampling rate yields better detection accuracies, i.e., 94.0% and 87.0%. On the other hand, employing pervasive wearable sensors (Kim et al. 2008) and low sampling rate (Egilmez et al. 2017) only, the performance decreases to 66.1% and 59.1%, respectively. Some researchers try to leverage the sensor fusion technique, i.e., using the fusion of LSR and HSR devices, to improve the stress detection performance (Healey and Picard 2005; Montesinos et al. 2019). In this paper, we categorize the LSR devices having the range of sampling rate between 5 and 25 Hz and HSR devices having the sampling rate above 25 Hz. One of the possible reasons for lower accuracy using LSR devices might be the feature engineering techniques. Most of these features were designed to record the variability using HSR devices with 250–1000 Hz (Rani et al. 2006; Jonghwa Kim and Andre 2008; van den Broek et al. 2009; Wen et al. 2014), whereas the LSR devices acquire the data at low sampling rates which makes it difficult to achieve such accuracy. Another reason is the use of longer time windows; existing studies use shorter time windows and collect the data for 1 min (Egilmez et al. 2017) to predict the stress condition. In this study, we intend to use the longer time windows (60 s) and collect the data for 3 min to detect the stress condition.

Many of the existing works in Table 1 use single independent classifiers for detecting stress except for some which employ multiple soft computing techniques, but they do not combine the results from multiple classifiers. Various studies have shown that by combining the results from

Table 1 Comparison of the existing works

Study	Sensors ^a	Induction method ^b	Sampling rate ^c	Results ^d
Setz et al. (2010)	EDA	MIST	16 Hz	82.8%
Zhai and Barreto (2006)	GSR, BVP, PD, and ST	Paced Stroop test	NP	90.1%
Kim et al. (2008)	HRV	Self-reporting	NP	66.1%
Valenza et al. (2012)	ECG	IAPS	250 Hz	NP
Arnrich et al. (2010)	GSR, Acc and pressure	MIST	25 Hz	73.75%
Mokhayeri et al. (2011)	PD, ECG and PPG	Stroop test	NP	78.5%
Healey and Picard (2005)	ECG, EMG, RSP and GSR	Self-reporting	ECG (496 Hz), GSR and RSP (31 Hz) and EMG (15.5 Hz)	97.0%
Muaremi et al. (2014)	ECG, RSP, Acc, BVP, GSR and ST	PSS (Self-reporting)	ECG (250 Hz), RSP (25 Hz), Acc (100 Hz), BVP (64 Hz), GSR (4 Hz) and ST (2 Hz)	73%
Sandulescu et al. (2015)	EDA and PPG	TSST	1000 Hz	83.08%
Zhai et al. (2005)	BVP, GSR and PD	Stroop test	BVP and GSR (360 Hz) and PD (60 Hz)	80.0%
Choi and Gutierrez-Osuna (2009)	HRV	Stroop test	500 Hz	83.0%
Wijmsman et al. (2011)	ECG, RSP, GSR and EMG	PSS and puzzles	ECG and RSP (250 Hz), GSR (100 Hz) and EMG (1000 Hz)	79.26%
Han et al. (2017)	ECG and RSP	MIST	250 Hz	94.0%
Egilmez et al. (2017)	HRV and GSR (event-based)	Intended stress/self-reported stress	5 Hz	59.2%/78.8%
Lawanont et al. (2019)	Activity tracking sensors and heart rate	PSS and GSE	NP	78.95%
Jebelli et al. (2019)	EEG sensor	Self-reporting	256 Hz	87.0%
Montesinos et al. (2019)	ECG, RSP, Acc, ST, BVP and EDA	EMDB	ECG, RSP and Acc (512 Hz), BVP (64 Hz) and EDA and ST (4 Hz)	84.13%

^aElectrodermal activity (EDA), galvanic skin response (GSR), blood volume pulse (BVP), pupil diameter (PD), skin temperature (ST), heart rate variability (HRV), electrocardiogram (ECG), accelerometer (Acc), photoplethysmography (PPG), electromyogram (EMG), electroencephalogram (EEG) and respiration (RSP)

^bMontreal imaging stress task (MIST), international affective picture system (IAPS), perceived stress scale (PSS), general self-efficacy scale, emotional movie database (EMDB) and tier social stress test (TSST)

^c Hz and NP which are defined as Hertz and Not provided, respectively

single independent classifiers results in better detection rates (Deng et al. 2012; Khowaja et al. 2017). In this regard, the main contributions of our work are to introduce a new feature set that can cope with the variability of LSR devices, considering the data from longer time windows and a new decision-level fusion method to combine the results from multiple classifiers.

3 Proposed stress detection system

This section presents our process flow for stress detection, as illustrated in Fig. 1. It includes training/testing data acquisition, data preprocessing, feature extraction from longer time windows, evaluation and selection of classification models in the training stage, and use of the classification model in the testing stage. Furthermore, we

propose a decision-level fusion method, VSBF, for combining results from two classifier models to improve the accuracy of the detection system. Each of these building blocks is explained in the later subsections.

3.1 Data acquisition

There are several methods to induce the stress emotion for a subject, and one of the popular approaches for emotion elicitation is to use an IAPS database. This database provides survey readings for valence and arousal using 1–9-point scales. Depending on Russell's model of emotion (Russell and Pratt 1980), distress is mapped as high arousal and negative valence as shown in Fig. 2. The distress term here is referred to as a severe or protracted stress condition. An android application was developed to record HRV and GSR measurements from the wearable sensors. The

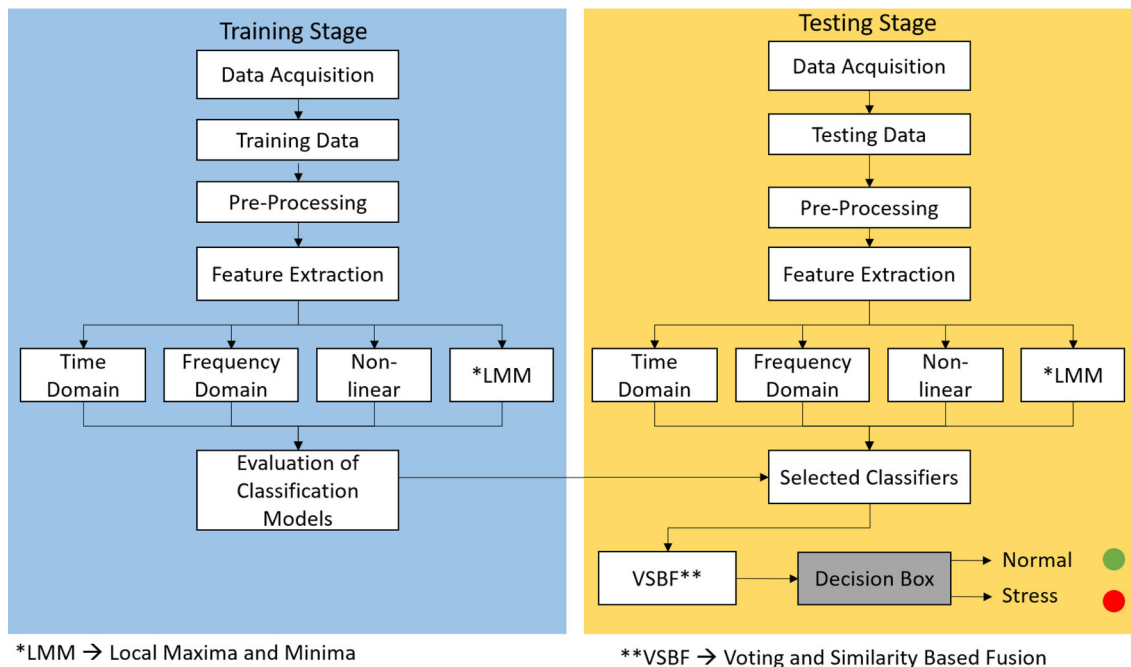


Fig. 1 Process flow of our stress detection system

measurements from these sensors were recorded as two-column readings and stored in .csv files for each subject and state. The data from each subject were labeled based on the ratings of IAPS images, accordingly. The collected data are then divided into two sets, i.e., training and testing data having 70% and 30% of the share, respectively.

3.2 Data preprocessing

The main task of this step includes the normalization of GSR values. As the values of GSR for each subject are of variable range, they should be normalized in the range between 0 and 1 for each subject before combining the data. For normalizing the values of GSR measurements, we used the min–max normalization as shown in Eq. (1), where x refers to the immediate sample value, and min and

max refer to the minimum and maximum values of GSR measurements for each subject, respectively. Once the values are normalized, the samples from all the subjects are combined for feature extraction.

$$\text{normalized}_{\text{values}} = \frac{x - \min}{\max - \min} \tag{1}$$

3.3 Feature extraction

The main objective of feature extraction is to transform the measurements into meaningful representation. We used the commonly derived features for stress recognition proposed in the related studies (Pan and Tompkins 1985). As we employ LSR devices in our study, we need to extend the time windows so that the extracted features can capture the variations between stress and normal states. Using shorter time windows for LSR devices while increasing the number of feature instances has not helped in improving the detection performance (see Table 1). A total of 85 features are extracted from each 60-s time window, and the prediction is made at every 180 s which contains a sequence of data values from HRV and GSR sensors. The features are extracted in a sliding fashion from each 60 s time window with 1 s of overlapping. Our features are categorized into two classes: 65 existing features which have been widely used in existing works and 20 LMM features that are engineered to be suited for processing data from LSR devices. The existing feature space includes 45 features from time-domain, 18 features from frequency-domain and

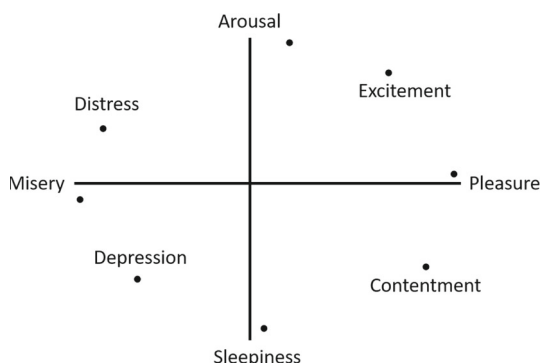


Fig. 2 Stress and normal state mapped onto the two-dimensional space of valence and arousal (Russell and Pratt 1980)

2 nonlinear features that are considered to be appropriate for our framework.

3.3.1 Existing features

3.3.1.1 Time-domain features Time-domain features are often classified into two groups, i.e., statistical and geometrical features. The statistical features are comprised of general and specific ones for HRV and GSR sensors. The former includes mean, first difference of the mean values, second difference of the mean values, standard deviation, first difference of the standard deviation values, second difference of the standard deviation values, median, covariance, interquartile range, 25th percentile, 75th percentile, skewness, kurtosis, root mean square value and minimum and maximum values. These features are generally derived from both the sensors, whereas the latter category of features includes some specific features of HRV, such as standard deviation of RR intervals (SDNN), percentage of all RR intervals which have more than 50 ms of difference (pNN50), and root mean square of successive difference (RMSSD). In addition, two variants of SDNN were computed, namely the standard deviation of the averages of RR intervals (SDANN) and the standard deviation of the i th RR interval (SDNN $_i$). Features dedicated to GSR include mean, standard deviation, minimum, maximum and median values of amplitudes and frequency responses of filtered window signals passed through fourth-order elliptic low-pass filter at 4 Hz. The most commonly derived features based on geometric characteristics of the HRV are based on the histogram of RR intervals. The first feature is the triangular interpolation of RR interval histogram (TINN), and the other is the ratio of a total number of RR intervals to the magnitude of the histogram of all RR intervals using the bin size of 1/128 s, which is known as a triangular index (TI) of HRV (Niskanen et al. 2004).

3.3.1.2 Frequency-domain features Similar to the time-domain features, frequency-domain features consist of general and specific features for both sensors. General features in this category include the number of peaks, the magnitude of the first five components, spectral peak features and spectral power features. HRV features specific to this category were computed using the Lomb periodogram method (Ruf 1999). Parameters are based on three frequency bands, i.e., high frequency (HF) 0.15–0.4 Hz, low frequency (LF) 0.04–0.15 Hz and very low frequency (VLF) 0.0–0.04 Hz. From the absolute values of these frequencies, the measures of spectral power and percentage of the sum of absolute values of high and low frequencies were computed. Subsequently, from the normalized value of the frequencies, the relative value of each power component, i.e., from HF, LF and its difference with VLF was

recorded, respectively. The ratio of LF/HF was also recorded as it is considered to be a well-known indicator of sympathovagal balance. High values of this ratio indicate the transition toward the dominance of sympathetic activity, whereas low values refer to the dominance of parasympathetic activity (Cinaz et al. 2013). A specific feature to GSR for this category includes only the signal power of the skin conductance (SCP).

3.3.1.3 Nonlinear features This category includes only specific features related to HRV as suggested in (Tulppo et al. 1996). The features are derived from the Poincare plot, which is referred to as the scatter plot of RR values of index n in the horizontal axis and RR values of index $n + 1$ in the vertical axis. The features which are computed from Poincare plots are $SD1$ and $SD2$, representing the standard deviation of long-term HRV as a major axis and that of short-term HRV as a minor axis, respectively.

3.3.2 Local maxima and minima (LMM) features

The proposed approach is based on the two-step transformation of the values. The first step transforms the raw HRV and GSR values to different probability distributions, such as geometric means (f_{geo}), Gaussian distribution (f_{Gauss}), harmonic means (f_{harmm}), extreme value distribution (f_{evd}) and central moments of fourth order (f_{cm}). The probability distributions used in this study were chosen empirically based on their performance on stress detection. However, the secondary reasons for choosing the said probability distributions are stated below. The reasons stated are strictly assumed for stress detection data.

- The geometric means might provide information regarding accruing stress levels over the period of time based on its characteristics (McNichol 2018).
- The Gaussian distribution is the most commonly used probability density function in the field of data science (Team 2017; Crooks 2019).
- The harmonic mean is a stacking of the division/multiplication layer over the geometric mean to deal with the varying periods of stress within the dataset (McNichol 2018).
- The extreme value distribution provides the likelihood of the occurrence of extreme values from the observed data within the detection period (Benstock and Cegla 2017).
- The fourth-order central moments provide the information regarding the occurrence of outliers from the observed data within the detection period (Imdadullah 2012; Chaudhary 2017).

The second step computes the local maxima and minima from the transformed values within a specified window. Let M, N , and K be the total number of samples acquired from HRV and GSR measurements, the number of sliding windows and the sliding window size, respectively. The raw data stream, $X = (X_1, \dots, X_M)$, is first transformed into the probability distribution, $S = (S_1, \dots, S_M)$, using $S_{pd,i} = f_{pd}(X_i) (1 \leq i \leq M)$ where $f_{pd}(\cdot)$ represents the transformation function based on various probability distributions. We evaluated several probability distributions, and the ones yielding good results in terms of accuracy were included in our study. The computation of each transformation is expressed in Eqs. (2)–(6):

$$S_{geo,i} = f_{geo}(X_i) = \left(\prod_{m=1}^i X_m \right)^{\frac{1}{i}}, \quad \text{for } i = 1, \dots, M \quad (2)$$

$$S_{Gauss,i} = f_{Gauss}(X_i) = \frac{1}{\sigma_M \sqrt{2\pi}} e^{-\frac{(X_i - \mu_M)^2}{2\sigma_M^2}}, \quad \text{for } i = 1, \dots, M \quad (3)$$

$$S_{harmm,i} = f_{harmm}(X_i) = \frac{i}{\sum_{m=1}^i \frac{1}{X_m}}, \quad \text{for } i = 1, \dots, M \quad (4)$$

$$S_{evd,i} = f_{evd}(X_i) = \frac{1}{\sigma_M} e^{\frac{X_i - \mu_M}{\sigma_M}} e^{-e^{\frac{X_i - \mu_M}{\sigma_M}}}, \quad \text{for } i = 1, \dots, M \quad (5)$$

$$S_{cm,i} = f_{cm}(X_i) = \frac{1}{i} \sum_{m=1}^i (X_m - \mu_M)^4, \quad \text{for } i = 1, \dots, M \quad (6)$$

In the above expressions, μ_M and σ_M represent the mean and standard deviation of the whole data stream S for the probability distribution, respectively. For the transformed data stream $S_{pd} = (S_{pd,1}, S_{pd,2}, \dots, S_{pd,M})$, in a specific probability distribution pd , we can get $N = M - K + 1$ sliding windows since the sliding interval is set to one second. For the n th sliding window having window size K , $Swin_{pd,n} = (S_{pd,n}, S_{pd,n+1}, \dots, S_{pd,n+K-1})$, local maxima and minima, $Lmax_{pd,n}$ and $Lmin_{pd,n}$ are defined as below:

Definition 1 *Local maxima* Local maxima ($Lmax$) of the n th sliding window, $Swin_{pd,n} = (S_{pd,n}, S_{pd,n+1}, \dots, S_{pd,n+K-1})$, denoted by $Lmax_{pd,n}$ is defined in Eq. (7)

$$Lmax_{pd,n} = \begin{cases} 1 & \text{if } S_{pd,n} < \text{avg}_{pd,n} \wedge \text{avg}_{pd,n} < S_{pd,n+K-1} \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } n = 1, 2, \dots, N \quad (7)$$

Definition 2 *Local minima* Local minima ($Lmin$) of the n th sliding window, $Swin_{pd,n} = (S_{pd,n}, S_{pd,n+1}, \dots, S_{pd,n+K-1})$, denoted by $Lmin_{pd,n}$ is defined in Eq. (8)

$$Lmin_{pd,n} = \begin{cases} 1 & \text{if } S_{pd,n} > \text{avg}_{pd,n} \wedge \text{avg}_{pd,n} < S_{pd,n+K-1} \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } n = 1, 2, \dots, N \quad (8)$$

where $\text{avg}_{pd,n}$ is defined by Eq. (9)

$$\text{avg}_{pd,n} = \frac{\sum_{k=1}^{K-2} S_{pd,n+k}}{K-2} \quad (9)$$

Accordingly, we get two feature vectors, $Lmax_{pd} = (Lmax_{pd,1}, \dots, Lmax_{pd,N})$ and $Lmin_{pd} = (Lmin_{pd,1}, \dots, Lmin_{pd,N})$, for each probability distribution. The feature extraction process deals with four categories of features, i.e., time-domain, frequency-domain, nonlinear, and LMM. The feature values for the first three categories which are all existing features are computed from the raw measurements of HRV and GSR sensors, whereas LMM are computed from the transformed probability distribution values.

Figure 3 illustrates the overall process for generating LMM features in our work. Each of the raw measurements and the transformed values is divided into a sequence of sliding and overlapping windows. One discrete feature value from each window is obtained using existing feature computation. Similarly, one binary value for each of $Lmax$ and $Lmin$ is obtained from each window of the transformed values. As we have the transformed values from five probability distributions and two sensor modalities, a total of 20 LMM feature values were derived. We combine 65 discrete values obtained from existing feature computations and 20 binary values from LMM to make a complete feature space. The combined feature space yields the size of $N \times 85$, suggesting that the features will be extracted for N windows. The illustration of extracting LMM features from the Gaussian distribution is given as an example.

3.4 Classification

In order to obtain the detection accuracy and reliability of the system for real environments, it is necessary to choose the classifier which shows the best performance according to the varying parameters of classification algorithms. This will prove the worthiness of the stress detection model for the implementation of the soft real-time system. In this regard, this stage will derive the best classifier which can

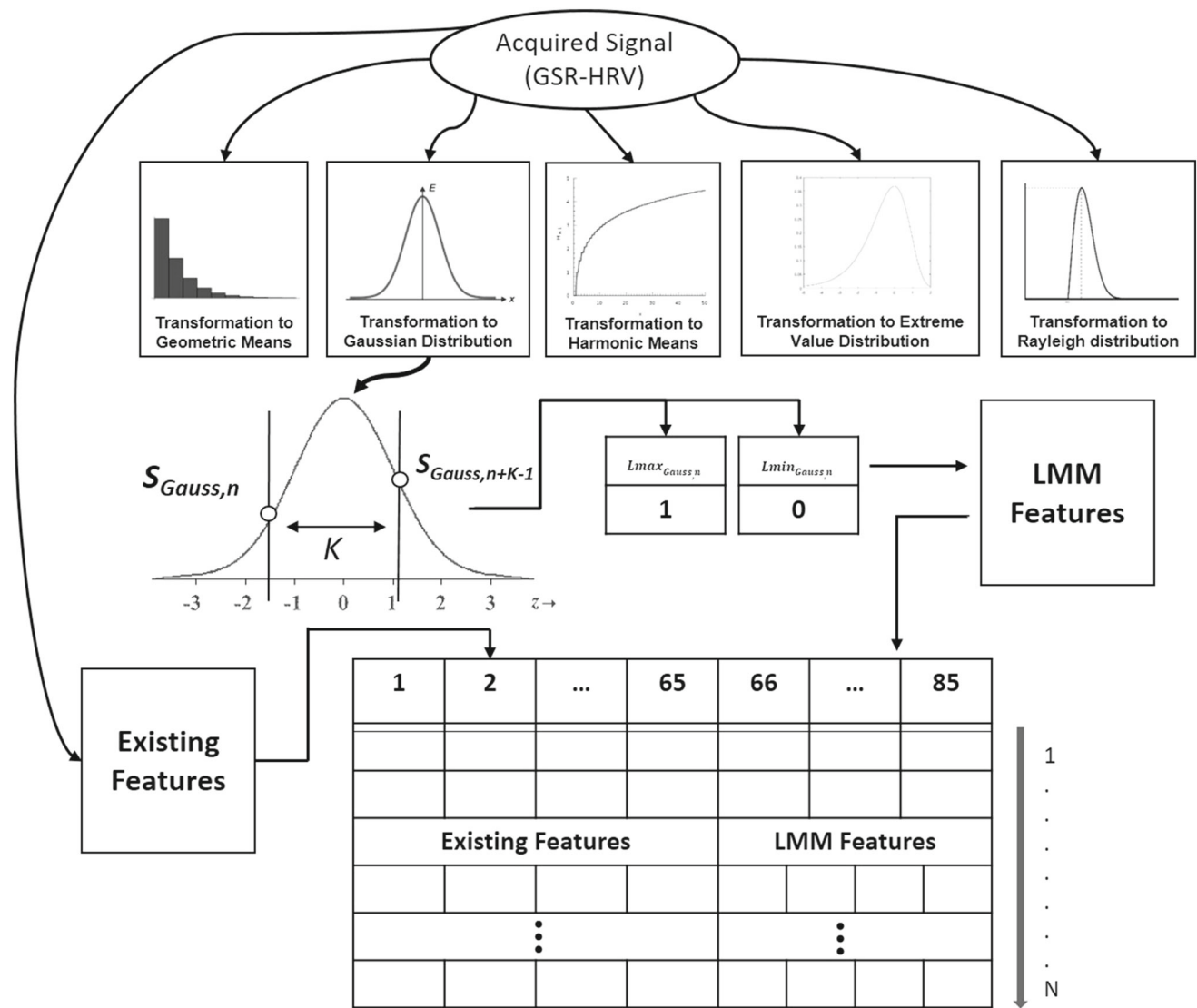


Fig. 3 Extraction of LMM and combining feature space

distinguish two physiological states, i.e., stress and normal. The classification algorithms which are analyzed and compared for the system are support vector machines (SVM) (Drucker et al. 2002), decision trees (DT) (Nakatsuka and Koishi 1995), logistic regression (LR) (Le Calve and Savoy 2000), random forest (RF) (Pal 2005) and ensemble boosting methods (EB) (Rooney et al. 2014). Each of the classification algorithms used in this study is briefly explained below.

SVM It is categorized as a large margin classifier as it finds the best hyperplane which separates two classes. The best hyperplane for SVM is the one having the maximal width of the plane between data points of the two classes. Support vectors are the points that are closest to the hyperplane separating the two classes. The linear hyperplane of SVM can be determined by Eq. (10)

$$f(x) = \langle w \cdot x \rangle + b \tag{10}$$

The variables w and x refer to the observed data point and normal vector, respectively. The sign “ $\langle \rangle$ ” represents the inner product of the two and b is the bias term. A linear hyperplane is defined by b and w , in such a way that it maximizes the margin between the samples.

DT It uses a top-down greedy approach to divide the input data space using trees. It uses the numerical process to split the data points iteratively, tests it using a cost function, and selects the best split which has the lowest cost. The sum of squared errors is often used as the cost function for all training samples as shown in Eq. (11)

$$y' = \sum (y - \text{predicted value})^2 \tag{11}$$

Variable y is the label from the training set and *predicted value* is the output from splits. Gini cost is used for finding the purity of the leaf nodes. The value of 0.5–0 characterizes the transition from worst to pure class representation. The computation for the Gini index is shown in Eq. (12)

$$G = 1 - \sum_{r=1}^R p_r^2 \tag{12}$$

where R refers to the number of classes and p_r is the ratio of instances classified as class r to the total number of instances.

LR This is conceptually similar to the linear regression, but it is used for dichotomous classification rather than the prediction. The goal of LR is to find the best fit model for a set of independent variables. LR estimates the probabilities using a cost function to choose the response levels. The cost function of LR is defined in Eq. (13)

$$J(\theta) = \frac{1}{Z} \sum_{i=1}^Z \text{Cost}(h_{\theta}(x_i), y_i) \tag{13}$$

The variables Z , y_i and x_i are the number of samples, labels and input samples, respectively. The hypothesis function $h_{\theta}(x_i)$ is the sigmoid function and is shown in Eq. (14), T refers to the transpose of weight matrix θ .

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{14}$$

RF This method constructs a number of trees for training using the input data. It selects the mode class from the outputs generated by individual trees. The main goal of RF is to reduce the variance from deep decision trees as they tend to overfit the data. To classify a given instance, RF generates multiple decision trees and chooses the output or label having the most votes. The generation of multiple trees is based on independent and identically distributed random vectors. The margin for RF is computed as shown in Eq. (15)

$$\text{margin}(X, y) = \text{av}_k I(h_k(X) = y) - \max_{j \neq y} \text{av}_k I(h_k(X) = j). \tag{15}$$

where $I(\cdot)$ is an indicator function, $h_k(X)$ is the hypothesis function, y is the label, and j is the mismatched label, respectively. The margin is computed iteratively until the average votes for one class exceed the average votes of another class. The average votes are represented by av_k .

AdaBoost Ensemble boosting is a method of combining weak classifiers using a set of prediction rules and assigns weights to them so that their prediction error gets decreased. Adaptive boosting (AdaBoost) (Freund and Schapire 1997) does the same. It tweaks the weak learners

which have misclassified the instances so that their error rate gets smaller, at least than a random guess, i.e., 0.5, for binary classification. The computation for AdaBoost can be given in Eq. (16)

$$f(x) = \sum_{r=1}^R \alpha_r h_r(x) \tag{16}$$

where $h_r(x)$ is the weak learner, α_r is the assigned weight, and R represents the number of classifiers. The main aim of the algorithm is to reduce the sum of training errors by taking into account the previous boosted classifier’s training error as shown in Eq. (17)

$$E_r = \sum_i E[f_{r-1}(x_i) + \alpha_r h(x_i)] \tag{17}$$

where E_r is the resulting error of r -stage classifier and $E[f_{r-1}(x_i)]$ is the error of the current training set on the sample.

GentleBoost Gentle AdaBoost (GentleBoost) (Friedman et al. 2000) is a variant of AdaBoost and works in a very similar way. The problem of AdaBoost is that it is very susceptible to noise. GentleBoost overcomes this problem by using a different cost function from AdaBoost. The remaining structure of the classification is the same as that of AdaBoost. The cost function is shown in Eq. (18)

$$f(x) = \sum_1^N d_n^t (y_n - h_t(x_n))^2 \tag{18}$$

where d_n^t are the observation weights at every step t and $h_t(x_n)$ are the hypotheses function from the regression model fitted to the target labels y_n .

As it is mentioned in the former subsection that we used 180 s data with a sliding window of 60 s and an overlapping window of 1 s for stress state detection, we obtain the resultant feature space of 121×85 from $121 = (180 - 60 + 1)$ windows. Each window with 60 samples is classified independently into stress or normal state. The probabilities of both states, $P(0)$ and $P(1)$, from the classification results are computed as $P(0) = \frac{\text{Number of windows classified as stress}}{121}$ and $P(1) = \frac{\text{Number of windows classified as normal}}{121}$. The state having a higher probability will be considered as the final decision.

3.5 Voting and similarity-based fusion (VSBF)

To further improve the classification accuracy of our stress detection system, we combine the results from two classifiers using a decision-level fusion method which elaborates the characteristics of voting and similarity measures. Figure 4 shows the complete process of our VSBF method. Let N and N' be the number of training and test set samples,

respectively. The test feature set is first classified using two individual classifiers and the average probability of each state is computed, accordingly. The “Classifier 1” and “Classifier 2” as shown in Fig. 4 are the classifiers with best recognition accuracies which will be determined when performing analysis on the acquired dataset. We assume that using classifiers with the best accuracies but with different characteristics can help improve the stress detection performance. Random N' samples based on LMM features from the training feature set having N samples for both states are extracted. Similarly, only LMM features from the test feature set are extracted to evaluate the similarity between them. The similarity values between the test feature set and the training feature set for stress and normal states are computed, respectively. The output label will be derived using the proposed VSBF function. As LMM are binary features, we employed the Jaccard similarity measure (Jaccard 1901) for VSBF. Though there are many other similarity measures for binary features (Choi 2008), the Jaccard similarity measure is widely used among all others (Choi et al. 2010). The computation for Jaccard similarity is given in Eq. (19)

$$\text{Sim}_{\text{jaccard}} = \frac{a}{a + b + c} \tag{19}$$

The definition of the variables in Eq. (19) is expressed in a 2×2 contingency table by Operational Taxonomic Units (OTU) (Dunn and Everitt 2004). The contingency table is shown in Table 2. The binary features in this table are represented by i and j , respectively. Variable a

represents the positive match suggesting that both values for i and j are 1. Variables b and c refer to the “ i th mismatch” and “ j th mismatch” represented by \bar{i} and \bar{j} , respectively. Variable d is the number of attributes having negative mismatched suggesting that both values for i and j are 0. The sum of all the matches and mismatches for all windows is represented by N .

The classification probabilities from both classifiers can be obtained using the probability expressions mentioned in the previous subsection. Considering two probability values for each state from both classifiers, we get the average probabilities, $P_{\text{avg}}(0) = \frac{P_{\text{classifier1}}(0) + P_{\text{classifier2}}(0)}{2}$ for stress state and $P_{\text{avg}}(1) = \frac{P_{\text{classifier1}}(1) + P_{\text{classifier2}}(1)}{2}$ for the normal state. The VSBF for drawing the final classification result is defined as follows.

Definition 3 *Voting and Similarity-based function (VSBF)* Voting and similarity-based function (VSBF) for stress state (0) and normal state (1), denoted by $\text{VSBF}(P_{\text{avg}}, \text{Sim}_{\text{jaccard}})$, is defined in Eq. (20):

$$\text{VSBF}(P_{\text{avg}}, \text{Sim}_{\text{jaccard}}) = \begin{cases} 0 & \text{if } P_{\text{avg}}(0) * \text{Sim}_{\text{jaccard}}(0) > P_{\text{avg}}(1) * \text{Sim}_{\text{jaccard}}(1) \\ 1 & \text{otherwise} \end{cases} \tag{20}$$

The performance of the classification algorithm is evaluated in terms of accuracy and F_1 scores. Evaluation

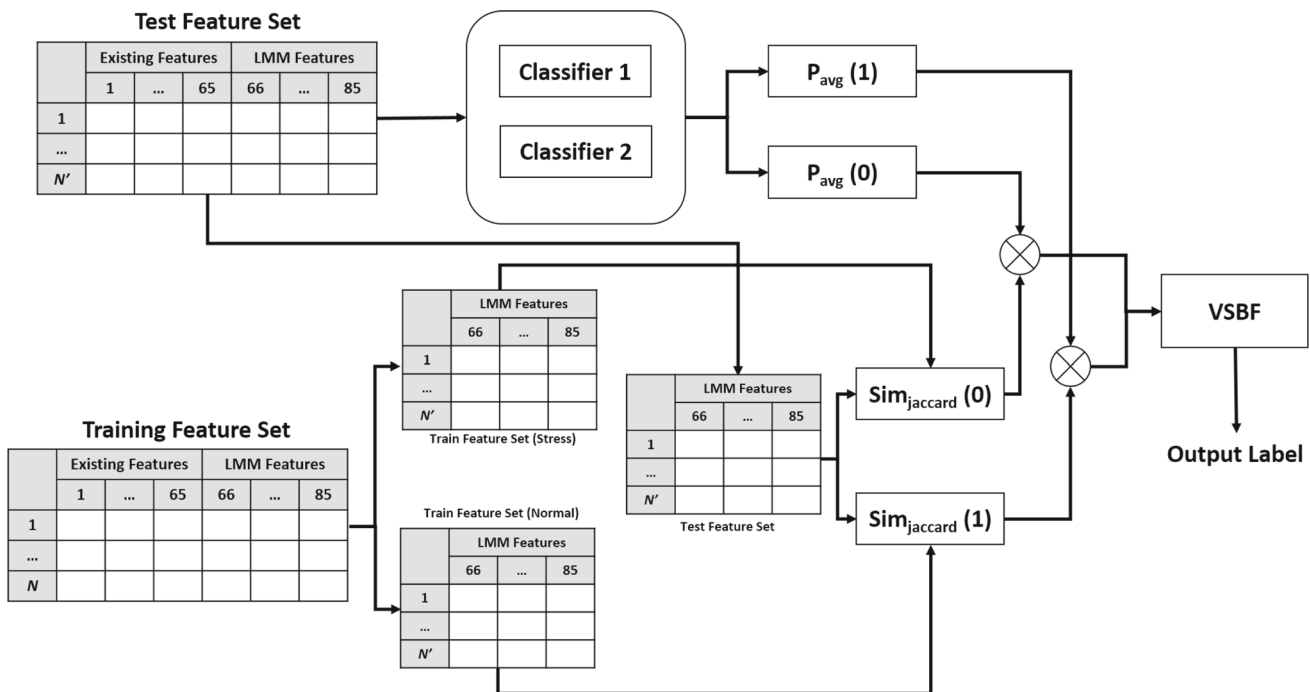


Fig. 4 Our proposed decision-level fusion method: VSBF

Table 2 OTUs expression of similarity measure for binary features (Dunn and Everitt 2004)

	1 (Presence)	0 (Absence)	Sum
1 (Presence)	$a = i \cdot j$	$b = \bar{i} \cdot j$	$a + b$
0 (Absence)	$c = i \cdot \bar{j}$	$d = \bar{i} \cdot \bar{j}$	$c + d$
Sum	$a + c$	$b + d$	$N = a + b + c + d$

parameters of both terms can be derived from prediction results. The computation for evaluation parameters is shown in Eq. (21)–(24):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \quad (21)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (22)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (23)$$

$$F_1 \text{ score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (24)$$

In the above computations, true positive (TP) refers to the positively predicted outcomes which actually belong to a positive class. True negative (TN) is the negatively predicted outcome that belongs to a negative class, while false positives (FP) and false negatives (FN) refer to the positively predicted outcome belonging to a negative class and the negatively predicted outcome belonging to a positive class, respectively.

4 Experiments and results

This section presents the experimental analyses to prove the effectiveness of the proposed method that uses LMM features and the VSBF method for stress detection. The experiments are carried out using two kinds of datasets. The first analysis is performed to evaluate the proposed feature set and our decision-level fusion method using the acquired dataset which is collected from LSR wrist-worn devices for workplace environments. The second analysis is done to evaluate the performance of the proposed method using the driveDB dataset which is a publicly available dataset and is suitable for comparison with existing studies.

4.1 Evaluation of LMM features and VSBF method using the acquired dataset

In general, groups of working can be classified into two: white collar and blue collar. Those two categorizations are commonly reflecting the type of occupations. White collar

is typically related to the workers who work behind a desk in a service industry. Researcher in a laboratory, clerk, and managerial-task-related person are some of the examples. Meanwhile, blue collar refers to the workers who engage in physical task such as construction workers, manufacturing or factory worker. The two different classes of participants (i.e., factory workers and graduate students) could be the samples to represent the work categorization. In this regard, a total of 14 healthy participants, 9 males and 5 females having ages 20–38 were considered for the data recording. Seven of the participants were graduate students, while 1 participant was a professor and 6 participants were factory workers. The reason for considering the samples from both the graduate students and the factory workers for designing the stress recognition system for human workspaces is threefold. Firstly, it is well established that the graduate students perceive higher stress levels and are quite vulnerable to such conditions in comparison with others (Johnson et al. 2008; Bhui et al. 2016; Levecque et al. 2017; Pain 2018). Secondly, as suggested in Johnson et al. (2008), Romer (2011), Bhui et al. (2016, 2019), the reasons for stress induction such as effort-reward imbalance, low salaries, excessive workloads, unclear performance expectations, low decision latitude, unfair treatment, lack of support, and unrealistic demands, are common for both factory workers and the graduate students. Thirdly, the survey from European Agency for Safety and Health at Work (Occupational Safety and Health) along with others has shown that the most problematic stress can be noticed in the people working in Education, Business Services, Financial Sector, Construction, and Telecommunication sectors (Conditions 2006; Milczarek et al. 2009; Romer 2011). Moreover, the people work in the above-mentioned sectors exhibit the common stressors which are workload, self-defeating beliefs and fear of conflict along with the stress indicators including burnout, absenteeism, insomnia, cardiovascular diseases and frequent interpersonal conflicts (SERV 2004; Conditions 2006; Billehoj 2007; Milczarek et al. 2009; Romer 2011; Thorsteinsson et al. 2014).

The intention of this study was clearly notified to the participants, informed consent was obtained, and the participants were asked to sign the research confidential agreement accordingly. The participants were allowed to stop the video during the emotion elicitation process if they feel the video content is highly intense or offensive. However, none of the participants left the video in between. A smartwatch was employed for acquiring the physiological measurements from HRV and GSR sensors, and a smartphone was used for the data collection using the self-designed android mobile application. The detailed process of data acquisition is illustrated in Fig. 5. IAPS is used as the video elicitation method for inducing stress

emotion. Many previous studies have used this database as an effective method for inducing the stress situation which is mapped as negative valence and high arousal on Russell’s model of emotion. The subjects were asked to wear the smartwatch having HRV and GSR sensors. The data recording was started after 2–3 min so that the smartwatch gets synchronized with the user’s skin and our android application. Depending on the smartwatch, it takes a while for the GSR values to be consistent, i.e., between 100 and 1200 kohm range. The data collection procedure started with a non-stressed (normal) video (compilation of IAPS images) followed by a black screen along with the 5-min break and simple IQ/math questions. The black screen and IQ/math questions are a part of the data collection protocol to normalize the emotion which has been followed by many existing studies (Valenza et al. 2012; Liu and Sourina 2014; Han et al. 2017; Khowaja et al. 2018b). The data recording for both states started after 10 s from the start of the video.

All participants were allowed to answer within 15 s, and subsequently, they were again asked to watch the stress state video starting with the black screen. The challenge of using the employed wearable device is the limitation of the sampling rate as it only provides 1 Hz for HRV. The characteristics of the video clip used for stress state elicitation are shown in Table 3.

We used various classification models for predicting stress and normal states. Six machine learning algorithms have been used to perform the analysis such as SVM, DT, LR, RF and EB (AdaBoost and GentleBoost). For more reliable investigation, we performed the analysis five times

with varying testing and training sets. Apart from varying parameters for the classification algorithms, we also applied the feature selection method, i.e., principal component analysis (PCA) (Subasi and Ismail Gursoy 2010). If the feature space compels the classification algorithm to overfit the data, then the feature selection method can help to reduce the variance. The PCA method is used for all the classification algorithms accordingly. The analyses shown in Tables 4, 5, 6 and 7 provide the average accuracies along with the parameters used for specific algorithms. It should be noted that we performed the analysis with various parameters of a specific learning algorithm; however, we only provide the results which achieved the best average accuracy from five trials with a specified set of parameters. Table 4 shows the results using DT. The constraints employed for this algorithm are split criterion, tree depth and the surrogating decision of splits. Standard

Table 3 Characteristics of stress state elicitation video clip

Characteristics	Description
Subjects	14
Clips	Stress and Normal
Duration of clips	4 min 25 s (Stress) 4 min 15 s (Normal)
The rate of change of pictures	5 s per picture
Number of pictures	53 images (Stress) 51 images (Normal)

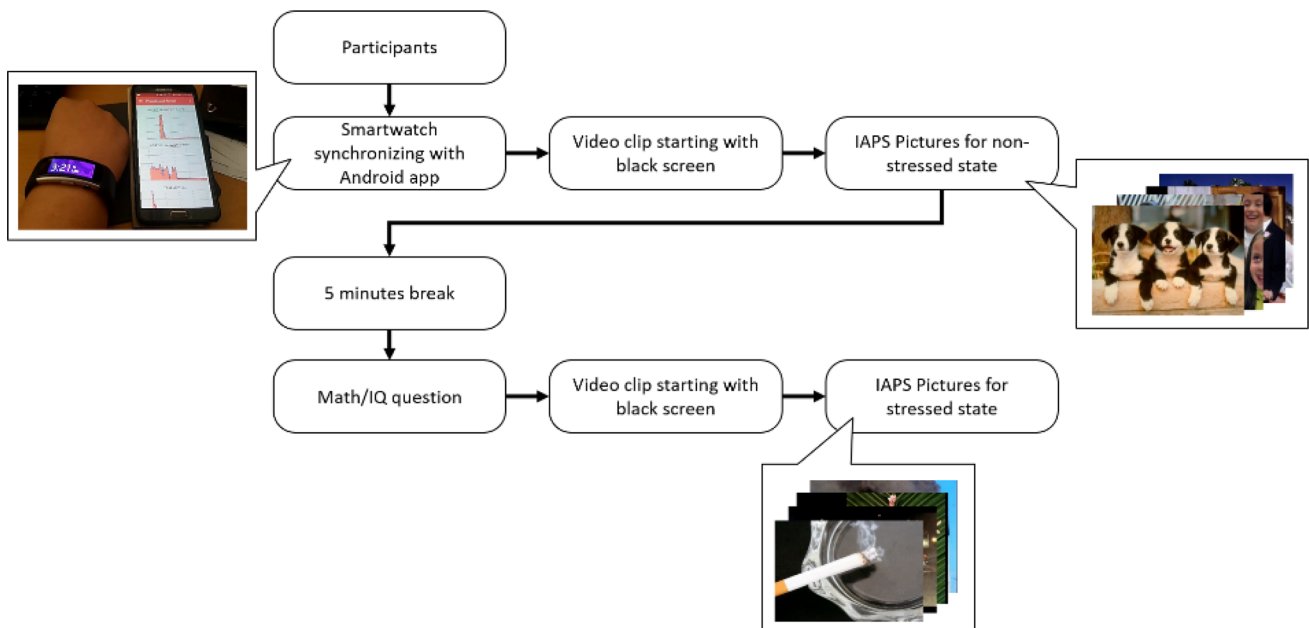


Fig. 5 Proposed data acquisition protocol

Table 4 Analysis using DT

DT model	Split criterion	Tree depth	Surrogate	PCA	Accuracy
DT1	Deviance	100	Off	Off	77.84
DT2	GINI index	100	Off	On	74.28
DT3	Deviance	100	Off	On	76.54

Table 5 Analysis by using LR and SVM

Method	Kernel	Kernel scale	Cost	Prior	Outlier fraction	PCA	Accuracy
LR	–	–	–	–	0	–	75.57
SVM1	Gaussian	1	–	Empirical	0	On	77.73
SVM2	Gaussian	0.5	–	Empirical	0	On	78.17
SVM3	Gaussian	27	[0 3; 1 0]	Uniform	0.05	Off	78.98

Table 6 Analysis using RF models

RF model	Max split	# of learners	PCA	Accuracy
RF1	20	30	On	79.55
RF2	20	50	On	76.73
RF3	20	100	On	75.64

classification and regression tree (CART) (Brezigar-Masten and Masten 2012) was employed for training the model. The split criterion is used to avoid overfitting. The methods used for it include the Gini index and maximum deviance reduction (also referred to as cross-entropy). The surrogating decision of splits measures the association between predictors for splitting rules. Table 5 shows the analysis results using LR and SVM, where there were no varying parameters used in the case of LR and it is also a kind of linear classifier. Therefore, the results of LR are merged with SVM. The constraints of SVM are defined as:

- Kernel function: to compute the inner product of the transformed predictors using this function.
- Kernel scale: to refer to the spread of the function which applies the appropriate kernel norm to compute the inner product of transformed predictors.
- Misclassification cost: to be used for penalizing the prior probabilities and to be incorporated in the form of a matrix.

Table 7 Analysis using EB models

EB method	Max splits	# of learners	Learning rate	PCA	Accuracy
AdaBoost1	12	30	0.1	Off	86.36
AdaBoost2	12	30	0.01	Off	86.09
GentleBoost1	20	30	0.1	Off	82.93
GentleBoost2	12	30	0.1	Off	84.66

- Outlier fraction: to assume a specific set of a portion of the training data as outliers.

Tables 6 and 7 show the analysis results using RF and EB, respectively. The common constraints for both methods are the maximum number of splits and the number of learners. The former controls the depth of the tree while the latter refers to the number of weak learners to be trained using ensemble methods. An additional constraint in the ensemble method is the learning rate which is commonly interchanged with the term step size. It determines the amount of newly acquired information to be overridden to the old information.

The qualitative results for the accuracy of these models are shown in Fig. 6. The experiment for each classifier was repeated 5 times with varying test set. The visualization is important to analyze the effect of parameters not only for achieving high accuracy but also for maintaining the consistency. DT achieves lower accuracy and fails to maintain consistency in obtaining the same results. Meanwhile, RF achieves relatively higher accuracy than SVM but does not exhibit consistent performance. SVMs, on the other hand, reveals a good consistency in obtaining the same performance but achieves lower accuracy than EB methods. GentleBoost exhibits better consistency, and AdaBoost achieves the highest accuracy as shown in Fig. 6. To apply our VSBF method, we need to choose the effective combination of the classifiers. The selection of classifiers was made on the basis of combining performance (higher accuracy) and consistency (maintaining the same level of

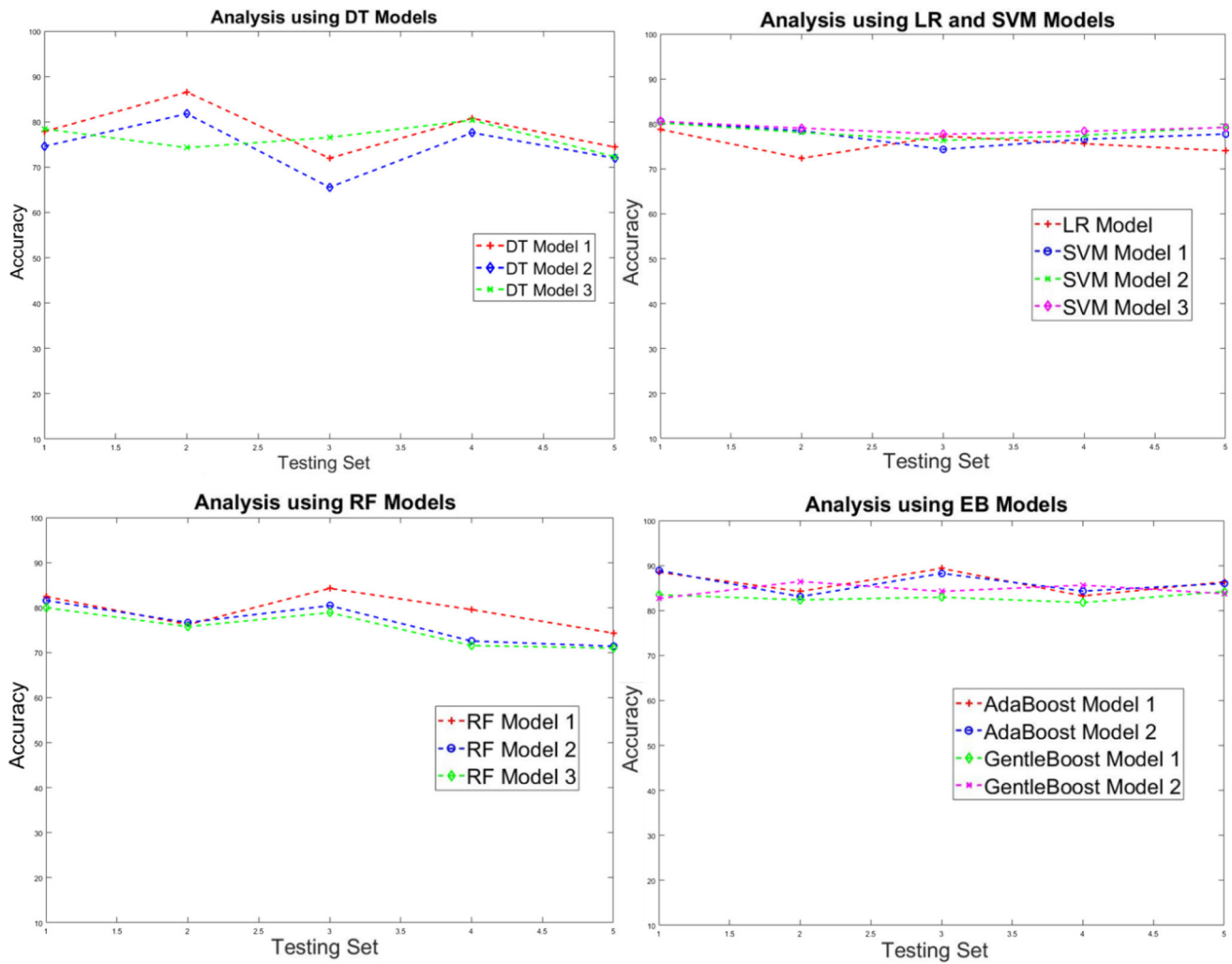


Fig. 6 Visualization of accuracies using different classification models

accuracy). As a result, AdaBoost and GentleBoost are qualified for the combination schema, but they both belong to the ensemble-based classification method. Intuitively, selecting the classifiers having the same characteristics might not improve the detection rates as their classification result will be mostly similar for individual samples. In this regard, we chose AdaBoost and SVM for the combination. The classification results from both the classifiers will be used for VSBF, our decision-level fusion method.

Table 8 shows the accuracy and F_1 scores (Becker et al. 2017) of each classification model with and without LMM features. The results show that by using LMM features, the detection accuracy is improved by at least 6.7% (LR) and up to 14.74% (SVM). It is evident from the analysis that LMM features have a positive impact on classifying the stress and normal states and are capable of capturing the variations even with LSR devices. SVM shows better performance in terms of F_1 scores compared to the RF model, which also supports our decision for selecting the

classifier combination. Table 8 also presents the accuracy and F_1 score of our VSBF method. The results can only be computed using LMM features as the VSBF method computes the similarity from binary features and the existing features do not have the binary characteristics. The results reveal that our proposed fusion method can improve the detection accuracy by 5.69% and 15.23% accuracy compared to the AdaBoost1, the best performer, with and without LMM features, respectively. These statistics prove that the proposed system has the ability to effectively detect the stress condition from LSR wrist-worn devices.

Figure 7 shows the receiver operating characteristics (ROC) curves for different classification models and our VSBF method. The curves demonstrate the relationship between false positive and true positive rates. Based on the analysis, an ensemble method, AdaBoost, proves to be the best model for detecting stress state in terms of accuracy and F_1 score since it has the least distant curve from the reference point (0, 1). Meanwhile, GentleBoost proves to

Table 8 Comparison of models with and without new features

Classification model	With new features		Without new features	
	Accuracy (%)	F_1 score	Accuracy	F_1 score
DT1	77.84	0.7194	70.86%	0.6232
LR	75.57	0.6861	68.87%	0.6043
SVM3	78.98	0.7448	64.24%	0.5630
RF1	79.55	0.7429	69.89%	0.6015
AdaBoost1	86.36	0.8378	76.82%	0.7050
GentleBoost2	84.66	0.8163	75.50%	0.6906
VSBF	92.05	0.9041	–	–

be the second-best classifier and has a better result compared to DT, SVM and RF. Our decision-level VSBF method outperforms all the independent classifiers by achieving the least distant curve from the reference point.

4.2 Evaluation of LMM features using the driveDB dataset

The second analysis is performed for validating the effectiveness of the proposed method using the publicly available dataset, driveDB. There are three main reasons for choosing this dataset for validation. First, a lot of existing studies have used this dataset for stress recognition. Secondly, it offers the raw readings from the same set of sensors, we used in our proposed stress detection system. Third, some of the features used for this dataset are also used in our proposed detection system. We have conducted a leave-one-subject-out (LOSO) analysis on the driveDB dataset which is compliant with the existing studies, suggesting that 13 subject’s data will be used for training and 1 subject’s data will be used for the testing purpose. By doing

so, the performance of our LMM features can be validated in terms of adaptability to new subjects. The original dataset consists of multimodal physiological data from ECG, RESP, HR, EMG and two GSR (placed on foot and palm of the left hand) sensors. Since our study mainly focuses on HRV and GSR sensors, we performed analyses using ECG and GSR (placed on the palm of the left hand) sensor measurements. The measurements from both sensor modalities are acquired with a sampling rate of 496 Hz and 31 Hz, respectively; thus, the sensors are categorized as HSR devices. An ECG sensor from the driveDB is used because the signals from this sensor can be used to derive HRV measurements. As the constrained environment of this dataset is different from our application, we need to perform the data preprocessing which is explained as below with the characteristics of the dataset.

The driving tasks comprise six sections: start of the driving (rest condition), driving through city before transiting to highway (city condition), driving through highway between first two tolls (highway condition), driving through highway between next two tolls (highway

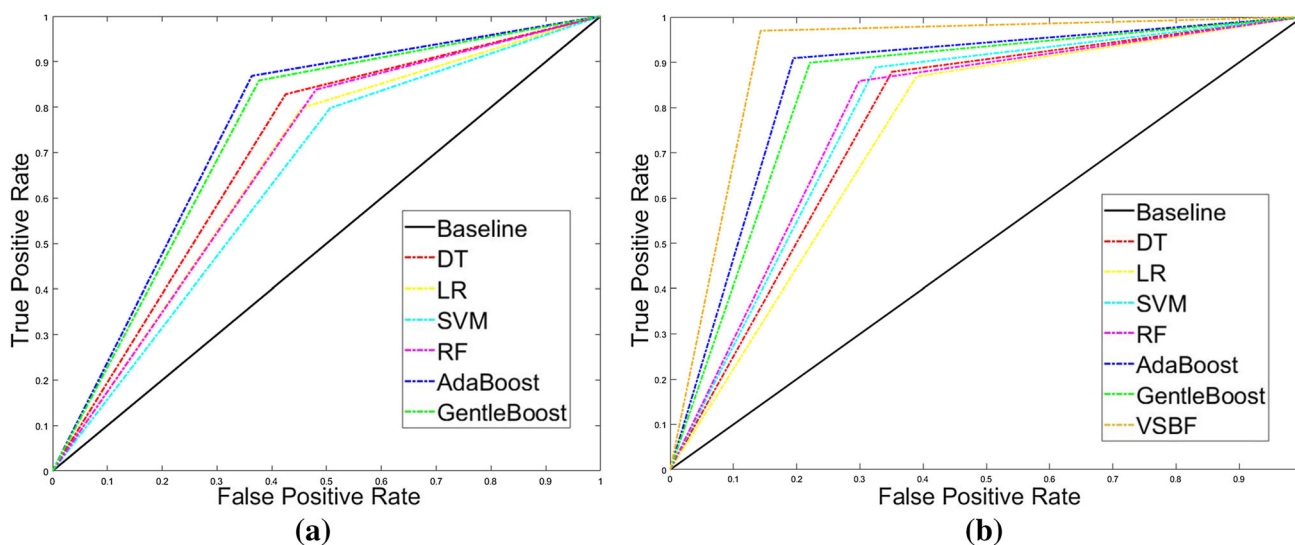


Fig. 7 ROC curves **a** without and **b** with using new features

condition), driving through city after the drive on highway (city condition) and end of driving (rest condition). Based on the traffic and road conditions, the total duration of the drive varied from 50 to 90 min. The rest periods were considered to be the baseline for the overall driving process. The available dataset is from 17 subjects, but we consider only 14 subjects in our study as the data record for one subject is incomplete and the data records for two subjects cannot be analyzed due to the unavailability of markers for the driving condition. The driveDB dataset only contains the data from physiological measures without the corresponding labels. However, from the literature (Healey and Picard 2005), it could be validated that the drive-through city, highway and resting periods yield high, medium and low levels of stress, respectively. Well-defined data samples were taken from each driving test, and the data were evenly distributed among the three driving conditions. Resting period segments were captured from the last 5 min, while high-stress segments were taken when the driver entered into the city and the signals showed high variations as compared to the resting state. Medium stress segments were taken when the subjects were driving on the highway between two tolls. The data of subject 1 from ECG and GSR sensor measurements annotated with the driving tasks are shown in Fig. 8. The recordings were divided into 100 s of sliding windows with 10 s of overlapping windows as proposed in (Healey and Picard 2005). In order to perform a fair comparison with the existing methods, we used the same set of features that have been proposed in the previous studies (Chen et al. 2017a).

A total of 53 features were included for the analysis of this dataset. Sixteen and 17 features out of those were extracted from GSR and ECG, respectively, using the existing method as proposed in Chen et al. (2017a). Additionally, 10 features for each sensor measurement are extracted using our proposed feature set. For reading the data and extracting the existing features from the driveDB dataset, we used the WFDB Toolbox for MATLAB (Silva and Moody 2014).

The performance of classification models for the driveDB dataset is assessed in terms of accuracy and ROC curves. The data represent three driving conditions, i.e., rest, highway, and city, yielding low, medium and high stress levels, respectively. These driving conditions were divided into three cases for performing the binary classification which includes rest condition versus others, highway condition versus others and city condition versus others, respectively. The results from the three cases are displayed in terms of accuracy in Table 9.

The same combination of classifiers from the previous experimental analysis was chosen for the VSBF method. Since none of the features from the existing feature set had binary characteristics, the performance of the VSBF

method on the existing feature set was not recorded. The classification model with the highest accuracy was AdaBoost, achieving 96.02%, 95.27% and 95.46% for individual cases and 95.58% on average, respectively. The proposed VSBF method achieves the best results with 98.86%, 98.11%, and 98.48% for individual cases and 98.48% on average, respectively. Figure 9 shows the ROC curves for rest versus other, highway versus other and city versus other conditions without and with the LMM features. It can be noticed that the LMM features contribute to better performance in detecting stress. Ensemble boosting methods (AdaBoost and GentleBoost) perform better than other classification algorithms, and the VSBF method achieves the lowest distant curve from the reference point.

The analysis conducted on the driveDB dataset proves the effectiveness of the proposed LMM features as they increase the accuracy at a maximum of 4.7% (LR) and on average by 3.49%. Table 9 also shows that the VSBF method increases the average accuracy by 2.9% and 7.01% compared to the AB, the best performer, with and without LMM features, respectively. The results from the LMM features and the VSBF method indicate an increasing trend in terms of detection accuracy and also prove its applicability to HSR devices achieving higher detection accuracy.

5 Implementation of soft real-time stress detection system

According to the analysis on stress detection, we implemented a soft real-time detection system which was designed to be compatible with the usage in real-life workplace environments in term of less intrusive and pervasive device choice (i.e., Microsoft Band 2 and Android-based smartphone). The conceptual diagram of our system is depicted in Fig. 10. The HRV and GSR sensors are equipped with Microsoft Band 2. It also provides a standard development kit (SDK) to transmit HRV and GSR data to a smartphone through Bluetooth low energy 4.0 protocol.

Microsoft Band 2 records HRV and GSR data in different sampling rates (i.e., 1 Hz for HRV and 5 Hz for GSR). Therefore, prior to transferring them to the server, their sampling rates need to be matched. In this regard, we down-sample the data from GSR to 1 Hz by averaging the corresponding values. Synchronized HRV and GSR data are then transferred to the server by the smartphone through a hypertext transfer protocol (HTTP) connection. The transmission rate is 180 rows per min. The passed data are then stored in a queue-structured dataset buffer with 180 slots. Each slot holds a one-row vector of HRV and GSR data. This buffer is updated continuously as new data arrive (i.e., the new data are appended to the queue, while

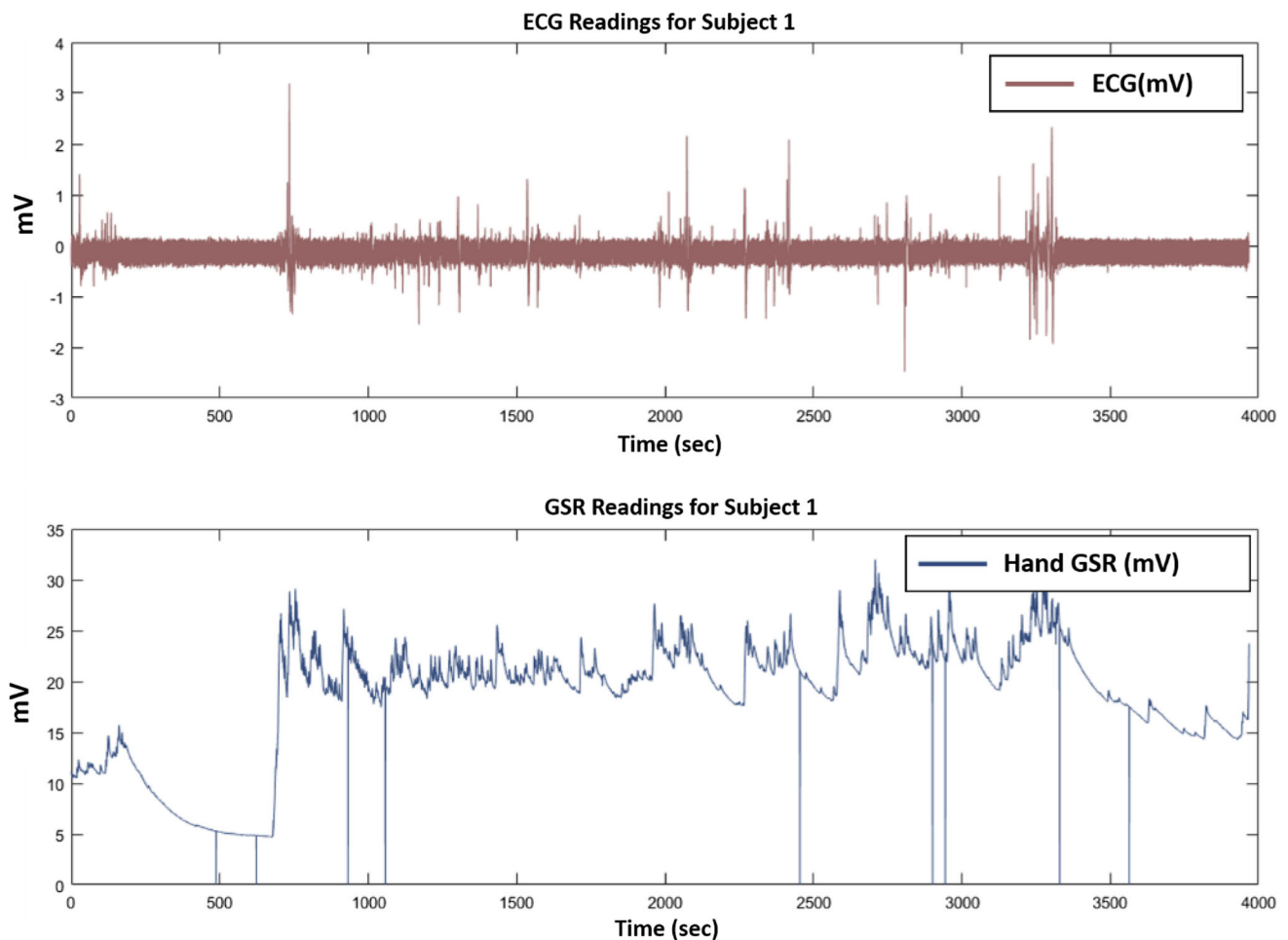


Fig. 8 Annotated sensor measurements for subject 1

Table 9 Stress detection accuracy from the driveDB dataset

Classification model	With LMM features			Without LMM features		
	Rest versus all (%)	Highway versus all (%)	City versus all (%)	Rest versus all	Highway versus all	City versus all
DT	90.34	86.74	88.64	88.32%	85.15%	86.90%
LR	85.23	85.42	86.36	80.73%	79.31%	82.88%
SVM	92.05	93.46	93.18	91.10%	89.17%	89.29%
RF	92.61	92.42	92.42	90.02%	88.32%	89.23%
AB	96.02	95.27	95.46	91.08%	91.84%	91.50%
GB	94.89	94.51	94.70	91.27%	90.25%	90.53%
VSBF	98.86	98.11	98.48	–	–	–

LMM local maxima and minima, DT decision trees, LR logistic regression, SVM support vector machines, RF random forest, AB AdaBoost, GB GentleBoost, VSBF voting and similarity-based fusion

the oldest ones are removed). In parallel, using the data, the server performs stress evaluation tasks by performing feature extraction and followed by classification.

The stress evaluation is done periodically in every 180 s. We implemented the same set of classification

algorithms using the Waikato Environment for Knowledge Analysis (Weka) framework (Hall et al. 2009), and all the APIs are implemented using the Java Server Page technology. After classification, the result is stored on the server. On the client side, the android application fetches

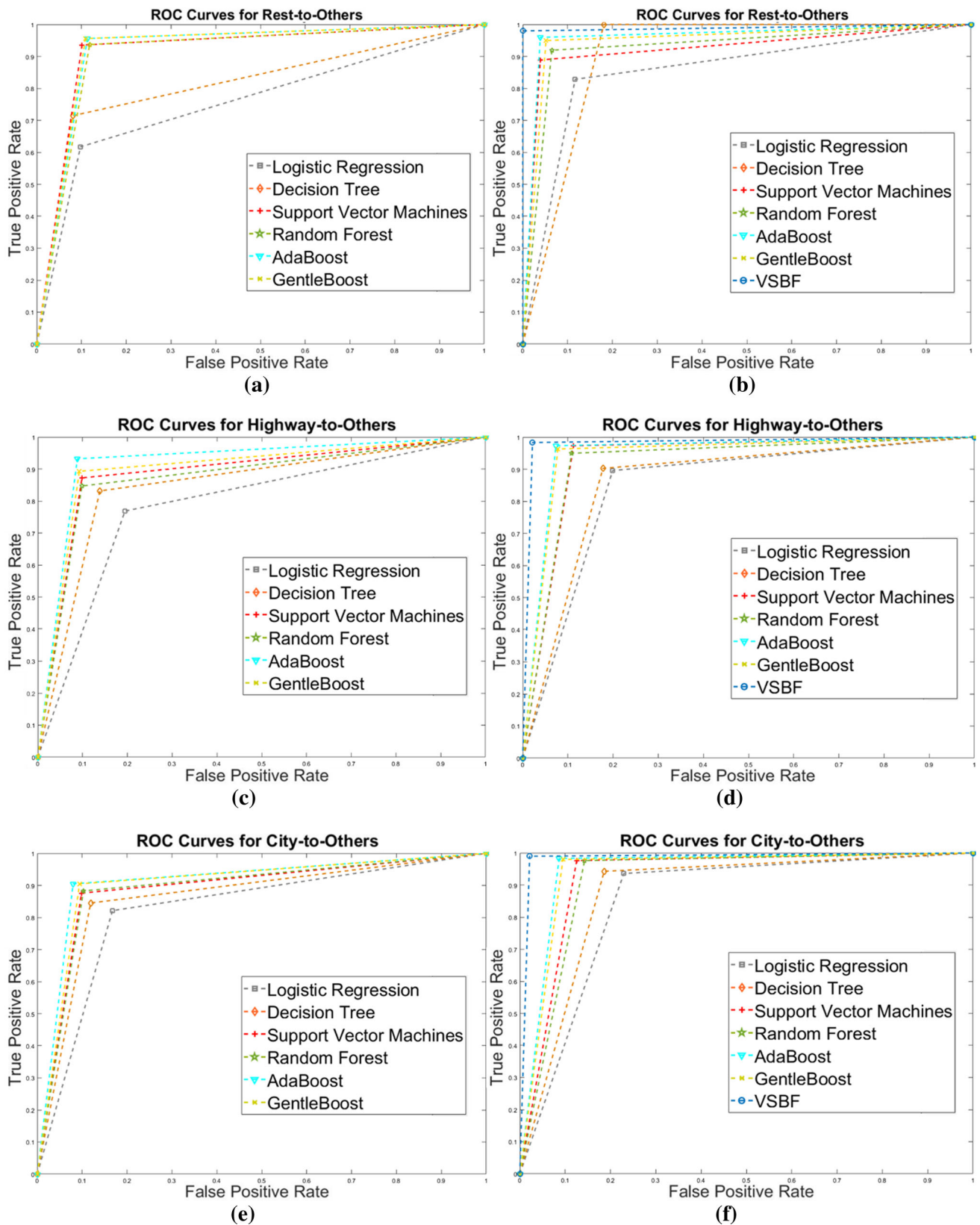


Fig. 9 ROC curves for **a** rest-to-others without LMM features. **b** Rest-to-others with LMM features. **c** Highway-to-others without LMM features. **d** Highway-to-others with LMM features. **e** City-to-others without LMM features and **f** City-to-others with LMM features

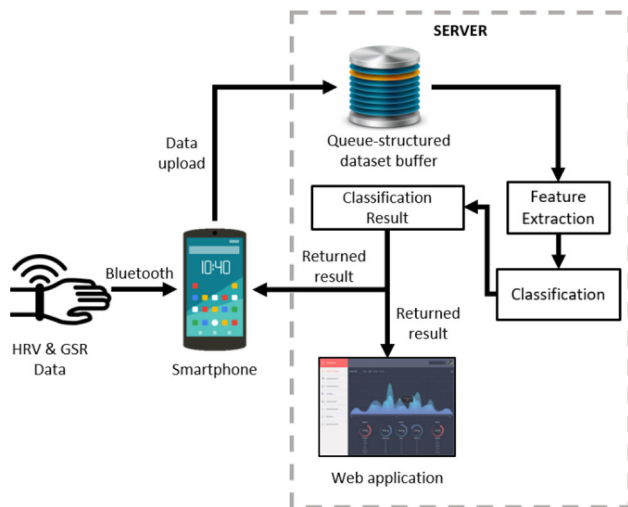


Fig. 10 Conceptual diagram of the real-time stress detection

the classification result in the server and displays it on the screen for users, while on the server side, a web page also displays the classification result for administrators or stakeholders as shown in Fig. 11. Following the notification of the web application, the administrators can report and arrange an emergency response in time for the affected individual. In addition, the web application provides a stress log in a bar chart as shown in Fig. 12. The web application depicts the results of our stress detection system for a whole day. Hence, it provides details at which time the person feels the most stressed. The results from the log can provide helpful insights and can be used in applications such as recommender systems and human behavior analyses. The x-axis indicates time scales in a day, while the y-axis indicates an amount of time in minutes a user feels stressed.

The web application also displays additional information which might be useful for stakeholders, such as location, skin temperature, outdoor temperature and weather condition. For the stress recognition system, the classification interval might be in minutes since human emotional

changes will not frequently occur in seconds. Time delay should be considered in the data transmitting process between sensors on the mobile phone.

Delay is one of the important indicators to measure network transmission capability and real-time system performance. (Chen 2012) identified several factors that affect delay in wireless sensor networks, such as packet size, physical environment and communication environment. To evaluate the system performance, we consider two types of delay: processing delay and transmission delay. The former is the delay associated with the time of execution on to the decision, whereas the latter is the time to send event details to the server. According to the test on an LTE network, the transmission delay for the sensor data is approximately 1.150 s.

Meanwhile, the processing delay to determine the stress state is almost instant, which is 1.8 ms. By summing up interval time (180 s) and transmission/processing delays, the system will roughly complete every classification task in 181.168 s, which means 1.168 s late compared to the ideal interval. Our soft real-time system is based on the assumption that there will be no catastrophic failure even though the system fails to meet the deadlines. Missing the deadlines will only cause slightly degraded performance and the stakeholders can still make interventions to the stressed workers. Moreover, the use of Internet connection for data transmission from a smartphone to a server cannot guarantee that the packet delivery will be exactly on-time since the transmission time will heavily depend on the Internet provider quality.

Battery consumption analysis is an important aspect when dealing with soft real-time systems. In our previous study (Khowaja et al. 2018b), we conducted the battery analysis for mainly two services, i.e., fall detection and stress detection, while using the same devices. The users were allowed to use the smartphones in their usual way provided that they do not play 3D games or watch high definition videos during this experiment. The percentage of battery usage was recorded for every minute, accordingly.

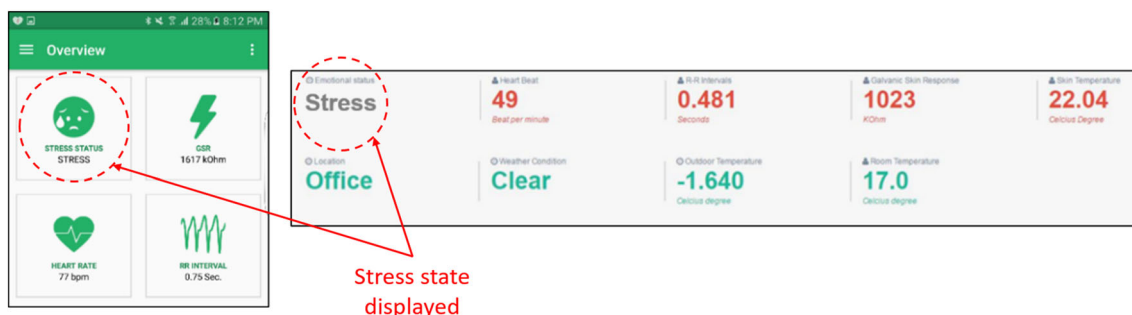


Fig. 11 Android (left) and web application (right) showing stress detection results

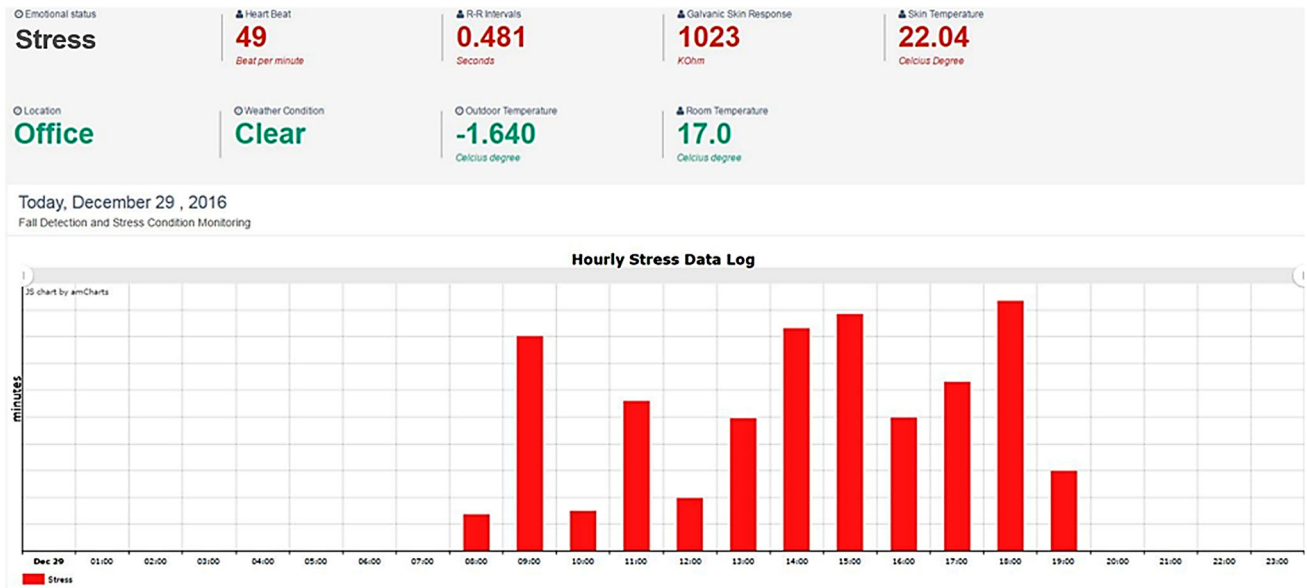


Fig. 12 Stress log chart displayed in the web application

It was revealed that the stress detection service utilizes very less amount of energy in comparison with the fall detection service, i.e., after 180 min of using stress detection service the smartphone only consumed less than 10% of the battery. Considering the prior findings, it can be assumed that the stress detection service qualifies for the real-world applicability based on low energy consumption.

Another important aspect of a soft real-time system is memory consumption. A snapshot from the android studio profiler in the “memory” section is shown in Fig. 13. It captures the dynamic of the occupied memory over several seconds. The snapshot was taken after all application services were running in the background, i.e., the peak of memory usage of the application. In total, the application takes only 17.9 MB space allocated, including the space for Java virtual machine, native code, graphics-related resource, stack, the application code and other OS-specific components. It is relatively efficient considering the capability of smartphones, nowadays. Note that our application only takes 6.3 MB out of the total allocated space. There are no obvious ups and downs in the memory usage graph, as the processes (data transmission and HTTP requests) take a very small portion of memory that they are barely observable. Besides, the app only stores temporary data with constant size (i.e., the buffer size), while the processing occurs in the server. Thus, there is no apparent dynamic memory allocation at run-time. Apart from memory profiler, we can also see network profiler. However, it is out of our concern since we target the users in South Korea with relatively fast Internet connection. Moreover, the application constantly sends and receives

approximately around 5 KB of data for every 180-s interval, which is trivial.

6 Discussion

6.1 Comparative analysis of the driveDB dataset

This subsection presents a quantitative analysis of the LMM features and the VSBF method. Table 10 provides a brief comparison for our proposed method and the existing works which were evaluated using the same dataset, driveDB, in terms of sensors used, methods adapted and accuracy. Healey and Picard (2005) and Chen et al. (2017a) achieved high accuracy by using all sensors in their study. Keshan et al. (2015) performed better on the driveDB dataset while using only ECG sensors. Their high accuracy is based on the average beat difference features extracted from QRS complexes with HSR devices. As our intended environment is a workplace that uses LSR devices for data acquisition, we cannot extract such complex features, accurately. However, the experimental analysis shows that the LMM features can improve the detection accuracy up to 95.58% for the driveDB dataset. Wang et al. (2013) achieved high accuracy using the combination of feature selection, i.e., PCA and kernel-based class separability (KBCS), and classifiers, i.e., linear discriminant analysis (LDA) and K-nearest neighbor (KNN). We proposed the VSBF method which can further improve the accuracy of the detection system. The analysis results show that by combining LMM features and the VSBF method, we can achieve the highest accuracy (98.48%) compared to the

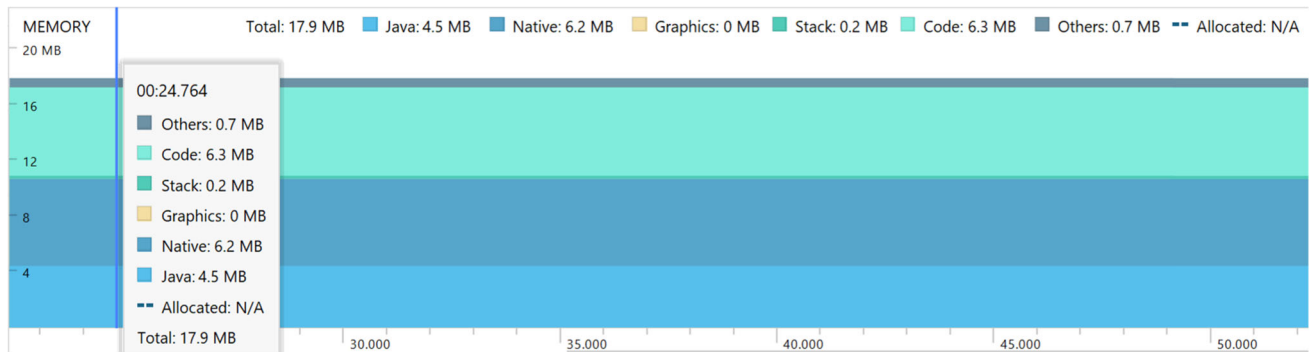


Fig. 13 Memory usage from android studio profiler

Table 10 Comparison with existing stress detection methods using the driveDB dataset

Authors	Sensors	Method	Accuracy (%)
Deng et al. (2012)	ECG, EMG, 2 GSR, RESP	PCA + LDA + DT + SVM + NB + KNN + EF	78.46
Singh and Bin (2013)	ECG, EMG, 2 GSR, RESP	NN + EF	88.75
Healey and Picard (2005)	ECG, EMG, 2 GSR, RESP	LDA + EF	97.4
Wang et al. (2013)	ECG	KBCS + LDA + PCA + KNN + EF	97.78
Zhang et al. (2010)	ECG, EMG, 2 GSR, RESP	SLBN + EF	90.53
Costin et al. (2012)	ECG	MV + MDC + EF	89.36
Keshan et al. (2015)	ECG	RT + ABD + EF	97.92
Ergin et al. (2011)	EMG	CVA + SSD + EF	69.24
Vanitha and Suresh (2013)	ECG	SOM + Hybrid SVM + HS + EF	91.0
Munla et al. (2015)	ECG	STFT + SVM + EF	83.0
Chen et al. (2017a)	ECG, RESP, 2 GSR	SBL + PCA + SVM + ELM + EF	~95.0
Proposed method	ECG, GSR	LMM + VSBF + EF	98.48

ECG electrocardiogram, GSR galvanic skin response, EMG electromyogram, RESP respiration, PCA principal component analysis, LDA linear discriminant analysis, DT decision tree, SVM support vector machines, NB Naïve Bayes, KNN K-nearest neighbor, EF existing features, NN neural networks, KBCS kernel-based class separability, SLBN structurally learned Bayesian network, MV morphological variability, MDC minimum distance classifier, RT random trees, ABD average beat difference, CVA common vector approach, SSD sub-space discriminant, SOM self-organizing maps, HS hierarchical structure, STFT short-term Fourier transform, SBL sparse Bayesian learning, ELM extreme learning machine, LMM local maxima and minima, VSBF voting and similarity-based fusion

existing works. Although we proposed the method to improve the accuracy using LSR devices the analysis shows that it can also be used to improve the detection accuracy for HSR devices. It is also noticed that we have used two sensor measurements only in our study while outperforming the other existing works which take into account all the complex sensor measurements.

6.2 Merits of the proposed system

The objective of this study is to design an automatic stress detection system based on LSR wrist-worn devices for workplace environments. The main contributions of our work can be summarized as follows:

(1) *Applicability in workplace environments* Unlike the existing studies which use electrodes and complex

wearable sensors for data acquisition in a controlled environment, it is a challenging and rewarding task to detect the stress using simple LSR sensors in real-life environments. The proposed stress detection system is designed to monitor the individual's mental status in real-world workplace environments, which makes detection results more practically applicable for use in human workspaces. However, the experiments performed in real-life workplaces may come across a few limitations such as continuous data acquisition and reliable detection. Taking consideration of these limitations, we used a wrist-worn device that is simple to wear and can acquire streams of data continuously.

(2) *High detection performance* We propose an efficient feature extraction from longer time windows and fusion methods to achieve high detection

performance. First, the LMM and VSBF method has been tested using the acquired dataset generated by LSR wrist-worn devices and then using the driveDB dataset which is acquired by HSR devices. Experimental analyses reveal that our proposed method not only achieves better detection accuracy on the dataset acquired by LSR devices but can also help in improving the accuracy of HSR devices.

- (3) *Implementation of a stress detection system* We implement a soft real-time stress detection system using the proposed method with LSR wrist-worn devices. The system was designed to prove the applicability of our method in real-life workplace environments using simple wearable and pervasive devices. Our stress detection system employs wrist-worn devices that overcome the limitations of mobility and complexity and uses a smartphone as a middleware for data transmission which is available and accessible to any factory worker. It was found that our system can return the recognition results within 1.168 s including feature extraction, classification and result fusion, which makes it possible to implement in real-life environments. The time computation is recorded by repeating and averaging the recognition results for ten times. Our current implementation only discusses the online solution for soft real-time stress detection; however, in the case of missing Internet connection, the system can still be used by moving the trained classifier (lightweight classification models) to the smartphone to perform the stress detection locally as proposed in various studies (Kwapisz et al. 2011; Uddin et al. 2016; Chen et al. 2017b; Ahmad et al. 2019).

6.3 Limitations and future work

Some of the limitations of our study are stated below:

- Limitation regarding the amount of data that is collected for our stress detection system. The data acquisition technique needs to be improved to a certain level to collect continuous data in working hours by taking full advantage of wrist-worn devices.
- The accuracy achieved on the acquired dataset is not as high as the detection accuracy achieved on the driveDB dataset. One of the possible reasons for not achieving such high accuracy is data acquisition using a low sampling rate. This limitation has been addressed by many researchers using pervasive wearable devices (Kim and Andre 2008; Setz et al. 2010; Egilmez et al. 2017); however, with the proposed set of features we have achieved higher detection accuracy compared to the early reported works.

- The analysis for stress recognition in the wild is missing in this study. The reason for not including such analysis is that this paper covers many aspects such as the feature engineering, VSBF fusion method, and a soft real-time stress detection system. The inclusion of analysis in the wild requires efforts with respect to the data collection, consent from the users, and implementation of more sophisticated machine learning methods such as 1D-CNNs or LSTMs. We assume that stress recognition in the wild can provide continuity to our work and can be considered as one of the potential future directions.

The stress detection system can be further integrated with diverse physical activities to analyze their relationship with stress conditions. It can give more insights into the subjective measure as the activity and behavior patterns of each subject will be different for prior and post-stress events. As the literature suggests that stress is a concept that is highly subjective, the subject-dependent analysis can be carried out by categorizing the characteristics of the subjects to make personalized models for stress detection.

7 Conclusions

In this study, we proposed a new feature set LMM from longer time windows (60 s), and a decision-level fusion method VSBF, to improve the accuracy of stress detection from LSR wrist-worn devices. The LMM features are first applied to the dataset acquired from LSR wrist-worn devices in compliance with the workplace environment. We chose the wrist-worn device having HRV and GSR sensors for the data acquisition so that the study might deviate toward realization as people do not prefer to wear complex sensor arrangements in real-life or at workspaces. We have also conducted an in-depth analysis using different classification algorithms for the VSBF method which combines the output of two classifiers and performs a similarity measure to draw the final result. Both of these methods were employed in our soft real-time system. For the acquired dataset, the result was incomparable with other approaches since the publicly available datasets are acquired using HSR wearable devices. For validating our proposed method, we tested the LMM features and VSBF method on the driveDB dataset which is widely used for stress detection. Experimental results showed that our method can improve the detection accuracy up to 15.23% for the acquired dataset and 7.01% for the driveDB dataset, respectively. Our analysis proved that the proposed feature set and fusion method can help considerably in improving the detection accuracy of the system. The performance of the proposed feature set and fusion method was compared

with the existing ones which used the same dataset. It revealed that the proposed method is capable of detecting stress with better accuracy by only employing ECG and GSR sensors only.

Although the classification accuracies are not very high on the acquired dataset, they are quite promising by keeping in view that the stress was detected using LSR wrist-worn devices. Besides the stress detection used for evaluation purposes, our soft real-time system has the potential for giving insights into an individual's behavior toward stress conditions. Our system can summarize the log of stress events on an hourly basis, providing details regarding the time of the day at which the person feels most stressed. These results can help to provide suggestions or recommendations to transit the stress state into a normal state. In addition, the battery and memory consumption analyses showed the efficiency of our soft real-time stress recognition system. In the future work, more analyses such as device energy consumption can be incorporated as the basis for further optimizing the system.

Acknowledgements This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07049113) and (2018R1D1A1B07047241).

Funding Dr. Seok-Lyong Lee is the recipient of the grant funded by the National Research Foundation of Korea (NRF) of the Ministry of Education (2018R1D1A1B07049113), and Dr. Bernardo Nugroho Yahya is the recipient of the grant funded by the National Research Foundation of Korea (NRF) of the Ministry of Education (2018R1D1A1B07047241).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All the procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee, following the British Psychological Society (BPS) ethical guidelines and policy for the use of electronic surveys/e-ethics from School Ethics Process, University of Nottingham.

Informed consent Informed consent was obtained from all the individual participants included in the study.

References

- (2016) WHO | World Health Organization. In: WHO. <http://www.who.int/en/>. Accessed 2 Mar 2017
- (2019) Coping with stress at work. In: Am. Psychol. Assoc. <https://www.apa.org/helpcenter/work-stress>. Accessed 1 Apr 2020
- Ahmad M, Khan AM, Mazzara M, et al (2019) Extended sammon projection and wavelet kernel extreme learning machine for gait-based legitimate user identification. In: Proceedings of the 34th

- ACM/SIGAPP symposium on applied computing—SAC'19. ACM Press, New York, New York, USA, pp 1216–1219
- Arnrich B, Setz C, La Marca R et al (2010) What does your chair know about your stress level? *IEEE Trans Inf Technol Biomed* 14:207–214. <https://doi.org/10.1109/TITB.2009.2035498>
- Becker K, Moreira VP, dos Santos AGL (2017) Multilingual emotion classification using supervised learning: comparative experiments. *Inf Process Manag* 53:684–704. <https://doi.org/10.1016/j.ipm.2016.12.008>
- Benstock D, Cegla F (2017) Extreme value analysis (EVA) of inspection data and its uncertainties. *NDT E Int* 87:68–77. <https://doi.org/10.1016/j.ndteint.2017.01.008>
- Bhui K, Dinos S, Galant-Miecznikowska M et al (2016) Perceptions of work stress causes and effective interventions in employees working in public, private and non-governmental organisations: a qualitative study. *BJPsych Bull* 40:318–325. <https://doi.org/10.1192/pb.bp.115.050823>
- Billehoj H (2007) Report on the ETUCE survey on teachers' work-related stress
- Boucsein W (2012) *Electrodermal activity*. Springer, Boston
- Bradley MM, Greenwald MK, Hamm AO (1993) Affective picture system. *Struct Emot* 25:48–65
- Bradley MM, Codispoti M, Cuthbert BN, Lang PJ (2001) Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion* 1:276–298. <https://doi.org/10.1037/1528-3542.1.3.276>
- Brezigar-Masten A, Masten I (2012) CART-based selection of bankruptcy predictors for the logit model. *Expert Syst Appl* 39:10153–10159. <https://doi.org/10.1016/j.eswa.2012.02.125>
- Chaudhary KM (2017) ML | raw and central moments. In: *Geeks-forGeeks*. <https://www.geeksforgeeks.org/ml-raw-and-central-moments/>. Accessed 6 Jan 2020
- Chen H (2012) End-to-end delay analysis and measurements in wireless sensor networks. Mid Sweden University, Sundsvall
- Chen L, Zhao Y, Ye P et al (2017a) Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst Appl* 85:279–291. <https://doi.org/10.1016/j.eswa.2017.01.040>
- Chen Z, Zhu Q, Soh YC, Zhang L (2017b) Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. *IEEE Trans Ind Inform* 13:3070–3080. <https://doi.org/10.1109/TII.2017.2712746>
- Cheng K-H, Fisher D, Canny J, Hartmann B (2011) How's my mood and stress? An efficient speech analysis library for nonobtrusive monitoring on mobile phones. In: Proceedings of 6th international conference on body area networks, pp 71–77
- Choi SS (2008) Correlation analysis of binary similarity measures and dissimilarity measures. *Dr Diss Pace Univ*
- Choi J, Gutierrez-Osuna R (2009) Using heart rate monitors to detect mental stress. In: Sixth international workshop on wearable and implantable body sensor networks. *IEEE*, pp 219–223
- Choi S-S, Cha S-H, Tappert CC (2010) A survey of binary similarity and distance measures. *J Syst Cybern Inform* 8:43–48
- Cinaz B, Arnrich B, La Marca R, Tröster G (2013) Monitoring of mental workload levels during an everyday life office-work scenario. *Pers Ubiquitous Comput* 17:229–239. <https://doi.org/10.1007/s00779-011-0466-1>
- Conditions EF for the I of L and W (2006) Fourth European working conditions survey
- Costin R, Rotariu C, Pasarica A (2012) Mental stress detection using heart rate variability and morphologic variability of EeG signals. In: 2012 International conference and exposition on electrical and power engineering. *IEEE*, pp 591–596
- Cox T, Griffiths A, Rial-González E, Work EA for S and H at (2000) Research on work-related stress. Office for Official Publications of the European Communities

- Crescentini C, Chittaro L, Capurso V et al (2016) Psychological and physiological responses to stressful situations in immersive virtual reality: differences between users who practice mindfulness meditation and controls. *Comput Human Behav* 59:304–316. <https://doi.org/10.1016/j.chb.2016.02.031>
- Crooks GE (2019) Field guide to continuous probability distributions. Berkeley Institute for Theoretical Sciences (BITS)
- Dedovic K, Renwick R, Mahani NK, Engert V (2005) The Montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *J Psychiatry Neurosci* 30:319–325
- Deng Y, Hsu DF, Wu Z, Chu C-H (2012) Feature Selection and combination for stress identification using correlation and diversity. In: 2012 12th international symposium on pervasive systems, algorithms and networks. IEEE, pp 37–43
- Dickerson SS, Kemeny ME (2004) Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol Bull* 130:355–391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Drucker H, Shahrany B, Gibbon DC (2002) Support vector machines: relevance feedback and information retrieval. *Inf Process Manag* 38:305–323. [https://doi.org/10.1016/S0306-4573\(01\)00037-1](https://doi.org/10.1016/S0306-4573(01)00037-1)
- Dunn G, Everitt BS (2004) An introduction to mathematical taxonomy. Courier Corporation, North Chelmsford
- Egilmez B, Poyraz E, Wenting Zhou, et al (2017) UStress: understanding college student subjective stress using wrist-based passive sensing. In: IEEE international conference on pervasive computing and communications workshops (PerCom Workshops). IEEE, pp 673–678
- Ergin S, Dag G, Gulmezoglu MB (2011) Stress level diagnosis for automobile drivers using common vector approach. In: 6th international advanced technologies symposium, pp 356–359
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann Stat* 28:337–407. <https://doi.org/10.1214/aos/1016218223>
- Future MR (2019) Wearable sensors market 2019—global trends, size, largest share, emerging technologies, competitive landscape, price trends, industry latest news, research report analysis By 2025. In: MedGadget. <https://www.medgadget.com/2019/08/wearable-sensors-market-2019-global-trends-size-largest-share-emerging-technologies-competitive-landscape-price-trends-industry-latest-news-research-report-analysis-by-2025.html>. Accessed 5 Jan 2020
- Gjoreski M, Gjoreski H, Luštrek M, Gams M (2016) Continuous stress detection using a wrist device. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing adjunct—UbiComp’16. ACM Press, New York, New York, USA, pp 1185–1193
- Habib M, Mohhtar M, Kamaruzzaman S et al (2014) Smartphone-based solutions for fall detection and prevention: challenges and open issues. *Sensors* 14:7181–7208. <https://doi.org/10.3390/s140407181>
- Hall M, Frank E, Holmes G et al (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newsl* 11:10–18
- Han L, Zhang Q, Chen X et al (2017) Detecting work-related stress with a wearable device. *Comput Ind* 90:42–49. <https://doi.org/10.1016/j.compind.2017.05.004>
- Handouzi W, Maaoui C, Pruski A, Moussaoui A (2014) Short-term anxiety recognition from blood volume pulse signal. In: IEEE 11th international multi-conference on systems, signals & devices (SSD14). IEEE, pp 1–6
- Hayward J (2018) Wearable sensors 2018–2028: technologies, markets & players. In: IDTechEx. <https://www.idtechex.com/en/research-report/wearable-sensors-2018-2028-technologies-markets-and-players/555>. Accessed 5 Jan 2020
- Healey JA (2000) Wearable and automotive systems for affect recognition from physiology. MIT
- Healey JA, Picard RW (2005) Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans Intell Transp Syst* 6:156–166. <https://doi.org/10.1109/TITS.2005.848368>
- Herbert BM, Pollatos O, Flor H et al (2010) Cardiac awareness and autonomic cardiac reactivity during emotional picture viewing and mental stress. *Psychophysiology* 47:342–354. <https://doi.org/10.1111/j.1469-8986.2009.00931.x>
- Hernandez J, Morris RR, Picard RW (2011) Call center stress recognition with person-specific models. In: Affective computing and intelligent interactions, pp 125–134
- Hernández J, Cabido R, Montemayor AS, Pantrigo JJ (2014) Human activity recognition based on kinematic features. *Expert Syst* 31:345–353. <https://doi.org/10.1111/exsy.12013>
- Hovsepian K, Al’Absi M, Ertin E, et al (2015) cStress: towards a gold standard for continuous stress assessment in the mobile environment. In: ACM international joint conference on pervasive and ubiquitous computing—UbiComp’15. ACM Press, New York, New York, USA, pp 493–504
- Imdadullah M (2012) The moments in statistics. In: Basic Stat. Data Anal. <http://itfeature.com/statistics/measure-of-dispersion/moments-in-statistics>. Accessed 6 Jan 2020
- Insights M (2019) Wearable sensors market to witness steady growth based on rising number of health & fitness monitoring applications Till 2025 | Million Insights. In: PR News Wire. <https://www.prnewswire.com/news-releases/wearable-sensors-market-to-witness-steady-growth-based-on-rising-number-of-health-fitness-monitoring-applications-till-2025-million-insights-300896156.html>. Accessed 5 Jan 2020
- Intelligence M (2019) Wearable sensors market—growth, trends, and forecast (2020 - 2025). In: Mordor Intell. <https://www.mordorintelligence.com/industry-reports/global-wearable-sensors-market>. Accessed 5 Jan 2020
- Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat* 37:547–579
- Jebelli H, Khalili MM, Lee S (2019) Mobile EEG-based workers’ stress recognition by applying deep neural network. In: Advances in informatics and computing in civil and construction engineering. Springer, Cham, pp 173–180
- Johnson B, Batia AS, Haun J (2008) Perceived stress among graduate students: roles, responsibilities, & social support. *VAHPERD J* 29:31+
- Keshan N, Parimi PV, Bichindaritz I (2015) Machine learning for stress detection from ECG signals in automobile drivers. In: 2015 IEEE international conference on big data (Big Data). IEEE, pp 2661–2669
- Khowaja SA, Lee S-L (2020) Semantic image networks for human action recognition. *Int J Comput Vis* 128:393–419. <https://doi.org/10.1007/s11263-019-01248-3>
- Khowaja SA, Dahri K, Kumbhar MA, Soomro AM (2015) Facial expression recognition using two-tier classification and its application to smart home automation system. In: International conference on emerging technologies (ICET). IEEE, pp 1–6
- Khowaja SA, Yahya BN, Lee S-L (2017) Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems. *Expert Syst Appl* 88:165–177. <https://doi.org/10.1016/j.eswa.2017.06.040>
- Khowaja SA, Khuwaja P, Ismaili IA (2018a) A framework for retinal vessel segmentation from fundus images using hybrid feature set

- and hierarchical classification. *Signal Image Video Process.* <https://doi.org/10.1007/s11760-018-1366-x>
- Khowaja SA, Prabono AG, Setiawan F et al (2018b) Contextual activity based Healthcare Internet of Things, Services, and People (HIoTSP): an architectural framework for healthcare monitoring using wearable sensors. *Comput Netw* 145:190–206. <https://doi.org/10.1016/j.comnet.2018.09.003>
- Kim J, Andre E (2008) Emotion recognition based on physiological changes in music listening. *IEEE Trans Pattern Anal Mach Intell* 30:2067–2083. <https://doi.org/10.1109/TPAMI.2008.26>
- Kim D, Seo Y, Cho J, Cho C-H (2008) Detection of subjects with higher self-reporting stress scores using heart rate variability patterns during the day. In: 30th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 682–685
- Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. *ACM SIGKDD Explor Newsl* 12:74. <https://doi.org/10.1145/1964897.1964918>
- Lang PJ, Bradley MM, Cuthbert BN (1999) International affective picture system (IAPS): affective ratings of pictures and instruction manual
- Laplante PA (2004) Real-time systems design and analysis. Wiley, New York
- Lawanont W, Mongkolnam P, Nukoolkit C, Inoue M (2019) Daily stress recognition system using activity tracker and smartphone based on physical activity and heart rate data. In: International conference on intelligent decision technologies, pp 11–21
- Le Calve A, Savoy J (2000) Database merging strategy based on logistic regression. *Inf Process Manag* 36:341–359. [https://doi.org/10.1016/S0306-4573\(99\)00036-9](https://doi.org/10.1016/S0306-4573(99)00036-9)
- Levecque K, Anseel F, De Beuckelaer A et al (2017) Work organization and mental health problems in PhD students. *Res Policy* 46:868–879. <https://doi.org/10.1016/j.respol.2017.02.008>
- LiKamWa R, Liu Y, Lane N, Zhong L (2011) Can your smartphone infer your mood. In: PhoneSense workshop, pp 1–5
- Liu Y, Sourina O (2014) Real-time subject-dependent EEG-based emotion recognition algorithm. In: Transactions on computational science XXIII, pp 199–223
- Lu H, Fraundorfer D, Rabbi M, et al (2012) StressSense: detecting stress in unconstrained acoustic environments using smartphones. In: ACM conference on ubiquitous computing—UbiComp'12. ACM Press, New York, New York, USA, pp 351–360
- Mariani S, Migliorini M, Tacchino G, et al (2012) Clinical state assessment in bipolar patients by means of HRV features obtained with a sensorized T-shirt. In: Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 2240–2243
- MarketsandMarkets (2016) Wearable technology market worth 51.60 Billion USD by 2022. <http://www.marketsandmarkets.com/PressReleases/wearable-electronics.asp>. Accessed 6 Apr 2017
- McNichol D (2018) On average, you're using the wrong average: geometric & harmonic means in data analysis. In: Towar. Data Sci. <https://towardsdatascience.com/on-average-youre-using-the-wrong-average-geometric-harmonic-means-in-data-analysis-2a703e21ea0>. Accessed 6 Jan 2020
- Melillo P, Bracale M, Pecchia L (2011) Nonlinear heart rate variability features for real-life stress detection. Case study: students under stress due to university examination. *Biomed Eng Online* 10:1–13. <https://doi.org/10.1186/1475-925X-10-96>
- Mental Health Foundation (2017) Good mental health for all
- Milczarek M, Schneider E, Rial-González E (2009) European agency for safety and health at work: EUROPEAN RISK OBSERVATORY REPORT
- Mokhayeri F, Akbarzadeh-T M-R, Toosizadeh S (2011) Mental stress detection using physiological signals based on soft computing techniques. In: 18th Iranian conference of biomedical engineering (ICBME). IEEE, pp 232–237
- Montesinos V, Dell'Agnola F, Arza A, et al (2019) Multi-modal acute stress recognition using off-the-shelf wearable devices. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 2196–2201
- Muaremi A, Bexheti A, Gravenhorst F, et al (2014) Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors. In: IEEE-EMBS international conference on biomedical and health informatics (BHI). IEEE, pp 185–188
- Munla N, Khalil M, Shahin A, Mourad A (2015) Driver stress level detection using HRV analysis. In: 2015 International conference on advances in biomedical engineering (ICABME). IEEE, pp 61–64
- Nakasuka S, Koishi T (1995) Automated extraction of attribute hierarchies for an improved decision-tree classifier. *Eng Appl Artif Intell* 8:391–399. [https://doi.org/10.1016/0952-1976\(95\)00022-S](https://doi.org/10.1016/0952-1976(95)00022-S)
- Niskanen J-P, Tarvainen MP, Ranta-aho PO, Karjalainen PA (2004) Software for advanced HRV analysis. *Comput Methods Programs Biomed* 76:73–81. <https://doi.org/10.1016/j.cmpb.2004.03.004>
- Pain E (2018) Graduate students need more mental health support, study highlights. In: Am. Assoc. Adv. Sci. <https://www.sciencemag.org/careers/2018/03/graduate-students-need-more-mental-health-support-new-study-highlights>. Accessed 5 Jan 2020
- Pal M (2005) Random forest classifier for remote sensing classification. *Int J Remote Sens* 26:217–222. <https://doi.org/10.1080/01431160412331269698>
- Palomba D, Stegagno L (1993) Physiology, perceived emotion and memory: responding to film sequences. *Struct Emot* 156–168
- Palomba D, Sarlo M, Angrilli A et al (2000) Cardiac responses associated with affective processing of unpleasant film stimuli. *Int J Psychophysiol* 36:45–57. [https://doi.org/10.1016/S0167-8760\(99\)00099-9](https://doi.org/10.1016/S0167-8760(99)00099-9)
- Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. *IEEE Trans Biomed Eng BME* 32:230–236. <https://doi.org/10.1109/TBME.1985.325532>
- Patel S, Park H, Bonato P et al (2012) A review of wearable sensors and systems with application in rehabilitation. *J Neuroeng Rehabil* 9:21. <https://doi.org/10.1186/1743-0003-9-21>
- Rai D, Kosidou K, Lundberg M et al (2012) Psychological distress and risk of long-term disability: population-based longitudinal study. *J Epidemiol Community Health* 66:586–592. <https://doi.org/10.1136/jech.2010.119644>
- Rani P, Liu C, Sarkar N, Vanman E (2006) An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Anal Appl* 9:58–69. <https://doi.org/10.1007/s10044-006-0025-y>
- Renaud P, Blondin J-P (1997) The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. *Int J Psychophysiol* 27:87–97. [https://doi.org/10.1016/S0167-8760\(97\)00049-4](https://doi.org/10.1016/S0167-8760(97)00049-4)
- Research ZM (2019) Global medical sensors market will reach USD 19.77 billion by 2025: zion market research. In: Globe news wire. <https://www.globenewswire.com/news-release/2019/05/23/1841434/0/en/Global-Medical-Sensors-Market-Will-Reach-USD-19-77-Billion-By-2025-Zion-Market-Research.html>. Accessed 5 Jan 2020
- Romer M (2011) Teachers' work-related stress: assessing, comparing and evaluating the impact of psychosocial hazards on teachers at their workplace
- Rooney N, Wang H, Taylor PS (2014) An investigation into the application of ensemble learning for entailment classification. Inf

- Process Manag 50:87–103. <https://doi.org/10.1016/j.ipm.2013.08.002>
- Ruf T (1999) The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biol Rhythm Res* 30:178–201. <https://doi.org/10.1076/brhm.30.2.178.1422>
- Russell JA, Pratt G (1980) A description of the affective quality attributed to environments. *J Pers Soc Psychol* 38:311–322. <https://doi.org/10.1037/0022-3514.38.2.311>
- Salahuddin L, Cho J, Jeong MG, Kim D (2007) Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In: 29th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 4656–4659
- Saleem K, Derhab A, Al-Muhtadi J, Shahzad B (2015) Human-oriented design of secure Machine-to-Machine communication system for e-Healthcare society. *Comput Human Behav* 51:977–985. <https://doi.org/10.1016/j.chb.2014.10.010>
- Sandi C (2013) Stress and cognition. *Wiley Interdiscip Rev Cogn Sci* 4:245–261. <https://doi.org/10.1002/wcs.1222>
- Sandulescu V, Andrews S, Ellis D, et al (2015) Stress detection using wearable physiological sensors. In: International work-conference on the interplay between natural and artificial computation, pp 526–532
- Sano A, Picard RW (2013) Stress recognition using wearable sensors and mobile phones. In: Humaine association conference on affective computing and intelligent interaction. IEEE, pp 671–676
- SERV S-I& A (2004) Flemish workability monitor
- Setiawan F, Khowaja SA, Prabono AG, et al (2018) A framework for real time emotion recognition based on human ANS using pervasive device. In: IEEE 42nd annual computer software and applications conference (COMPSAC). IEEE, pp 805–806
- Setz C, Arnrich B, Schumm J et al (2010) Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans Inf Technol Biomed* 14:410–417. <https://doi.org/10.1109/TITB.2009.2036164>
- Shepard C, Rahmati A, Tossell C et al (2011) LiveLab: measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Perform Eval Rev* 38:15–20
- Sierra ADS, Ávila CS, Casanova JG, Bailador G (2011) Real-time stress detection by means of physiological signals. In: Recent application in biometrics, pp 23–45
- Silva I, Moody GB (2014) An open-source toolbox for analysing and processing PhysioNet databases in MATLAB and octave. *J Open Res Softw* 2:1–4. <https://doi.org/10.5334/jors.bi>
- Singh M, Bin QA (2013) A novel method of stress detection using physiological measurements of automobile drivers. *Int J Electron Eng* 5:13–20
- Stawski RS, Sliwinski MJ, Smyth JM (2006) Stress-related cognitive interference predicts cognitive function in old age. *Psychol Aging* 21:535–544. <https://doi.org/10.1037/0882-7974.21.3.535>
- Steidl S (2009) Automatic classification of emotion related user states in spontaneous children's speech. University of Erlangen-Nuremberg, Erlangen
- Stroop JR (1935) Studies of interference in serial verbal reactions. *J Exp Psychol* 18:643–662. <https://doi.org/10.1037/h0054651>
- Subasi A, Ismail Gursoy M (2010) EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst Appl* 37:8659–8666. <https://doi.org/10.1016/j.eswa.2010.06.065>
- Systems B (2017) Data acquisition, loggers, amplifiers, transducers, electrodes | BIOPAC. <https://www.biopac.com>. Accessed 13 Jul 2017
- Taelman J, Vandepuit S, Spaepen A, Van Huffel S (2009) Influence of mental stress on heart rate and heart rate variability. In: 4th European conference of the international federation for medical and biological engineering. IFMBE proceedings, pp 1366–1369
- Technology T (2016) Thought technology Ltd. T7555M thought technology Ltd. thoughttechnology.com/index.php/flexcomp-system-with-biograph-infiniti-software-t7555m.html. Accessed 12 Jul 2017
- Team AV content (2017) 6 Common probability distributions every data science professional should know. In: Anal. Vidhya. <https://www.analyticsvidhya.com/blog/2017/09/6-probability-distributions-data-science/>. Accessed 6 Jan 2020
- Thorsteinsson EB, Brown RF, Richards C (2014) The relationship between work-stress, psychological stress and staff health and work outcomes in office workers. *Psychology* 05:1301–1311. <https://doi.org/10.4236/psych.2014.510141>
- Tulppo MP, Mäkikallio TH, Takala TE et al (1996) Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *Am J Physiol* 271:H244–H252. [https://doi.org/10.1016/0735-1097\(94\)90177-5](https://doi.org/10.1016/0735-1097(94)90177-5)
- Uddin MT, Billah MM, Hossain MF (2016) Random forests based recognition of human activities and postural transitions on smartphone. In: 5th international conference on informatics, electronics and vision (ICIEV). IEEE, pp 250–255
- Valenza G, Allegrini P, Lanatà A, Scilingo EP (2012) Dominant Lyapunov exponent and approximate entropy in heart rate variability during emotional visual elicitation. *Front Neuroeng* 5:1–7. <https://doi.org/10.3389/fneng.2012.00003>
- van den Broek EL, Lisý V, Janssen JH, et al (2009) Affective man-machine interface: unveiling human emotions through biosignals. In: Biomedical engineering systems and technologies. BIOSTEC, pp 21–47
- Vanitha L, Suresh GR (2013) Hybrid SVM classification technique to detect mental stress in human beings using ECG signals. In: 2013 international conference on advanced computing and communication systems. IEEE, pp 1–6
- Wang J-S, Lin C-W, Yang Y-TC (2013) A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing* 116:136–143. <https://doi.org/10.1016/j.neucom.2011.10.047>
- Wen W, Liu G, Cheng N et al (2014) Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Trans Affect Comput* 5:126–140
- Wijsman J, Grundlehner B, Hao Liu, et al (2011) Towards mental stress detection using wearable physiological sensors. In: Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 1798–1801
- Wijsman J, Grundlehner B, Liu H, et al (2013) Wearable physiological sensors reflect mental stress state in office-like situations. In: Humaine association conference on affective computing and intelligent interaction. IEEE, pp 600–605
- Wikgren M, Maripuu M, Karlsson T et al (2012) Short telomeres in depression and the general population are associated with a hypocortisolemic state. *Biol Psychiatry* 71:294–300. <https://doi.org/10.1016/j.biopsych.2011.09.015>
- World Health Organisation (2013) Guidelines for the management of conditions specifically related to stress. *Assess. Manag. Cond. Specifically Relat. to Stress mhGAP Interv. Guid. Modul. (version 1.0)* 1–273
- Zenonos A, Khan A, Kalogridis G, et al (2016) HealthyOffice: mood recognition at work using smartphones and wearable sensors. In: 2016 IEEE international conference on pervasive computing and communication workshops (PerCom Workshops). IEEE, pp 1–6
- Zhai J, Barreto A (2006) Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In: International conference of the IEEE engineering in medicine and biology society. IEEE, pp 1355–1358
- Zhai J, Barreto AB, Chin C, Li C (2005) Realization of stress detection using psychophysiological signals for improvement of

- human-computer interaction. In: Proceedings. IEEE SoutheastCon. IEEE, pp 415–420
- Zhang L, Tamminedi T, Ganguli A, et al (2010) Hierarchical multiple sensor fusion using structurally learned Bayesian network. In: Wireless Health 2010. ACM, pp 174–183
- Zimmermann a opinion: watchmakers get creative in face of smart threat. In: CNN bus. <https://money.cnn.com/2016/03/18/luxury/watches-smartwatches-industry/index.html>. Accessed 5 Jan 2020

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