[Turkish Journal of Electrical Engineering and Computer Sciences](https://journals.tubitak.gov.tr/elektrik)

Volume 30 | Number 1 Article 21

1-1-2022

Stressed or just running? Differentiation of mental stress and physical activityby using machine learning

YEKTA SAİD CAN

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2022) 30: 312 – 327
© TÜBİTAK
doi:10.3906/elk-2102-138

Stressed or just running? Differentiation of mental stress and physical activity by using machine learning

Yekta Said CAN*
History Department, College of Social Sciences and Humanities, Koç University, İstanbul, TURKEY

Received: 27.02.2021 • **Accepted/Published Online:** 09.11.2021 • **Final Version:** 19.01.2022

Abstract: Recently, modern people have excessive stress in their daily lives. With the advances in physiological sensors and wearable technology, people's physiological status can be tracked, and stress levels can be recognized for providing beneficial services. Smartwatches and smartbands constitute the majority of wearable devices. Although they have an excellent potential for physiological stress recognition, some crucial issues need to be addressed, such as the resemblance of physiological reaction to stress and physical activity, artifacts caused by movements and low data quality. This paper focused on examining and differentiating physiological responses to both stressors and physical activity. Physiological data are collected in the laboratory environment, which contain relaxed, stressful and physically active states and they are differentiated successfully by using machine learning.

Key words: Machine learning, stress detection, affective computing, smart band, PPG, physical activity detection

1. Introduction

Stress is one of the significant issues in modern society. It is generally experienced when individuals fail to respond to mental, emotional, or physical demands competently [1]. Hence, the major causes of stress can be listed as the exposure to the demands and pressures from physical or mental activities or self-imposed requirements, obligations and self-criticism [2]. Severe illnesses can be caused by stress such as depression, cardiovascular diseases, infectious diseases, and cancer [3] and therefore, continuous measurement of stress in daily life gains importance. In behavioral science, researchers usually employ self-reports by periodically collecting instantaneous assessments of perceived stress. However, it is not possible to recognize stress levels with this approach continuously because of its burden and obtrusiveness. Therefore, researchers investigated different sensing techniques for recognizing stress levels automatically, such as facial expressions, physiological signals, speech and gestures. Due to their noninvasiveness, privacy and ease of use in daily lives, detecting stress by using smartbands and smartwatches from physiological signals attracted more attention.

Notwithstanding the high potential of smartbands and smartwatches in stress detection, researchers need to deal with several issues in exploiting physiological reactions. The first one is the similar physiological reactions to stressors and changes in physical or mental conditions. A physiological pattern might not necessarily indicate a stress level change [4, 5]. Similar physiological reactions can be seen under demanding physical activities, physical discomfort, noise, changes in posture, lighting conditions, demanding mental task and emotional stress [1]. The second issue is the limitations of these devices for detecting physiological signals. They are sensitive to noise, might miss samples, and have problems with battery life. The third issue is the different types of stress

*Correspondence: ycan@ku.edu.tr

can have varying physiological expressions. Under physical stress, the sympathetic nervous reaction is more dominant, but under cognitive stress, the adrenal reaction is more dominant. Finally, physiological reactions to stressors depend on individuals. Personalizing stress models might be necessary for accurate systems.

With the widespread usage of these wrist-worn wearables, health applications have been developed for improving the life quality of users. However, these devices are not without challenges. Periodical stress detection apps were developed by consumer smartwatch companies, but they require users to be still during these periodic measurements. The exact instruction is "during a stress test, wear your watch correctly and keep still". This means that these applications are not appropriate for recognizing the stress levels of individuals in their daily lives because the movements are unlimited. The reasons for this requirement could be artifacts caused by movements or difficulty in differentiating physical activity and stress from physiological responses.

In this study, a stress and physical activity differentiation system is developed, which is implemented for unobtrusive smart wearable devices. To the best of our knowledge, this study is the first one to differentiate stress and physical activity system works with smart bands. The system is tested on 14 participants in the laboratory environment. Heart rate variability (HRV) and electrodermal activity (EDA) signals obtained from an Empatica E4 smartband are used for this purpose. After cleaning the artifacts, most discriminative features were extracted. A variety of classifiers were tested for differentiating stressful, relaxed, physically active states.

The rest of the paper is organized as follows: In Section 2, the related work is provided for automatic stress and physical activity recognition systems that use smart bands or smartwatches. In Section 3, the data collection procedure is explained. In Section 4, the proposed smart band-based stress and physical activity differentiation scheme is presented. In Section 5, the experimental results of the proposed system are discussed. In Section 6, the findings of the study are evaluated and future work of the current research is discussed.

2. Literature review

After the emergence of smartbands and smartwatches, researchers used them to improve the quality of life. In the literature, there are studies for recognizing physical activity and mental stress levels. Physical activity intensity level detection studies were accelerated with the emergence of smartphones. The motion sensors (accelerometer and gyroscope) were used for this purpose [6]. After smartphones, smartwatches and smartbands were also employed for physical activity detection. Degroote et al. [7] used Polar M600, Huawei watch and Asus Zenwatch for detecting physical activity levels. The accelerometer sensors were employed in these devices. In daily life, they achieved approximately 72.25% accuracy. Dobbins and Rawssizadeh [8] developed a physical activity level recognition system. They used an LG watch and Samsung Gear S watch and utilized the combination of acceleration data of smartwatches and a smartphone for physical activity level detection. They achieved 78.61% accuracy. Davoudi et al. [9] conducted a validity study of Samsung Gear S devices for physical activity level recognition and showed that they have 87% accuracy.

Another widely conducted research field with smartwatches and smartbands is automatic stress recognition. Stress has a physiological reaction and it can be measured from physiological signals such as electrodermal activity (EDA), HRV, skin temperature (ST) and acceleration (ACC). Samsung Gear S, S2 and Microsoft Band 2 are among the widely used commercial smartwatches and Empatica E4 is a commonly used research-oriented smartband for stress level recognition. Hao et al. [10] used Empatica E4 PPG sensor for extracting HRV information and detected stress levels in real-life settings. De Arriba-Perez et al. [11] used Microsoft Band 2 for detecting mental stress in laboratory settings and achieved 85% accuracy. Siirtola et al. [12] used a public dataset WESAD and performed experiments. This dataset contains physiological data of participants in a laboratory environment collected with Empatica E4. Can et al. [13] used a combination of smartwatches and smartbands for collecting data in a real-life event. By using HRV and EDA, they achieved 90% accuracy for detecting mental stress. They also analyzed the difference in physiological reactions between cognitive load and mental stress. As it can be seen from the literature, physical activity intensity and mental stress are studied separately in several studies. However, to the best of our knowledge, there is not any study that investigates both physical activity and mental stress and tries to differentiate them.

3. Laboratory data collection

The laboratory data is collected through a psychological experiment by using an implementation of the Trier Social Test (TSST) [14]. TSST is commonly used to induce stress in the literature. Controlled laboratory data were obtained from the 14 subjects who participated in the experiment. All participants followed the same process. They were college students whose ages are between 20 and 25. Nine of them were males and five of them were females. The experiment took approximately 1 h and it consists of baseline (prestress), mental stress, recovery and physical activity sessions (see Figure 1).

Interviewers and the participants communicated in Turkish which is the mother tongue of all participants. Furthermore, they can speak English as a foreign language. This fact is utilized for inducing stress on the participants. The participants wear the Empatica E4 smartbands during the experiment and the sessions were recorded with a camera. The physiological signals of a participant during the experiment are shown in Figure 2.

Figure 1. The experiment procedure is shown.

In total, we collected 14 h of physiological data and 56 self-reports in the laboratory environment. The data is imbalanced in terms of labels. Approximately 50% of the data is stressed data, 33% of the data is relaxed data, and the remaining 16% of the data is divided into physically active and recovery sessions equally. In order to overcome the class imbalance problem, the randomly undersampling method is used while extracting the results.

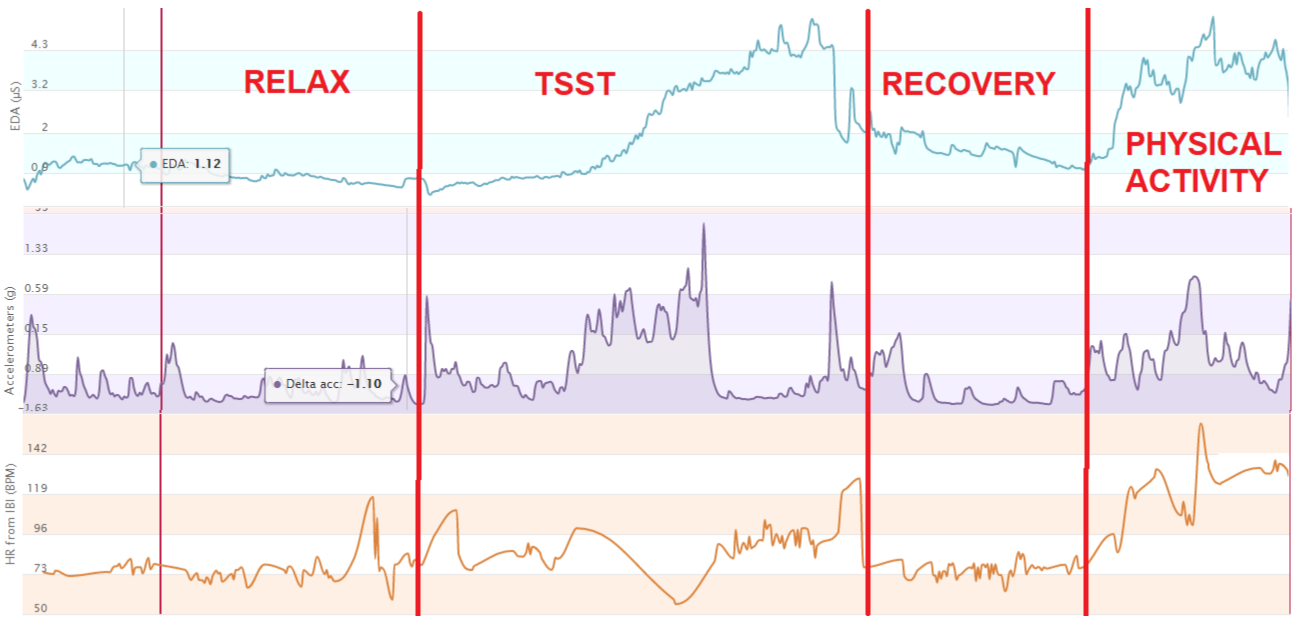

Figure 2. The changes in physiological signals of a participant during the experiment. The top signal is EDA, the middle signal is acceleration and the bottom one is heart rate signal.

3.1. Set up

In the set-up phase;

1. Experiment areas are prepared, the camera is set and we check the Empatica E4 smartband battery and functionalities.
2. Interviewers are prepared. They keep eye contact with the participant during the experiment. Their gestures and mimics should be neutral.
3. We informed the participants about the procedure and then they sign the consent forms.
4. The participant wears the smartband (Empatica E4).
5. The participants are asked to turn off their cellphones to refrain from distraction.

3.2. Prestress phase

In the prestress phase;

1. Perceived Stress Scale (PSS-14) form is filled.
2. The participants stay in the waiting area and relax for 10 min. We put emotionally neutral magazines such as home, garden, car, furniture, fashion related ones for them to read.

3.3. The TSST phase

The TSST is implemented as follows;

1. The participant enters the interview area.

2. For preparing the TSST speech, we read the following text to all the participants: “This is the speech preparation portion of the task; you are expected to prepare a five-min speech describing why you study [name of the degree that the participant studies/studied] and why you would be a good candidate for your ideal job. Your speech will be videotaped and reviewed by the psychologists that we conduct the research. You have five min to prepare and your time begins now.”

3. After that, the participants prepare their speech. The digital timer is set to five min. Interviewers leave the room during this phase.

4. At the end of the speech preparation phase we inform them by reading the following text: “This is the speech portion of the task. You should speak for the entire five-min time period. Your time begins now”. Interviewers should start the recording of the camera.

5. TSST speech performance period: If the participant stops during this period, interviewers allow him/her to stay silent for around 20 s and then prompt: “You still have time remaining.”

6. After the first 2 min of the speech period, the participants are interrupted and asked to continue their speech in English by telling them: “Could you continue in English from now on, please ?”

7. At the end of 2.5 min, if the participant does not attempt to reply to both questions, interviewers prompt the participant to answer the other question.

8. At the end of the speech performance period, the communication between interviewers and the participant resumes in Turkish. Interviewers reset the timer to 5 min and read the following to the participant: “During the final five-min math portion of this task, you will be asked to sequentially subtract 13 from 1022. You will verbally report your answers aloud and be asked to start over from 1022 if a mistake is made. Your time begins now.” If the participant makes any mistake, the interviewer says the following: ”That is incorrect, please start over from 1022.”

3.4. Poststress recovery phase

In order to decrease the stress levels of participants, we applied a biofeedback-based intervention technique which is the built-in breathing application of Apple Watch. The technique is applied as follows;

1. Participants sit on the couch as a relaxing place.

2. Participants wear the Apple Watch provided to them and follow the breathing exercise built in the Apple Watch for a minute and then follow a mindfulness video, for the remaining four min, on a comfortable couch, while sitting or lying as they prefer.

3. Interviewers leave the room after giving the Apple Watch to the participants.

4. At the end of the five-min-long recovery period, interviewers return to the room.

3.5. Physical activity phase

1. Interviewers leave the room in order not to induce stress during the physical-activity period.

2. Participant follows a five-min-long exercise on Youtube. It is an intense cardio workout without using any equipment. The link could be accessed at ”https://www.youtube.com/watch?v=Tz9d7By2ytQ”

3. Interviewers return to the room and return the pieces of equipment.

4. Mental stress and physical activity level detection method

An unobtrusive stress and physical activity detection system which is suitable for daily life use is proposed (see Figure 3). Empatica E4 smart wristband is selected as the unobtrusive device. The battery life of this device is around forty eight h and it has a 3D accelerometer sensor, a photoplethysmography (PPG) sensor, an skin temperature sensor and an EDA sensor. However, it has lower data quality than medical-grade devices and it is more sensitive to unrestricted motions in daily life. Therefore, state-of-the-art preprocessing, feature extraction and machine learning algorithms are required for daily life stress and physical activity level detection and differentiation system. Preprocessing and feature extraction modules were developed in our previous study [15].

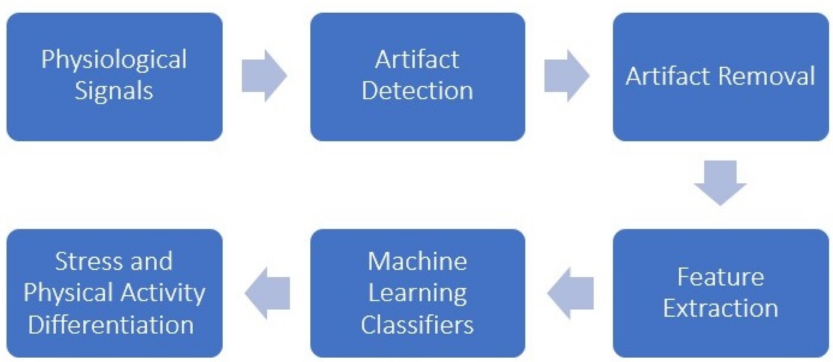

Figure 3. The high-level block diagram of the proposed system is shown.

4.1. Preprocessing

Physiological time-series data obtained from various sensors of the smartband were divided into nonoverlapping windows. The window sizes are determined as 120 s because stress stimulation and recovery processes lasted around 3-4 min and selected window lengths can capture them [1].

4.1.1. HRV artifact detection and removal module

An artifact detection percentage threshold between each R to R interval and the local average was applied for the heart rate signal. The threshold was selected as 20% because it is unlikely that RR intervals can deviate from the local mean with that amount [16]. The removed artifact data points were replaced with a cubic spline

[1]Harvard University (2021) Understanding the stress response [online]. https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response [accessed 27 July 2021].

interpolation function, as implemented in Kubios [16]. The interpolation technique achieved better results, and therefore it is preferred to the application of minimum consecutive time and sample constraints on the remaining data [17]. The windows that contain more than 10% artifacts were discarded.

4.1.2. EDA artifact detection and removal module

The toolbox developed by Taylor et al. [18] was used to detect and remove the artifacts in the EDA signal. The EDAExplorer toolbox employs the SVM algorithm and determines artifact data points in the EDA signal with less than 5% error by examining ST, ACC, and EDA signals. The artifacts were automatically removed after being detected by this tool. The batch processing feature was further added. Then, the cleaned signal is passed to the EDA feature extraction unit.

4.2. Feature extraction

For each modality, discriminative features are extracted. These state-of-the-art features were commonly used in the literature [19–21]. The feature extraction methodologies for each of the physiological signals are described in this section.

4.2.1. HRV feature extraction module

Thirteen HRV features were computed by analyzing the RR intervals. To extract frequency-domain features, two methods are applied. Firstly, RR intervals are resampled at 4 Hz [22] and Fast Fourier transform (FFT) is applied. In a second way, the Lomb-Scargle technique is applied, which is a special technique for converting nonequidistant sampled signals to the frequency domain. The extracted heart rate variability features are listed as: 1. Mean value of the inter-beat (RR) intervals 2. Standard deviation of the inter-beat interval 3. Root mean square of the successive difference of the RR intervals. 4. Percentage of the number of the successive RR intervals varying more than 50 ms from the previous interval 5. Total number of RR intervals divided by the height of the histogram of all RR intervals measured on a scale with bins of 1/128 s 6. Triangular interpolation of RR interval histogram 7. Power in the low-frequency band (0.04–0.15 Hz) 8. Power in the high-frequency band (0.15–0.4 Hz) 9. Ratio of LF to HF. 10. Prevalent low-frequency oscillation of the heart rate 11. Prevalent high-frequency oscillation of the heart rate 12. Power in the very low-frequency band (0.00–0.04 Hz) 13. Related standard deviation of successive RR interval differences.

4.2.2. EDA feature extraction module

EDA signal is decomposed into phasic and tonic components using the convex optimization-based EDAcvx tool [18]. EDAcvx also cleanses the Gaussian noise. Ten discriminative features from both the tonic and phasic components of the EDA signal were extracted. Percentiles are very usable for examining the distribution of number sets using various EDA graphs, and different percentiles were used. The following EDA features were selected from the literature [19, 24]. 1. Mean value 2. Maximum value 3. Standard deviation 4. Minimum value 5. Number of peaks 6. Derivative 7. Number of strong peaks 8. Twentieth percentile 9. Eightieth percentile 10. Quartile deviation.

4.2.3. ACC feature extraction module

The sensor data of the accelerometer is employed for two different purposes. First, features are extracted from this sensor. This sensor was further used to clean the EDA signal in the EDAExplorer tool. Extracted accelerometer features are Mean X (Mean acceleration over x-axis), Mean Y (Mean acceleration over y-axis), Mean Z (Mean acceleration over z-axis), Mean ACC MAG (Mean acceleration over acceleration magnitude axis).

4.3. Machine learning classification algorithms

To detect stress and physical activity levels separately and differentiate them, the Python Keras and Scikit libraries are employed. Several preprocessing tools were applied to the data before initiating the classification process. When the number of instances in each class is concerned, our dataset is not balanced. Extra samples of the majority class were removed, and this issue was solved. Therefore, the biasing of the classifiers towards the majority classes was prevented. Five different machine learning classifiers were used. They are the most widely applied classifiers in the literature [19], [25], Logistic Regression (LR), kNN, multi-layer perceptron (MLP), support vector machine (SVM) and random forest (RF). Python scikit library was used for the implementation of these classifiers. The default parameters were used. In order to evaluate the performance of the classifiers, 10-fold cross-validation was applied.

4.3.1. Support vector machines (SVM)

SVM creates decision planes that define decision boundaries. A decision plane can be defined as the plane that divides objects belonging to different classes. In some classification tasks, complex decision structures are needed to separate these objects into their classes correctly. Support vector machines are designed to cope with this kind of task. SVM rearranges objects using kernels which are a set of mathematical functions [26]. The objects are mapped or transformed so that they can be easily separated by less complex planes.

4.3.2. K nearest neighbors (kNN)

This method relies on memory for saving instances with known outputs. Labels of these instances are known. When a new test instance is to be decided, the output of the closest known objects is examined. The majority of votes rule is applied. The output that has emerged on the majority among these neighbors is assigned to the test instance. The distance formula (Euclidean, Mahalanobis, etc.) and the number of closest objects that are being evaluated (k number) are important parameters for the kNN algorithm.

4.3.3. Decision tree / random forest

Decision trees are ML tools that are used for regression or classification of both continuous and discrete variables [26]. The structure of this ML model inspired its name. The decision tree mechanism is as follows: For each iteration, local regions are created recursively. It is a supervised, and hierarchical model [27]. The decision tree comprises of decision nodes and leaves. Each decision divides the data. Low entropy divisions are created in this manner. Appropriate sized tree generation requires expert knowledge [26]. A random forest is a variation of a decision tree that uses multiple trees instead of a single one.

4.3.4. Logistic regression

Logistic regression (LR) is a method of representing the probability of a discrete outcome provided by an input [28]. The most widely used LR model is the model with a binary output, such as true/false, yes/no, and so on. Multinomial logistic regression can be employed if there are more than two possible discrete outputs. LR is also a helpful analysis tool for classification problems where researchers try to decide whether a new sample fits best into a category.

4.3.5. Multilayer perceptron

The Multilayer perceptron (MLP) is a feedforward artificial neural network [29]. It has a minimum of three layers which are the input layer, hidden layer(s) and the output layer. The hidden layer (s) uses the activation functions to capture nonlinear data relations. Therefore, MLPs can discriminate between classes that are nonlinearly separated. We selected them as representative of a shallow neural network to compare with the traditional machine learning algorithms. Unipolar sigmoid function was used as an activation function in hidden volumes of MLP. We used a plain MLP with only one hidden layer.

5. Experimental results and discussion

Three different states were differentiated which are physically active, stressed and relaxed states by using machine learning classifiers. The performance metrics were defined first the results which are obtained with data coming from EDA, ACC, and PPG sensors were presented separately.

5.1. Metrics

To evaluate our system performance, we used four different metrics. These metrics are widely used in stress detection studies.

5.1.1. Classification accuracy

The first metric is accuracy. It can be calculated by dividing the accurately classified windows by the number of all windows of all participants (it is calculated for all tuples separately).

5.1.2. Precision, recall and f_Measure

We also computed and presented the precision, recall and F_measure metrics. Recall and precision metrics are employed to compute the F_measure. Their formulas are as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{1}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2}$$

$$F_{measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

5.2. Differentiation of physically active and relaxed states

First, physically active and relaxed states were differentiated. The classification accuracies are presented in Tables 1, 2 and 3. Since the movements of participants are constrained in relaxed states, it is more convenient to use acceleration data to differentiate them from physically active states. Furthermore, the EDA signal is also sensitive to movements and it is the most distinctive signal along with the acceleration signal. The heart activity signal obtained from the PPG sensor achieves the lowest accuracy for distinguishing these two states. This could be explained by the contamination of this signal by the increased artifacts during physical activity.

Table 1. The differentiation results of physically active and relaxed sessions by using the accelerometer sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.926	0.926	0.926	92.647	0.982
RF	0.899	0.897	0.897	89.706	0.957
kNN	0.905	0.904	0.904	90.124	0.911
SVM	0.921	0.912	0.911	91.176	0.914
LR	0.907	0.904	0.904	90.441	0.966

Table 2. The differentiation results of physically active and relaxed sessions by using the PPG sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.863	0.826	0.822	82.638	0.833
RF	0.851	0.833	0.831	83.333	0.912
kNN	0.917	0.917	0.917	91.667	0.954
SVM	0.894	0.889	0.889	88.889	0.848
LR	0.808	0.806	0.805	80.556	0.865

Table 3. The differentiation results of physically active and relaxed sessions by using the EDA sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.916	0.899	0.898	89.923	0.908
RF	0.860	0.860	0.860	86.046	0.938
kNN	0.869	0.868	0.868	86.821	0.864
SVM	0.927	0.915	0.914	91.473	0.913
LR	0.884	0.884	0.884	88.372	0.922

5.3. Differentiation of physically active and stressed states

Physically active and stressed states were differentiated. This classification is the most challenging one since the physiological reactions to them are similar to each other. The classification accuracies are presented in Tables 4, 5 and 6. The acceleration signal achieved the best performance while differentiating between these two states, which is expected because the physical activity intensity level is different in these sessions. Around 70% accuracies with heart activity signal were obtained and 60% accuracies with EDA signal were obtained. This proves the similarity of physiological reactions.

Table 4. The differentiation results of physically active and stressed sessions by using the accelerometer sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.872	0.871	0.871	87.143	0.948
RF	0.909	0.907	0.907	90.491	0.959
kNN	0.818	0.814	0.814	81.428	0.897
SVM	0.911	0.907	0.907	90.714	0.907
LR	0.886	0.886	0.886	88.571	0.958

Table 5. The differentiation results of physically active and stressed sessions by using the PPG sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.702	0.660	0.641	65.97	0.659
RF	0.756	0.722	0.713	72.22	0.845
kNN	0.748	0.740	0.739	74.028	0.792
SVM	0.709	0.688	0.679	68.75	0.633
LR	0.774	0.764	0.762	76.389	0.762

Table 6. The differentiation results of physically active and stressed sessions by using the EDA sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.686	0.683	0.681	68.254	0.696
RF	0.694	0.690	0.689	69.047	0.709
kNN	0.688	0.659	0.645	65.873	0.696
SVM	0.635	0.635	0.635	63.492	0.635
LR	0.682	0.675	0.671	67.46	0.681

5.4. Differentiation of relaxed and stressed states

Finally relaxed and stressed states were differentiated. The classification accuracies are presented in Tables 7, 8 and 9. The accuracies with all modalities are around 75% and 80%. The results could be higher but since participants' perceived stress levels could be different than the expected stress level of the known context, this decreases the performance of our system. As an example, some participants might have lower stress levels in the TSST session, which will cause a false ground truth and misleading training data segments.

Table 7. The differentiation results of relaxed and stressed sessions by using the accelerometer sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.715	0.712	0.711	71.212	0.783
RF	0.783	0.780	0.780	78.03	0.852
kNN	0.770	0.765	0.764	76.5153	0.746
SVM	0.690	0.689	0.689	68.939	0.689
LR	0.680	0.681	0.682	68.261	0.690

Table 8. The differentiation results of relaxed and stressed sessions by using the PPG sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.810	0.809	0.809	80.9497	0.845
RF	0.824	0.819	0.818	81.918	0.886
kNN	0.849	0.841	0.840	84.119	0.911
SVM	0.825	0.822	0.821	82.232	0.822
LR	0.791	0.791	0.791	79.088	0.884

Table 9. The differentiation results of relaxed and stressed sessions by using the EDA sensor.

Method	Precision	Recall	F_measure	Accuracy	Area under ROC
MLP	0.735	0.720	0.715	71.969	0.811
RF	0.781	0.780	0.780	78.03	0.882
kNN	0.854	0.833	0.831	83.333	0.895
SVM	0.844	0.773	0.760	77.272	0.773
LR	0.697	0.689	0.686	68.949	0.776

5.5. Discussion

The main purpose of this study is to differentiate stress and physical activity. The most successful signal is the acceleration for this purpose. Heart activity has also around 75% accuracy for differentiating these two signals. Combining these two signals will achieve a more robust system for differentiating stress and physical activity. For detecting stress from relax sessions, heart activity achieves the best results. This proves that stress detection systems must monitor heart activity. The acceleration signal is especially needed in the presence of potential physical activity. EDA signal is affected by environmental conditions such as temperature and physical activity, and it could be easily contaminated, and that is why it could not achieve higher results than the other signals. The easiest task is to differentiate physical activity and relax states, and all signals can easily detect physical activity with over 90% accuracy. SVM and random forest algorithms are the most successful classifiers in these tasks. Furthermore, I selected the best classifiers for differentiating each tuple (relax-stress, physical activity-stress and relax-physical activity) and draw the ROC curves (see Figures 4,5 and 6) to compare the results visually.

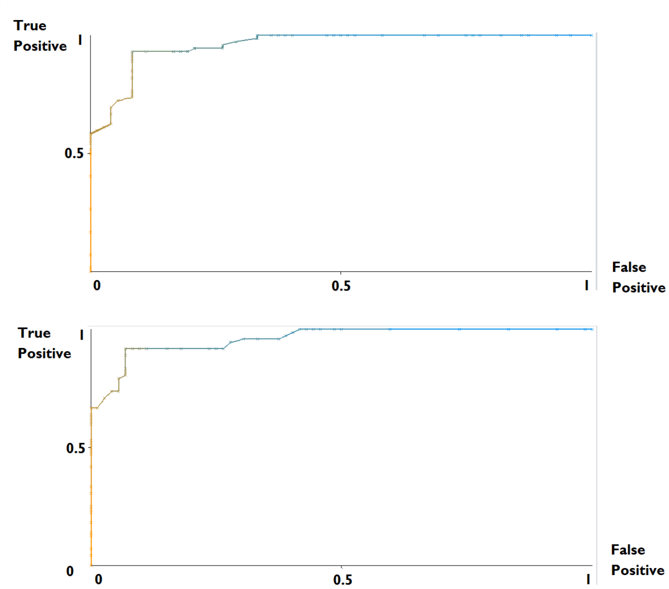

Figure 4. The ROC curves of the best performing classifier (kNN) using heart activity signal for differentiating physical activity and relax states. The top ROC curve belongs to the relax label and the bottom ROC curve belongs to the physical activity label.

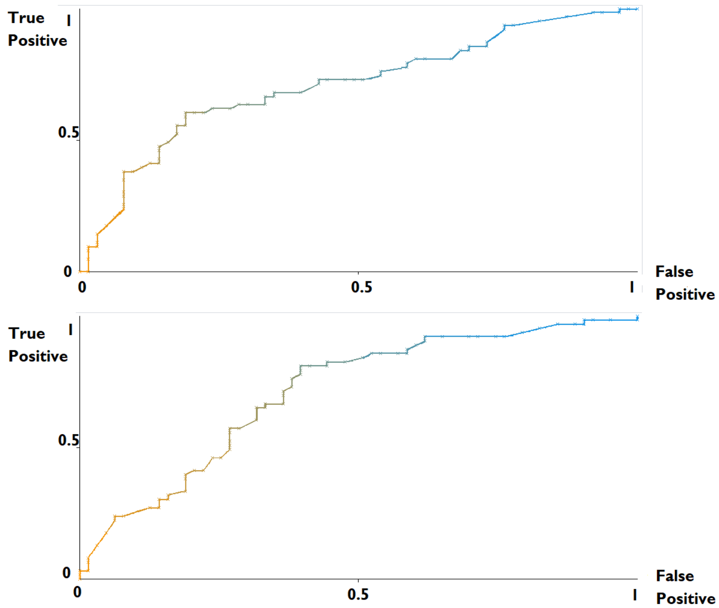

Figure 5. The ROC curves of the best performing classifier (random forest) using electrodermal activity signal for differentiating physical activity and stress states. The top ROC curve belongs to the stress label and the bottom ROC curve belongs to the physical activity label.

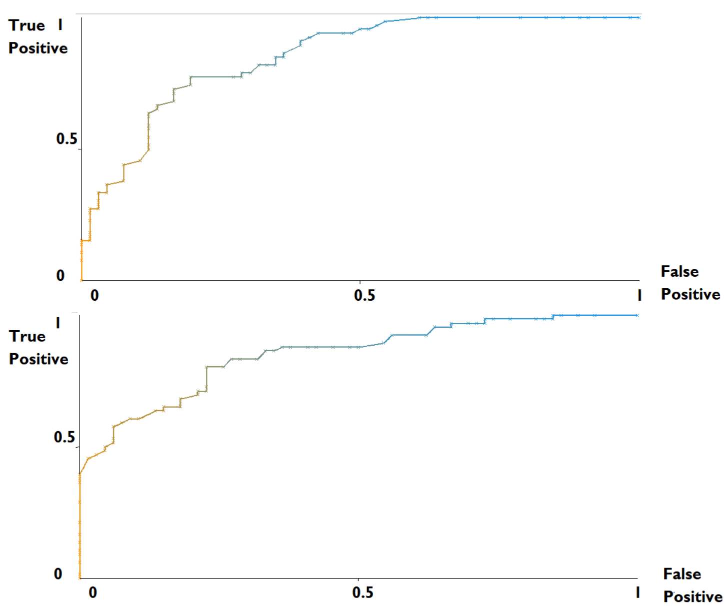

Figure 6. The ROC curves of the best performing classifier (random forest) using acceleration signal for differentiating relax and stress states. The top ROC curve belongs to stress label and the bottom ROC curve belongs to relax label.

We further compared our work with the studies that detect physical activity or stress levels (see Table 10). To the best of our knowledge, this study is the first one to differentiate stress and physical activity from physiological signals. I also achieved over 90% accuracy, which was aligned with the best-reported results in the literature for recognizing mental stress levels or physical activity levels alone. This promising result shows that the proposed system could be used to differentiate these two physiological responses with success.

Table 10. Comparison of studies that detect physical activity and mental stress levels.

Article	Wearable devices	Signals	Mental stress	Physical activity	Environment	Accuracy
[8]	LG watch and Samsung Gear S	Accelerometer	X	✓	Laboratory	%78.61
[7]	Polar M600, Huawei Watch, and Asus Zenwatch	Accelerometer	X	✓	Daily Life	%72.25
[9]	Samsung Gear S	Accelerometer	X	✓	Laboratory	%87
[13]	Empatica E4, Samsung Gear S, S2	HRV, EDA	✓	X	Real-Life Event	%90
[12]	Empatica E4	BVP, ST, HR	✓	X	Laboratory	%87.4
[10]	Empatica E4	BVP	✓	X	Real-Life Event	%81
[11]	Microsoft Band 2	HR, ST, EDA	✓	X	Laboratory	%85
Our study	Empatica E4	HRV, EDA,ST,ACC	✓	✓	Laboratory	%90.71

6. Conclusion

In this study, a physiological signal-based system that can differentiate physically active, relaxed and stressed states was developed. Artifact detection and removal, feature extraction and machine learning classification modules were implemented. The results for differentiating different tuples of these three states by using acceleration, heart activity and electrodermal activity signals were presented. When the physically active session is tried to be differentiated from the other two sessions, the accelerometer signal played an important role. Acceleration signal achieved around 90% accuracy for these classifications. Around 90% accuracies for classifying physically active and relaxed sessions with the EDA and 70% accuracy with the HRV signal were achieved. Stressed and relaxed sessions can be differentiated with approximately 80% accuracy with all modalities. The accuracies could be improved by analyzing the expected stress level of context and the actual perceived stress level of participants. The most challenging classification is between physically active and stressed states. A maximum of 70% accuracy by using HRV signal was achieved which shows that the heart activity pattern of physical activity and stress has differences. Multilayer perceptron and random forest are generally the most successful classifiers. Furthermore, this study is not without limitations. In order to generalize the conclusions, more experiments based on larger sample groups should be conducted. Furthermore, stressed, physically active and relaxed sessions were differentiated and the success of each modality is shown. However, the physiological signals were not analyzed when mental stress occurs during physical activity. It is planned to conduct more experiments by including this state. It is believed that the contribution of this research will be beneficial to both academia and industry, and it could be used to improve the stress detection algorithms running on commercial smartbands. This study could be used for a deeper understanding of physiological reactions of stressed and physically active states and it will guide researchers in developing more robust continuous daily life stress detection algorithms.

Acknowledgment

I would like to thank Professor Cem Ersoy for his contributions in this study and for managing the project.

References

[1] Hong JH, Ramos J, Dey A. Understanding physiological responses to stressors during physical activity. Proceedings of the ACM conference on ubiquitous computing 2012; 270-279.

[2] Liao W, Zhang W, Zhu Z, Ji Q, Gray WD. Toward a decision-theoretic framework for affect recognition and user assistance. Elsevier International Journal of Human-Computer Studies 2006; 64 (9): 847-873. doi: 10.1016/j.ijhcs.2006.04.001

[3] Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. Journal of American Medical Association 2007; 298 (14): 1685-1687. doi: 10.1001/jama.298.14.1685

[4] Plarre K, Raij A, Hossain S, Monowar A, Amin A et al. Continuous inference of psychological stress from sensory measurements collected in the natural environment. In Proceedings of the 10th ACM/IEEE international conference on information processing in sensor networks; Chicago, IL, USA; 2011. pp. 97-108.

[5] Li M. Multimodal physical activity recognition by fusing temporal and cepstral information. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2010; 18 (4): 369-380. doi: 10.1109/TNSRE.2010.2053217

[6] Wu W, Dasgupta S,Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. Journal of medical Internet research 2012; 14 (5): e130. doi: 10.2196/jmir.2208

[7] Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR mHealth and uHealth 2018; 6 (12): e10972. doi: 10.2196/10972

[8] Dobbins C, Rawassizadeh R. Towards Clustering of Mobile and Smartwatch Accelerometer Data for Physical Activity Recognition. Informatics 2018; 5 (2). doi: 10.3390/informatics5020029

[9] Davoudi A. Accuracy of samsung gear s smartwatch for activity recognition: Validation study. JMIR mHealth and uHealth 2019; 7 (2): e11270. doi: 10.2196/11270

[10] Hao T, Walter KN, Ball MJ, Chang HY, Sun S et al. StressHacker: towards practical stress monitoring in the wild with Smartwatches. In American Medical Informatics Association Annual Symposium Proceedings; Washington D.C., USA; 2017. pp. 830.

[11] de Arriba-Pérez F, Santos-Gago JM, Caeiro-Rodríguez M, Ramos-Merino M. Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables. Journal of Ambient Intelligence and Humanized Computing 2019; 10 (12): 4925-4945.

[12] Siirtola P. Continuous Stress Detection Using the Sensors of Commercial Smartwatch. In Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; London, UK; 2019. pp. 1198-1201.

[13] Can YS, Chalabianloo N, Ekiz D, Fernandez-Alvarez J, Riva G et al. Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches. IEEE Access 2020; 38146-38163. doi: 10.1109/ACCESS.2020.2975351

[14] Kirschbaum C, Pirke KM, Hellhammer D. The ‘trier social stress test’–a tool for investigating 20 psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993; 28 (1-2):76–81. doi: 10.1159/000119004

[15] Can YS, Chalabianloo N, Ekiz D, Fernandez-Alvarez J, Repetto C et al. Real-life stress level monitoring using smart bands in the light of contextual information. IEEE Sensors Journal 2020; 20 (15):8721-8730. doi: 10.1109/JSEN.2020.2984644

[16] Tarvainen MP et al. Kubios HRV - A Software for Advanced Heart Rate Variability Analysis. In 4th European Conference of the International Federation for Medical and Biological Engineering; Berlin, Germany; 2009. pp. 1022-1025.

[17] Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N et al. How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life. Sensors 2020; 20 (3).

[18] Taylor S, Jaques N, Chen W, Fedor S, Sano A et al. Automatic identification of artifacts in electrodermal activity data. In 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); San Fransico, CA, USA; 2015. pp. 1934-1937.

[19] Gjoreski M, Luštrek M, Gams M, Gjoreski H. Monitoring stress with a wrist device using context. Journal of Biomedical Informatics 2017; 73 (1): 159-170.

[20] Tarvainen MP, Niskanen J, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV–heart rate variability analysis software. Computer methods and programs in biomedicine 2014; 113 (1): 210-220.

[21] Alberdi A, Aztiria A, Basarab A. Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of biomedical informatics 2016; 59 (1): 49-75.

[22] Singh D, Vinod K, Saxena SC. Sampling frequency of the RR interval time series for spectral analysis of heart rate variability. Journal of Medical Engineering & Technology 2004; 28 (6): 263-272.

[23] Greco A, Marzi C, Lanata A, Scilingo EP, Vanello N. Combining Electrodermal Activity and Speech Analysis towards a more Accurate Emotion Recognition System. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019; 229-232.

[24] Setz C, Arnrich B, Schumm J, La Marca R, Tröster G et al. Discriminating stress from cognitive load using a wearable EDA device. IEEE Transactions on Information Technology in Biomedicine 2010; 14 (1): 410-417.

[25] Cinaz B, Arnrich B, Marca R, Tröster G. Monitoring of mental workload levels during an everyday life office-work scenario. Personal Ubiquitous Computing 2013; 17 (2): 229-239.

[26] Hill T, Lewicki P. Statistics: methods and applications: a comprehensive reference for science, industry, and data mining. Boston USA: Statsoft Incorporation, 2006.

[27] Alpaydın E. Introduction to machine learning. Cambridge, Massachusettes: The MIT Press, 2004.

[28] Edgar TW, Manz DO. Research methods for cyber security. Amsterdam, The Netherlands: Syngress, 2017.

[29] Can T. Gait Analysis and Fall Risk Assessment with Wearable Inertial Sensors. Ph.D., Bogazici University, Istanbul, TURKEY, 2019.