Measurement of multimodal physiological signals for stimulation detection by wearable devices

Measurement of multimodal physiological signals for stimulation detection by wearable devices

Gloria Cosoli[1], Angelica Poli[2], Lorenzo Scalise[1], Susanna Spinsante[2]

[1] Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, v. Brecce Bianche 60131 Ancona (Italy); g.cosoli@staff.univpm.it, l.scalise@staff.univpm.it
Corresponding author: Gloria Cosoli, g.cosoli@staff.univpm.it
[2] Department of Information Engineering, Università Politecnica delle Marche, v. Brecce Bianche 60131 Ancona (Italy); a.poli@staff.univpm.it, s.spinsante@staff.univpm.it

Highlights

- Acoustic stimulation detection useful in many application fields (e.g. assistive technologies)
- Signals generated by wearable devices can be used to detect human reactions
- Multimodal physiological signals are more effective and reliable respect to single assessment
- Machine learning algorithms effective in the stimulus presence identification through arousal

Abstract: The presence of stimuli and the consequent reactions undoubtedly reflect in experience-related changes of physiological parameters, which can be monitored by wearable devices. Generally, reactions related to the sympathetic nervous system activity are assessed through heart rate variability analysis. However, the exploitation of multimodal physiological signals provides a broader fingerprint. This study aims to identify the elicitation of acoustic stimulation through a wearable device; physiological signals, including electrodermal activity and skin temperature, were measured on a test population wearing a wrist-worn medical device. Eight machine learning algorithms were evaluated in a binary classification (presence/absence of stimuli), using 22 meaningful metrics from the collected data. The experimental results showed that Linear Regression (LR) algorithm, followed by Support Vector Machine (SVM), performed satisfactorily across all the evaluation metrics, achieving 75.00% and 72.62% of accuracy rate, respectively. Finally, the trained LR and SVM algorithms have been validated on a publicly available dataset (WESAD).

Keywords: Acoustic stimulation detection; wearable devices; measurement systems; multimodal physiological signals; features selection; machine learning.

---

1. Introduction

Human emotions can be usefully classified on a discrete scale according to Plutchick, who defined a taxonomy based on eight distinct emotions: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation [1]. Later on, given that an emotion can have different intensities, multi-dimensional space-models have been developed, taking into account both valence (pleasant/unpleasant) and arousal (high/low) [2]; subsequently, also dominance (submissive/dominant, reflecting the control ability of people) has been added [3]. The research on emotions, their recognition, and their elicitation through specific stimuli is an active field of research, where affective computing plays a pivotal role [4], even if there are also different application contexts, such as safe driving, health care, and social security [5]. Emotions can be thought as the subject's reaction to a stimulus, and this reaction undoubtedly reflects in unconscious changes in the subject's physiological state. Therefore, the monitoring of physiological signals could provide useful information on the presence/absence of stimuli influencing the subject's state. To this aim, different physiological signals can be considered [6], given that human psycho-physiological mental state is always correlated with physical and physiological reactions to internal/external stimuli [7]. In fact, if it is true that a subject can mask or pretend her/his facial and/or behavioural appearance (e.g. face expressions [8] or body gestures [9]), on the other hand it is undoubtable that the fluctuations of physiological signals are under the control of the sympathetic nervous system (SNS), thus cannot be controlled voluntarily [10]. For this reason, physiological sensors should be preferred to systems like cameras, which can push someone towards hiding emotional reactions, also depending on her/his own cultural habits [11]). To this aim, several signals have been used in the literature, such as the electrocardiogram (ECG) [12], the electroencephalogram (EEG) [13], the electromiogram (EMG) [14], the photoplethysmogram (PPG) [15], the electro-dermal activity signal (EDA) [16], and the skin temperature (SKT) signal [17]. All these signals can be obviously measured in ambulatory conditions by means of medical devices, but current healthcare provisioning paradigm is shifting towards remote monitoring and telemedicine [18], [19]. In particular, in the latest years wearable devices have gained more and more popularity, not only for activity tracking or fitness applications, but also in telemedicine and with clinical purposes [20]. Examples consist in their use in Ambient Assisted Living (AAL) domain [21], [22] or for promoting healthy and active ageing [23], [24], up to possible applications during a pandemic emergency [25]. However, the accuracy requirements should be always taken into account [20] in order to give a valuable contribution to decision-making processes [26], depending on the purpose of the measurement itself. Wearable devices allow to easily collect a great amount of data also in non-controlled environments [27], [28], representing a double-edge sword: on the one hand, they enable continuous (remote) monitoring, on the other hand the "big data" represent a challenge both for their accuracy and the computational requirements for their analysis. In the recent years, data coming from wearable devices have met the potentiality of Artificial Intelligence (AI) and Machine Learning (ML) approaches [29], empowering the capability of these devices to analyse physiological data for deriving significant parameters describing different human spheres, from physiological, through behavioural, to psychological ones, also for applications in Industry 4.0 context [30]. Furthermore, in the present COVID-19 pandemics, the combination of AI and big data is giving a valuable contribution in tracking people and trying to limit the contagion [31], [32].

Multimodal recordings of physiological signals can provide a broader fingerprint with respect to a single signal [33]. For example, Zhao et al. [34] used a wearable device to measure Blood Pressure Volume (BVP), EDA and SKT for emotion recognition through Support-Vector Machine (SVM) classifier; also Gjoreski et al. used the same model of wearable (Empatica E4), considering the same three signals plus accelerometer for stress monitoring [35]. It is worth underlining the high subjectivity of emotions, since their perception depends on many factors, among which we can mention experience, gender, age and culture [36]. Fear, surprise and stress were classified by Park et al. using SKT, EDA, ECG and PPG and different ML algorithms, namely linear discriminant analysis, classification and regression tree, self-organizing map, and Naïve Bayes, with 10-fold cross-validation [37]. In [38], after selecting the optimal set of features, joy, anger, sadness and pleasure were correctly discriminated, reaching a high recognition performance (i.e. 100% for joy and anger). Multimodal acquisition systems can also improve the motion artefacts identification, largely affecting the data quality, and consequently the analysis. Indeed, the reduction of noise and the increase of measurement accuracy are generally outcomes of a data fusion algorithm, including both physiological signals and a 3-axis accelerometer used as reference sensor to detect and estimate the movements [39], [40].

In this context, this article aims at investigating the possibility to correctly identify the presence or the absence of an acoustic stimulus by the use of signals measured by a wrist-worn wearable device, and exploring different ML approaches (namely Random Forest, Decision Tree, Naïve Bayes, K-nearest neighbour, Bagging, Boosting, Support Vector Machine, and Linear Regression), using features extracted from the measured data, whose effectiveness has been evaluated by means of the correlation-based feature selection method [41]. In our study, a wrist-worn, multimodal, medical device (Empatica E4) has been used to acquire a collection of physiological data (hereinafter referred as "Lab_dataset") from which features have been extracted to feed ML algorithms aiming at the detection of an acoustic stimulation. The overall workflow of the proposed work is described in Fig. 1.

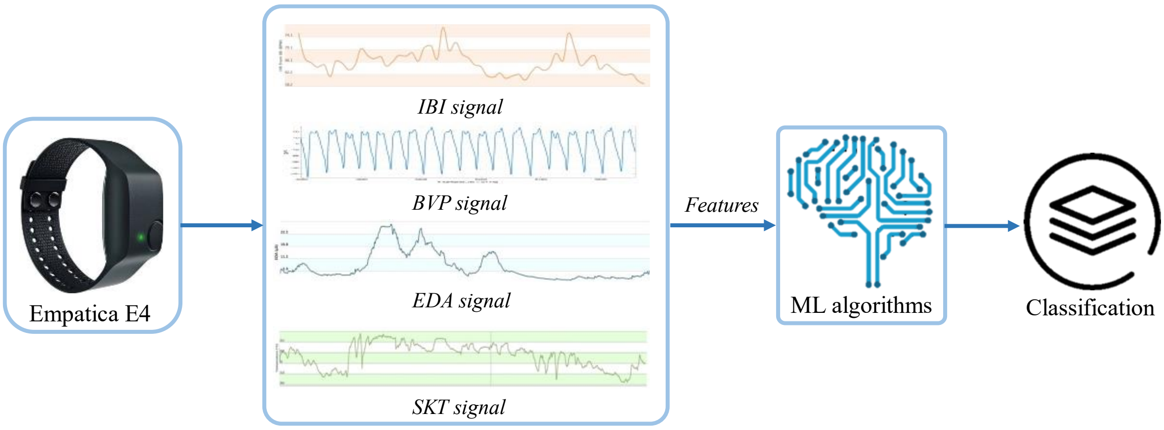

Fig. 1 Overall workflow of the proposed study.

In our study, we have considered the following physiological signals: Inter-Beat Intervals (IBI), BVP, EDA and SKT. In particular, the BVP signal is obtained through a PPG sensor and permits to derive Heart Rate Variability (HRV, i.e. the physiological variability in time intervals between consecutive heartbeats), which, among other purposes, can be exploited for emotion recognition [42]; it is worthy to note that, since PPG sensor is prone to motion artifacts (for example, in the literature Empatica E4 is said to correctly detect heartbeats during sitting and household work for 68% and 9% of the cases, respectively [43]), proper correction algorithms should be adopted [44]. Electrodermal activity reflects eccrine sweat gland activity [45] and can consequently provide the measurement of the so-called “emotional sweating”, independent from the thermoregulatory system, hence evaluating SNS function and limbic activity. EDA can be used for stress detection and emotion recognition [10], [46]; in particular, the relaxed state is characterised by a low variability signal with a decreasing trend, whereas the stressed state is characterised by a high variability and an increasing trend. It is important to note that after the electrodes application the conductivity increases over time until reaching the subject’s skin conductivity value; this process takes time, hence before recording it should be waited at least for 20 minutes [46] – even if others consider only 5 minutes [47]. The signal acquired with wearable devices is vulnerable to several types of disturbances; artifacts can derive from electronic noise or variations in contact between electrodes and skin [48] and they need to be recognised and corrected. The EDA signal (and particularly the magnitude of its changes [49]) seems to be associated more to arousal (i.e. the level of emotions, low/high) than to valence (i.e. the pleasantness of emotions, positive/negative). Moreover, given that arousal can have similar intensity of response for positive/negative stimuli (this can be confirmed through Self-Assessment Manikin – SAM – questionnaire, enabling to assess valence, arousal and dominance associated to a presented stimulus [50]), the authors have chosen to consider a binary classification of presence/absence of stimuli starting from the measured physiological signals, also considering that the accuracy of arousal discrimination seems to be higher than that of valence differentiation [49]. In the literature it has been evidenced that the combination of EDA and HRV measurements can provide an improved quantification of sympathovagal balance, since EDA is related to sympathetic activity, whereas HRV spectral estimates to parasympathetic one [51]. Moreover, also SKT data have been considered, since it can vary consequently to changes in blood flow associated to the modulation of local vascular resistance operated by the smooth muscle tone, which in turn is mediated by the SNS [34]. For the data processing, the authors used the Kubios software tool [52] for HRV analysis and algorithms for BVP, EDA and SKT, employing the Biosignal-Specific Processing (Bio-SP) Tool from Matlab® [53], [54] for the analysis of EDA and the extraction of its related features. In order to elicit emotions, audio stimuli chosen from the International Affective Digitized Sound system (IADS-2) database, which are already classified in terms of valence and arousal through the SAM scale, were used [55].

Finally, in order to validate the proposed measurement method, it was tested on a publicly available multimodal dataset, the Wearable Stress and Affect Detection (WESAD), which is available online and it is commonly used for stress and affect detection from data acquired by means of wearables [56]. Specifically, the validation phase aims to provide an additional analysis, thanks to the WESAD dataset collected under daily external stimuli in free living conditions.

The paper is organized as follows: Section 2 describes materials and methods adopted in this research, whereas results are presented in Section 3. Finally, in Section 4 the authors comment on the results and provide conclusions.

2. Materials and Methods

A. Participants

The experimental tests involved a population of 7 subjects (5 females and 2 males) in healthy conditions, aged between 15 and 52 years and with a Body Mass Index (BMI) between 20.8 and 26.8 kg/m^2. All the participants were informed about the purpose and the methods of the experiment and decided voluntary to be part of the study, by signing the informed consent before starting the tests (which were performed following the principles outlined in the WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [57]).

B. Experiment procedure and data collection

Since the measurement protocol was intended to acquire data in a relaxed situation, the participants were tested individually at their own home, in a quiet room with lights turned off to reduce any potential interfering input. In particular, they were in supine position on a bed with closed eyes, in order to focus on the presented stimuli and avoid disturbances on the measured data due to movements.

In this study, the individual physiological changes were simultaneously measured by wearing a wrist-worn, medical device (Empatica E4 - Fig. 2) [58], on the dominant wrist.

Fig. 2 Wrist-worn medical device (mod. Empatica E4 device): front (left) and back (right) view.

In order to induce an emotional reaction, audio recordings were extracted from the International Affective Digitized Sounds 2nd Edition (IADS-2) standardized database, in which 167 natural sounds of daily life are categorized in terms of arousal, valence and dominance [55]. The following sounds were selected: 1 neutral event (i.e. Walking sound no. 722), 1 pleasant event (i.e. Rock'n'roll sound no. 815) and 1 unpleasant event (i.e. Scream sound no. 275). Because of the short duration (6 s) of audio recordings in the database, all sounds were repeated to obtain 1-minute long stimuli. Participants completed three trials, each lasting 10 minutes. In detail, after recording 5 minutes of baseline of physiological parameters, one sound was randomly selected among the others, and played through a Bluetooth speaker for 1 minute to collect physiological changes through the wearable device. It is worthy to underline that the sensors were positioned at least 5 minutes before starting the acquisition, in order to reach stable measured data, e.g. for EDA signal. Then, 4-minute-long signals were collected in rest condition. All the subjects listened to the three audio clips twice, for a total of 42 recordings. Simple self-annotations, by pressing the event-marker button on E4 device, were performed by the subjects to label the beginning and the end of each acoustic stimulation. Thanks to data-labelling, for each measurement we well-categorized the first and the second part of acquisitions, i.e. in absence and in presence of stimulus, respectively. From the beginning of the stimulation to the end of the recording, the data portion was labelled as presence of stimulus due to the potential prolonged psychological reaction after the stimulus [59].

C. Acquisition device

The wrist-worn medical device utilised in our experiment is the Empatica E4, a wireless multi-sensor device (Class IIA Medical Device according to the 93/42/EEC Directive) for comfortable and real-time data acquisition. The E4 has four embedded sensors, namely PPG, EDA, 3-axial MEMS accelerometer and infrared (IR) thermometer. According to the definition of the acquisition protocol, i.e. at rest physical conditions of users, the signals acquired from the accelerometer sensor were not included in this study. All the others were included in the analysis, specifically:

- PPG sensor samples at 64 Hz. In particular, BVP is the input signal to the algorithm that provides the Heart Rate (HR) and the IBI signal as outputs. The digital sensor output, with a resolution of 0.9 nW/Digit, is generated by the light produced with 4 light emitting diodes (LEDs, 2 green and 2 red ones) and 2 photodiodes with a total sensitive area of 14 mm^2. The light during the green exposure mainly contains the information on the heartbeats, while the red exposure helps the reduction of motion artifacts that are dynamically compensated by firmware.
- EDA sensor measures the changes in skin electrical conductance with a sampling rate of 4 Hz, in the range of [0.01, 100] μS and with a resolution of 900 pS. Through the Ag/AgCl electrodes placed on the ventral wrist, a small alternating current (8 Hz frequency – max 100 μA) is applied to the user's skin.
- IR thermometer is configured with a sample frequency of 4 Hz. SKT values are measured by an optical thermopile sensor. The reported accuracy within the range of human skin temperature (i.e. 36-39 °C) is ± 0.20°C. Calibration is valid in the range [-40, 115] °C.

The multi-parameter sensor can operate either in streaming mode for real-time data visualization using a Bluetooth Low Energy (BLE) interface and the *E4 Realtime* app from supported mobile devices, or in-memory recording mode storing temporarily data in the internal flash memory (memory capacity > 48 hours of continuous data). In any case, the recording sessions associated to the serial number of the used Empatica E4 are saved as available from the *E4 Connect* remote platform for data management and can be download as .csv files.

D. Data pre-processing

As mentioned above, the approach proposed in this study analysed the following data: IBI, EDA, SKT and BVP. The pre-processing pipeline, first in MATLAB environment and then in WEKA [60], included mainly the following phases: segmentation, filtering, feature extraction and features selection. Then, these features were used to feed 8 different machine learning algorithms, which will be detailed in the following.

Firstly, raw data were split into two segments, representing the first and the second part of acquisitions, labelled as absence and presence of stimulus, respectively. Each segment lasted 5 minutes. Secondly, to reduce the artifacts and interference recorded during the acquisition phase, each measured signal was separately filtered according to the related literature.

IBI data, consisting in the duration (in ms) of successive heartbeats, were correctly and successfully reconstructed in our previous work [44]. Therefore, we used the same artifact correction method and extracted the same meaningful features through the Kubios toolbox [61]. EDA signals were pre-processed by using the Bio-SP toolbox [62] to reduce noise and artifacts attributable to wrist motion and physiological properties at skin level. More specifically, a Gaussian low-pass filter (with a 40-point window and a sigma of 400 ms), as recommended in a previous study, was applied for the filtering process of EDA [63]. Since SKT data change slightly and slowly under rest physical condition, as in the defined acquisition protocol, no filter was applied. The same for BVP data, also to avoid potential signal distortion.

E. Features extraction

In the features extraction phase, meaningful metrics were extracted in both time and frequency domains, from each data portion (i.e. absence/presence of stimulus) to characterize the different segments of data. Additional non-linear and information theory-based parameters were extracted from IBI signal according to the Kubios toolbox. Regarding the BVP signal, the features were extracted from the blood volume amplitude (BVA) representing the blood flow, while EDA features were computed on both the filtered EDA signal and the skin conductance response (SCR) component, strictly related to the SNS activity and stimuli response. Table 1 details the 33 statistical features, successfully applied in previous works [64], [65] and selected to quantify each signal.

Table 1: Features extracted from the physiological signals grouped by the domain.

Signals	**Features**	
IBI	Time Domain	IBIs mean (RR_mean, ms), IBIs standard deviation (RR_std, ms), HR mean (HR_mean, bpm), HR min (HR_min, bpm), HR max (HR_max, bpm), root mean square of successive IBIs (RMSSD, ms)
	Frequency domain	Absolute powers in the VLF (0-0.04 Hz), LF (0.04-0.15 Hz), HF (0.15-0.40 Hz) frequency bands (VLF_abs, LF_abs, HF_abs, ms^2), total power (P_tot, ms^2)
	Non-linear and information theory-based measures	HRV short- and long-term variability (SD1, SD2, ms) and their balance SD2/SD1, approximate and sample entropy (ApEn, SampEn), short- and long-term fluctuation analysis (alpha1, alpha2)
BVP	Time Domain	Mean and standard deviation of BVA (BVA_mean, nW, and BVA_std, nW, respectively), mean and standard deviation of the BVP 1^{st} derivative (BVA'_mean, nW/s, and BVA'_std, nW/s, respectively), mean and standard deviation of the BVA 2^{nd} derivative (BVA''_mean, nW/s^2), and BVA''_std, nW/s^2, respectively)
EDA	Time Domain	SCR mean duration (SCR_D_mean, s), SCR mean amplitude (SCR_A_mean, µS), SCR mean rise-time (SCR_RT_mean, s), EDA mean signal (EDA_mean, µS), no. of SCRs (SCR_n)
SKT	Time Domain	Mean (SKT_mean, °C), standard deviation (SKT_std, °C), minimum (SKT_min, °C) and maximum (SKT_max, °C) values

F. Features selection

A preliminary investigation was conducted using the full set of features. In fact, although a higher number of features means greater information available to the algorithms, the use of redundant and irrelevant features results in a poor classification performance [66]. Such a strategy was applied in [67], where the correlation-based algorithm, along with the genetic algorithm, was able to reduce the features from more than 300 to 69, limiting the computing time while increasing the ML algorithm performance. Therefore, in order to optimize the learning accuracy, we performed a feature selection analysis with the correlation-based filter [41], a common and reliable technique for features selection [68]. Such algorithm evaluates the worth of an attribute by measuring the (Pearson's) correlation of the attribute with the class to predict. The output of Pearson's correlation varies between high positive correlation (values close to 1) and high negative correlation (values close to -1) – 0 means no correlation. Generally, the threshold from which defining two variables correlated can be arbitrarily selected

[69]; hence, after running the correlation filter in WEKA tool to generate a rank for each feature, all the attributes with a rank below 0.10 were discarded.

G. Emotions and physiological reactions: stimulation detection by machine learning algorithms

After estimating the informative content of the physiological signals by extracting the related features, the correlation-based filter was used to establish the goodness of features extracted and to select a subset according to the high correlation between features and class. The resulting subset of features was used as input to the ML classifiers, from which an output identifying the class label related to the presence/absence of a stimulus is returned.

For the model evaluation, the ML algorithms were tested with the 10-fold cross validation configuration setting, in which the entire features dataset was randomly divided into 10 subsamples, namely 9 subsamples as training data and 1 as validation data for testing the model. After using the available subsamples as validation data, the resulting accuracy percentage is the average over the 10 iterations.

Given the large inter-individual variability for the physiological signals, especially considering the emotion recognition (indeed, individual differences can hinder emotion pattern discrimination [34]), several well-known classifiers were selected in WEKA tool for the classification task: Support Vector Machine (implemented with SMO algorithm in WEKA, that is Sequential Minimal Optimization), Random Forest (RF), Decision Tree (J48), Naïve Bayes (NB), K-nearest neighbour (kNN), Bagging, Boosting (LogitBoost) and Linear Regression (SimpleLogistic).

Finally, the performance of these algorithms was evaluated in terms of the ability of classifying the presence or absence of a stimulation, as a binary classification task. In detail, the classification performance was evaluated in terms of Accuracy (1), Sensitivity (2), Precision (3) and F-measure (4), as defined in [70]:

$$Accuracy = \left(1 - \frac{|N_{cci} - N_{ti}|}{N_{ti}}\right) \cdot 100 \tag{1}$$

being N_{cci} the number of correctly classified instances and N_{ti} the number of total instances considered by the classifier. The Sensitivity was computed as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

and Precision as:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

where TP denotes True Positives, FN False Negatives, and FP False Positives, respectively.

$$F - measure = \frac{(1 + \beta^2) \cdot Sensitivity \cdot Precision}{\beta^2 \cdot Sensitivity + Precision} \tag{4}$$

F-measure is the harmonic mean of Sensitivity (or equivalently Recall) and Precision, with the weight coefficient $\beta = 1$ in WEKA tool.

Additionally, for a visual interpretation among the classification results for the two classes, a confusion matrix has been realized to compare the predicted and actual classes.

H. Validation of the best performing algorithms on a different dataset

After having determined the best subset of features and the two best ML algorithms in terms of stimulation detection, trained and tested on the dataset acquired through Empatica E4 as described in Section 2. B, the authors validated their performance on a public dataset, namely WESAD. WESAD includes data measured on 15 participants (three females and twelve males, aged 27.5 ± 2.4 years, expressed as mean ± standard deviation) during laboratory studies aimed to detect whether stress or affect (both representative of daily external stimuli) from data acquired by means of wearable devices. In particular, Empatica E4 wrist-worn device was used to measure BVP, EDA, SKT and acceleration signals (as already described in Section C). Moreover, the chest-strap device RespiBAN [71] was employed for the measurement of ECG, EDA, EMG, SKT and acceleration signals; the sampling frequency was equal to 700 Hz for all of them. Laboratory studies were performed during three different affective states, namely neutral, stress and amusement; in particular, funny video clips were showed during the amusement conditions, whereas the Trier Social Stress Test (TSST) was used to cause a high mental load and, consequently, a stress condition. In order to have a ground truth, self-assessment questionnaires were administered to the participants (also to verify the effective elicitation of the different affective states); specifically, Positive and Negative Affect Schedule (PANAS, for positive and negative affect assessment) and some items from the State-Trait Anxiety Inventory (STAI, for the anxiety level quantification) and from the Short Stress Questionnaire (SSSQ, for the type of stress determination) were considered. Since the authors aimed to analyse the WESAD dataset following the same procedure described above for Lab_dataset, only data acquired through the Empatica E4 on 7 subjects in baseline and the stress conditions were selected among the signals collected. Additionally, in this case the ML algorithms were trained on our dataset and tested on WESAD one. This means that the ML algorithms were trained with data collected in rest physical state under acoustic stimulation, and then tested with data recorded in tasks typical of free-living conditions, which can be associated to daily external stimuli. More specifically, in the WESAD experimental protocol, Schmidt et al. [56] made the volunteers perform a public speaking (a 5-minute speech in front of a three-person panel, speaking of their own personality, highlighting strengths and weaknesses) and a mental arithmetic task (to count from 2023 to zero, with steps of 17) to elicit stress. This widens the application context where the algorithms were trained (i.e. laboratory controlled conditions with emotions elicited by audio stimuli).

3. Results

Data measured from Empatica E4 were pre-processed and analysed to extract features from each recording. More specifically, attributes were computed for the two data portions, which were labelled as absence (i.e. before playing the sound clip) and presence of stimulus (i.e. during and after playing the sound clip). Then, the selected features for feeding the ML algorithms are reported, together with the ranks obtained with the correlation-based feature selection method.

A. Features selection

Before the application of ML classifiers, the correlation-based feature selection algorithm was used to establish the importance of features over the whole dataset. The resulting subset of features included 22 metrics (from the highest to lowest ranked attributes) as shown in Fig. 3.

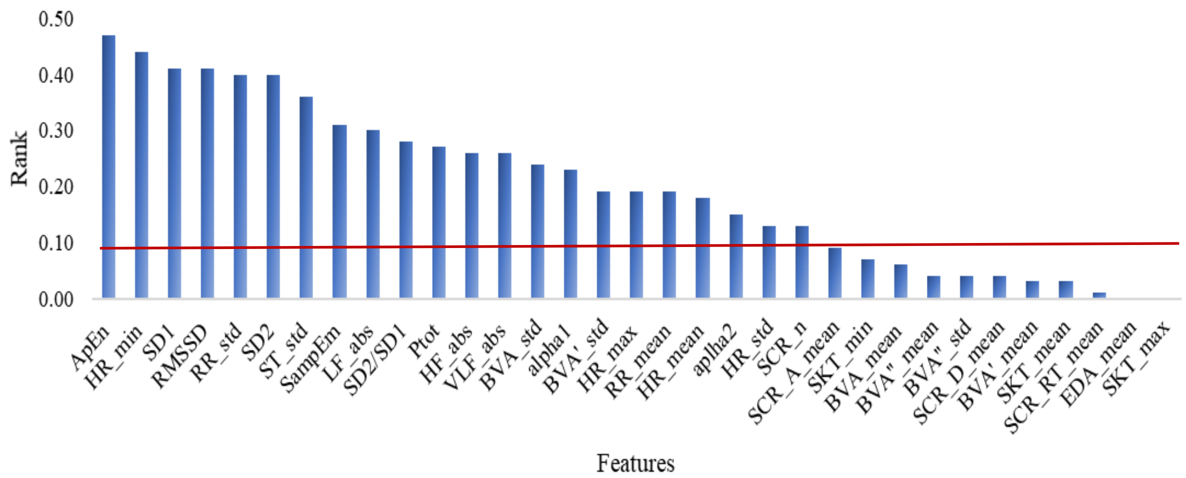

Fig. 3 Ranks listed in order of importance for each feature extracted from the Empatica E4 signals (the red horizontal line indicates the threshold of 0.10 over which the features are considered significant).

In summary, the features selected to feed ML algorithms were the following ones:

- HRV analysis (from IBI signal): mean and standard deviation of RR intervals (RR_mean and RR_std), mean, standard deviation, minimum and maximum values of HR (HR_mean, HR_std, HR_min, HR_max), root mean square of successive RR intervals (RMSSD) in the time domain; absolute powers in the VLF, LF and HF frequency bands (VLF_abs, LF_abs, HF_abs) and total power (P_tot) in the frequency domain; short- and long-term variabilities (SD1 and SD2) and their ratio as balance (SD2/SD1), approximate and sample entropies (ApEn and SampEn) and short- and long-term fluctuation analysis (alpha1 and alpha2) as non-linear and information theory-based features;
- BVP signal: the standard deviation of vasoconstriction (BVA_std) and the standard deviation of the 1st derivative of BVP signal (BVA'_std);
- EDA signal: the number of SCRs (SCR_n);
- SKT signal: the standard deviation of skin temperature (ST_std).

Such potential optimal subset of features, composed by the parameters more sensitive to stimuli, was examined for correctly discriminating the absence or presence of stimuli by means of different ML algorithms.

B. Classification of the presence/absence of acoustic stimuli

Eight different machine learning classifiers, trained with a 10-fold cross-validation scheme, were used for the binary classification of stimulation events. Herein, the performance of the proposed scheme to classify the presence or absence of acoustic stimulation (i.e. pleasant, unpleasant and neutral sound clips) are listed. Each algorithm was evaluated in terms of Accuracy percentage, Sensitivity, Precision and F-measure (Table 2).

Table 2: Performance of classifiers in terms of Accuracy (%), Sensitivity, Precision and F-measure.

Classifiers	Accuracy (%)	Sensitivity	Precision	F-measure
Support Vector Machine	72.62	0.78	0.73	0.71
Random Forest	70.24	0.70	0.72	0.70
Decision Tree	71.43	0.71	0.71	0.71
Naïve Bayes	67.86	0.68	0.78	0.65
K-nearest neighbour	58.33	0.58	0.58	0.58
Bagging	70.24	0.70	0.71	0.70
Boosting	71.43	0.71	0.72	0.71
Linear Regression	75.00	0.75	0.76	0.75

The highest average accuracy (75.00%) was achieved for the Linear Regression classifier (grey-highlighted raw in Table 2), while an average accuracy of 58.33% was found concerning the worst classifier, i.e. K-nearest neighbour. On the other hand, the highest sensitivity was reported for Support Vector Machine classifier, even if its value (0.78) is very close to the Linear Regression one (0.75). Similar considerations can be done for precision: the highest value (0.78) was reached by Naïve Bayes, but Linear Regression one is comparable (0.76).

Regarding the two best classifiers, i.e. the Linear Regression followed by the Support Vector Machine, the number of correctly classified and misclassified instances is reported also by their confusion matrices (

Table 3 and Table 4, respectively).

Table 3: Confusion matrix related to Linear Regression.

		Predicted class	
		Absence of stimulus	**Presence of stimulus**
Actual class	**Absence of stimulus**	28	14
	Presence of stimulus	7	35

Table 4: Confusion matrix related to Support Vector Machine.

		Predicted class	
		Absence of stimulus	**Presence of stimulus**
Actual class	**Absence of stimulus**	21	21
	Presence of stimulus	2	40

A similar behaviour of misclassification was found in both the classifiers: poor ability to distinguish the absence of stimulus, that were often predicted as presence of stimulus (Linear Regression: 14 wrong instances; Support Vector Machine: 21 wrong instances, resulting in a lower FP number for the former). Contrarily, the presence of stimulus was well-classified with a low number of misclassification (i.e. Linear Regression: 7 wrong instances; Support Vector Machine: 2 wrong instances, resulting in a lower FN number for the latter).

C. Validation of Linear Regression and Support Vector Machine algorithms on WESAD dataset

The discrimination performance of the two best ML algorithms, i.e. Linear Regression and Support Vector Machine algorithms, resulted good also in the analysis of signals from the WESAD dataset. In particular, the classification accuracy was equal to 71.43% and 64.29%, respectively, which are values not so far to those found on our laboratory data (75.00% and 72.62%, respectively). The related performance metrics are reported in Table 5. Finally, the confusion matrices are reported in

Table 6 and Table 7, respectively.

Table 5: Performance of classifiers in terms of Accuracy (%), Sensitivity, Precision and F-measure (WESAD dataset).

Classifiers	Accuracy (%)	Sensitivity	Precision	F-measure
Support Vector Machine	64.29	0.71	0.63	0.67
Linear Regression	71.43	1.00	0.64	0.69

Table 6: Confusion matrix related to Linear Regression – WESAD dataset.

		Predicted class	
		Absence of stimulus	**Presence of stimulus**
Actual class	**Absence of stimulus**	7	0
	Presence of stimulus	4	3

Table 7: Confusion matrix related to Support Vector Machine – WESAD dataset.

		Predicted class	
		Absence of stimulus	Presence of stimulus
Actual class	Absence of stimulus	5	2
	Presence of stimulus	3	4

4. Discussion and Conclusions

In this study, a multimodal physiological system was proposed to analyse physiological signals measured with a wearable device on a test population before and after listening to three acoustic stimuli twice. In order to evaluate the effect of stimuli, causing reactions in the subjects' physiological state, a binary classification was considered to identify the absence and presence of acoustic stimulation More specifically, after extracting the meaningful features from the physiological signals, the correlation-based feature selection algorithm was used to both improve the detection rate of the acoustic stimulation and to limit the number of involved parameters. Specifically, 22 features were selected from the whole dataset. An interesting finding is that most of the features that achieved the best results (considering a rank > 0.20) were inherent to the HRV analysis derived from the IBI signal, except for the SKT_std, from SKT signal, which reached a quite high rank. Also, note that among the BVP features only the standard deviation of vasoconstriction value (i.e. BVA_std) and the standard deviation of the peaks detected on the 1st derivative of BVP signal were above the selected rank threshold (i.e. 0.10), while the others resulted not meaningful according to the correlation-based features selection method. This means that, for the stimulation detection, the dispersion of both SKT and BVP attributes resulted to be more important than their average values. Furthermore, it is worthy to note that the whole HRV analysis starts from BVP signal, from which IBI is directly derived by Empatica E4. Similarly, for EDA signal only the number of peaks in the phasic component (i.e. SCR_n) had a rank higher than the threshold: this confirms that SCR_n is the main expression of SNS activity from the EDA point of view. All the features below the threshold may not provide significant improvement in the classifiers performance.

Although the attributes related to HRV signal are the most discriminating for the classification of the presence/absence of acoustic stimuli events, by considering only a single physiological signal (i.e. HRV) the classification performance reaches low values. For example, comparing the multimodal recordings system proposed in this work and our previous study [44], the classification performance of SVM algorithm improved from 66.67% to 72.62%, as it is evident also looking at the related confusion matrices. This underlines the importance of using multimodal recordings to have a more complete description of the analysed situation, which can be particularly useful when data were acquired in free-living conditions, hence compensating also for the effect of movement artifacts (in particular on signals like PPG). However, comparing the results obtained for each classifier, Linear Regression reached the highest results followed by the SVM (which was employed in the authors' previous study [44]), which therefore confirms to be one of the most powerful in this type of analysis. Both Linear Regression, building linear logistic regression models, and SVM, implementing the optimization algorithm, are basically two-classification methods [72], [73]. As a result, they allowed to well-separate and distinguish the presence/absence classes of stimulation in our binary classification system, aligned with the literature [74], [75].

Further confirmation was observed from precision, sensitivity and from their harmonic mean (F-measure) that were evaluated for each single classifier. F-measure values confirmed the accuracy trend, achieving simultaneously the highest accuracy and F-measure with the Linear Regression approach. The good performances of the Linear Regression and SVM classifiers were also observed in detail, by examining both classified and misclassified instances (see the confusion matrices for more details). A point of interest is to notice how the presence of a stimulus is better classified with respect to the absence of a stimulus. A possible motivation is that the performance of stimulation detection across subjects depends on many factors, especially on individual characteristics (and the related reactions to stimuli), both physical and psychological ones measured by the wearable device. More specifically, the first portion of signal (absence of stimulus) can be considered as a physiological baseline (where the inter-subject variability plays a dominant role), while the second portion of signal (presence of stimulus) is a sum of the physiological baseline and physiological responses elicited by the specific stimulus, which prevails in determining the physiological response.

The remaining classifiers achieved lower values for all the considered metrics; a possible motivation of these unexpected low percentages may be a consequence of the small dataset size (the tested population consisted of 7 subjects, each performing 3 tests repeated twice), which affects the classification performance. On the other hand, this assumption could influence the high percentages of Linear Regression and SVM algorithms. For this reason, the second experiment implied the validation phase on signals extracted from the WESAD dataset. Not surprisingly, a discrepancy between the performances was observed, but the good performances of both Linear Regression and Support Vector Machine algorithms were confirmed, reporting accuracies of 71.43% and 64.29%, respectively. This means that the subset of features and the selected algorithms trained on the data collected by the authors in laboratory during elicitation with audio stimuli can be successfully applied to detect the presence of different stimuli, such as stressors from the TSST.

Despite the above-mentioned limitations, the proposed measurement method achieved a high performance in the detection of different types of daily external stimulation (e.g. acoustic and visual). Specifically, this work confirms that it is possible to recognize the elicitation of a stimulus through the variations in measured physiological signals acquired in a living home environment through a wearable device. This means that, similarly, the presence of stress can be easily detected and inferred in daily life, since the signals involved in the analysis are nearly the same. Therefore, this point opens to some developments for the study of multimodal physiological systems to assess the human psychological, physical and social well-being. In particular, future works may be conducted by evaluating the proposed methodology on data gathered during emotional tasks from a wider population (which, moreover, would be useful to counterbalance eventual side effects, e.g. movement artifacts in PPG signals), also using different models of wearable devices in order to evaluate the influence of their metrological characteristics on the results. Furthermore, tests performed on a wider population could provide a database large enough to train the classifier to distinguish the valence and arousal among external stimuli of different intensities, thus discriminating emotions, which was outside the scope of the present study.

References

[1] R. Plutchik, "The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89. Sigma Xi, The Scientific Research Honor Society, pp. 344–350.

[2] P. J. Lang, "The Emotion Probe: Studies of Motivation and Attention," *Am. Psychol.*, vol. 50, no. 5, pp. 372–385, 1995.

[3] A. Mehrabian, "Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression," *J. Psychopathol. Behav. Assess.*, vol. 19, no. 4, pp. 331–357, 1997.

[4] E. Yadegaridehkordi, N. F. B. M. Noor, M. N. Bin Ayub, H. B. Affal, and N. B. Hussin, "Affective computing in education: A systematic review and future research," *Comput. Educ.*, vol. 142, p. 103649, 2019.

[5] L. Shu *et al.*, "A review of emotion recognition using physiological signals," *Sensors (Switzerland)*, vol. 18, no. 7, 2018.

[6] J. A. Domínguez-Jiménez, K. C. Campo-Landines, J. C. Martínez-Santos, E. J. Delahoz, and S. H. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomed. Signal Process. Control*, vol. 55, p. 101646, 2020.

[7] W. James, "What is an emotion?," *Mind*, vol. 9, no. 34, pp. 188–205, 1884.

[8] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "ScienceDirect Emotion recognition using facial expressions," *Procedia Comput. Sci.*, vol. 108, pp. 1175–1184, 2017.

[9] S. Saha, S. Datta, A. Konar, and R. Janarthanan, "A study on emotion recognition from body gestures using Kinect sensor," in *International Conference on Communication and Signal Processing, ICCSP 2014 - Proceedings*, 2014, pp. 56–60.

[10] A. Martínez-Rodrigo, R. Zangróniz, J. M. Pastor, and M. V. Sokolova, "Arousal level classification of the aging adult from electro-dermal activity: From hardware development to software architecture," *Pervasive Mob. Comput.*, vol. 34, pp. 46–59, 2017.

[11] A. Murata, J. S. Moser, and S. Kitayama, "Culture shapes electrocortical responses during emotion suppression," *Soc. Cogn. Affect. Neurosci.*, vol. 8, no. 5, pp. 595–601, 2013.

[12] K. N. Minhad, S. H. M. Ali, and M. B. I. Reaz, "Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness," *J. Transp. Heal.*, vol. 7, pp. 75–89, 2017.

[13] J. Zhou, S. Chu, X. Li, F. Xiao, and L. Sun, "An EEG Emotion Recognition Method Based on Transfer Learning and Echo State Network for HilCPS," *Microprocess. Microsyst.*, p. 103381, 2020.

[14] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Inf. Fusion*, vol. 59, pp. 103–126, 2020.

[15] A. Goshvarpour and A. Goshvarpour, "Poincaré's section analysis for PPG-based automatic emotion recognition," *Chaos, Solitons and Fractals*, vol. 114, pp. 400–407, 2018.

[16] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, and P. Zontone, "Driver's stress detection using Skin Potential Response signals," *Meas. J. Int. Meas. Confed.*, vol. 122, pp. 264–274, 2018.

[17] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2004, vol. 3068, pp. 36–48.

[18] O. H. Salman, M. I. Aal-Nouman, and Z. K. Taha, "Reducing waiting time for remote patients in telemedicine with considering treated patients in emergency department based on body sensors technologies and hybrid computational algorithms: Toward scalable and efficient real time healthcare monitoring syste," *J. Biomed. Inform.*, vol. 112, p. 103592, 2020.

[19] A. S. Albahri *et al.*, "IoT-based telemedicine for disease prevention and health promotion: State-of-the-Art," *J. Netw. Comput. Appl.*, vol. 173, p. 102873, 2021.

[20] G. Cosoli, S. Spinsante, and L. Scalise, "Wrist-worn and chest-strap wearable devices: systematic review on accuracy and metrological characteristics," *Measurement*, p. 107789, Apr. 2020.

[21] Ó. Belmonte-Fernández, A. Puertas-Cabedo, J. Torres-Sospedra, R. Montoliu-Colás, and S. Trilles-Oliver, "An Indoor Positioning System Based on Wearables for Ambient-Assisted Living," *Sensors*, vol. 17, no. 12, p. 36, Dec. 2016.

[22] A. Almeida, R. Mulero, P. Rametta, V. Urošević, M. Andrić, and L. Patrono, "A critical analysis of an IoT—aware AAL system for elderly monitoring," *Futur. Gener. Comput. Syst.*, vol. 97, pp. 598–619, 2019.

[23] J. Helbostad *et al.*, "Mobile Health Applications to Promote Active and Healthy Ageing," *Sensors*, vol. 17, no. 3, p. 622, 2017.

[24] H. Lewy, "Wearable devices- from healthy lifestyle to active ageing," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015, vol. 2015-Novem, pp. 7748–7751.

[25] G. Cosoli, L. Scalise, A. Poli, and S. Spinsante, "Wearable devices as a valid support for diagnostic excellence: lessons from a pandemic going forward," *Health Technol. (Berl).*, 2021.

[26] H. Fouad, A. S. Hassanein, A. M. Soliman, and H. Al-Feel, "Analyzing patient health information based on IoT sensor with AI for improving patient assistance in the future direction," *Measurement*, vol. 159, p. 107757, 2020.

[27] X. Hu, F. Li, and J. T. D. Ng, "On the relationships between music-induced emotion and physiological signals," *Proc. 19th Int. Soc. Music Inf. Retr. Conf. ISMIR 2018*, pp. 362–369, 2018.

[28] A. Poli, G. Cosoli, L. Scalise, and S. Spinsante, "Impact of Wearable Measurement Properties and Data Quality on ADLs

Classification Accuracy," *IEEE Sens. J.*, p. 1, 2020.

[29] Z. Xu, B. Yu, and F. Wang, "Chapter 4 - Artificial intelligence/machine learning solutions for mobile and wearable devices," in *Digital Health*, S. Syed-Abdul, X. Zhu, and L. Fernandez-Luque, Eds. Elsevier, 2021, pp. 55–77.

[30] G. Cosoli, G. Iadarola, A. Poli, and S. Spinsante, "Learning classifiers for analysis of Blood Volume Pulse signals in IoT-enabled systems," in *IEEE MetroInd4.0&IoT*, 2021, p. accepted for presentation.

[31] S. Whitelaw, M. A. Mamas, E. Topol, and H. G. C. Van Spall, "Applications of digital technology in COVID-19 pandemic planning and response," *Lancet Digit. Heal.*, vol. 2, no. 8, pp. e435–e440, 2020.

[32] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3176, 2020.

[33] Y. Dai, X. Wang, P. Zhang, and W. Zhang, "Wearable biosensor network enabled multimodal daily-life emotion recognition employing reputation-driven imbalanced fuzzy classification," *Measurement*, vol. 109, pp. 408–424, 2017.

[34] B. Zhao, Z. Wang, Z. Yu, and B. Guo, "EmotionSense: Emotion recognition based on wearable wristband," *Proc. - 2018 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/UIC/ATC/ScalCom/CBDCo*, pp. 346–355, 2018.

[35] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, 2017.

[36] M. Egger, M. Ley, and S. Hanke, "Emotion Recognition from Physiological Signal Analysis: A Review," in *Electronic Notes in Theoretical Computer Science*, 2019, vol. 343, pp. 35–55.

[37] B. J. Park, E. H. Jang, S. H. Kim, M. A. Chung, and J. H. Sohn, "A study on autonomic nervous system responses and feauture selection for emotion recognition: Emotion recognition using machine learning algorithms," *BIOSIGNALS 2014 - 7th Int. Conf. Bio-Inspired Syst. Signal Process. Proceedings; Part 7th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2014*, pp. 116–121, 2014.

[38] P. Gong, H. T. Ma, and Y. Wang, "Emotion recognition based on the multiple physiological signals," in *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2016, pp. 140–143.

[39] Y. Zhang *et al.*, "Motion Artifact Reduction for Wrist-Worn Photoplethysmograph Sensors Based on Different Wavelengths," *Sensors*, vol. 19, no. 3, 2019.

[40] S. Nabavi and S. Bhadra, "A Robust Fusion Method for Motion Artifacts Reduction in Photoplethysmography Signal," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9599–9608, 2020.

[41] N. Gopika and A. M. Kowshalaya M.E., "Correlation Based Feature Selection Algorithm for Machine Learning," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, 2018, pp. 692–695.

[42] A. F. Bulagang, N. G. Weng, J. Mountstephens, and J. Teo, "A review of recent approaches for emotion classification using electrocardiography and electrodermography signals," *Informatics Med. Unlocked*, vol. 20, p. 100363, 2020.

[43] J. Pietilä *et al.*, "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities.," in *EMBEC & NBC 2017*, 2017, pp. 145–148.

[44] G. Cosoli, A. Poli, L. Scalise, and S. Spinsante, "Heart Rate Variability Analysis With Wearable Devices: Influence of Artifact Correction Method on Classification Accuracy for Emotion Recognition," in *I2MTC 2021 - 2021 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Proceedings*, 2020, accepted for presentation.

[45] M. Asahina, A. Poudel, and S. Hirano, "Sweating on the palm and sole: physiological and clinical relevance," *Clin. Auton. Res.*, vol. 25, no. 3, pp. 153–159, 2015.

[46] I. V. Bornoiu, R. Strungaru, and O. Grigore, "Intelligent system for emotion recognition based on electrodermal activity processing," in *IFMBE Proceedings*, 2015, vol. 45, pp. 70–73.

[47] M. Sperduti *et al.*, "The distinctive role of executive functions in implicit emotion regulation," *Acta Psychol. (Amst).*, vol. 173, pp. 13–20, 2017.

[48] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, “Automatic identification of artifacts in electrodermal activity data,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 1934–1937, 2015.

[49] J. Kim and E. André, “Emotion recognition based on physiological changes in music listening,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, 2008.

[50] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[51] S. Ghiasi, A. Greco, R. Barbieri, E. P. Scilingo, and G. Valenza, “Assessing Autonomic Function from Electrodermal Activity and Heart Rate Variability During Cold-Pressor Test and Emotional Challenge,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.

[52] “Biosignal Analysis and Medical Imaging Group - Kubios HRV.” [Online]. Available: http://kubios.uef.fi/.

[53] S. Ostadabbas, “Biosignal-Specific Processing (Bio-SP) Tool.” Sarah Ostadabbas (2020). Biosignal-Specific Processing (Bio-SP) Tool (https://www.mathworks.com/matlabcentral/fileexchange/64013-biosignal-specific-processing-bio-sp-tool), MATLAB Central File Exchange. Retrieved November 27, 2020., 2020.

[54] M. Nabian, A. Nouhi, Y. Yin, and S. Ostadabbas, “A biosignal-specific processing tool for machine learning and pattern recognition,” in *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, 2017, pp. 76–80.

[55] M. M. Bradley and P. J. Lang, “The International Affective Digitized Sounds (2nd Ediotion; IADS-2), Affective Ratings of Sounds and Instruction Manual.” NIMH Center for the Study of Emotion and Attention, 2007.

[56] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 400–408.

[57] W. – T. W. M. Association, “WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects.” .

[58] “E4 wristband | Real-time physiological signals | Wearable PPG, EDA, Temperature, Motion sensors.” .

[59] A. Poli, A. Brocanelli, S. Cecchi, S. Orcioni, and S. Spinsante, “Preliminary Results of IoT-Enabled EDA-Based Analysis of Physiological Response to Acoustic Stimuli,” in *EAI Healthy IoT Conference Proceedings*, 2020.

[60] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software,” *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, p. 10, Nov. 2009.

[61] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen, “Kubios HRV – Heart rate variability analysis software,” *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, 2014.

[62] M. Nabian, Y. Yin, J. Wormwood, K. S. Quigley, L. F. Barrett, and S. Ostadabbas, “An open-source feature extraction tool for the analysis of peripheral physiological data,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, 2018.

[63] A. Greco, G. Valenza, and E. P. Scilingo, “Electrodermal Phenomena and Recording Techniques,” *Advances in Electrodermal Activity Processing with Applications for Mental Health*. Springer International Publishing, pp. 1–17, 2016.

[64] P. J. Bota, C. Wang, A. L. N. Fred, and H. Plácido Da Silva, “A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals,” *IEEE Access*, vol. 7, pp. 140990–141020, 2019.

[65] A. Goshvarpour and A. Goshvarpour, “The potential of photoplethysmogram and galvanic skin response in emotion recognition using nonlinear features,” *Phys. Eng. Sci. Med.*, vol. 43, no. 1, pp. 119–134, 2020.

[66] A. Poli, S. Spinsante, C. Nugent, and I. Cleland, “Improving the Collection and Understanding the Quality of Datasets for the Aim of Human Activity Recognition,” Springer, Cham, 2020, pp. 147–165.

[67] G. De Leonardis *et al.*, “Human Activity Recognition by Wearable Sensors : Comparison of different classifiers for real-time applications,” in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018, pp. 1–6.

[68] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, “How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life,” *Sensors*, vol. 20, no. 3, p. 838, 2020.

[69] S. Rosati, C. M. Gianfreda, G. Balestra, L. Martincich, V. Giannini, and D. Regge, “Correlation based Feature Selection impact on the classification of breast cancer patients response to neoadjuvant chemotherapy,” in *2018 IEEE International Symposium*

on Medical Measurements and Applications (MeMeA), 2018, pp. 1–5.

[70] D. J. Cook, *Activity learning : discovering, recognizing, and predicting human behavior from sensor data*. John Wiley & Sons, 2015.

[71] “biosignalsplux | respiBAN Pro.” .

[72] V. Vapnik, S. Golowich, and A. Smola, “Support Vector Method for Function Approximation, Regression Estimation and Signal Processing,” in *Advances in Neural Information Processing Systems*, 1997, vol. 9, pp. 281–287.

[73] T. M. Mitchell, *Machine Learning*, 1st editio. McGraw-Hill Education, 1997.

[74] Vanlalhruaia, Y. K. Singh, and N. D. Singh, “Binary face image recognition using logistic regression and neural network,” in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2017, pp. 3883–3888.

[75] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, “A Deep-Learning Model for Subject-Independent Human Emotion Recognition Using Electrodermal Activity Sensors,” *Sensors*, vol. 19, no. 7, p. 1659, 2019.