



Mood classification through physiological parameters

Filippo Cavallo^{1,2} · Francesco Semeraro¹ · Gianmaria Mancioppi^{1,2} · Stefano Betti¹ · Laura Fiorini^{1,2}

Received: 10 December 2018 / Accepted: 22 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Future smart agents, like robots, should produce personalized behaviours based on user emotions and moods to fit more in ordinary users' activities. Besides, the emotions are also linked to human cognitive systems, thus their monitoring could be extremely useful in the case of neurodegenerative diseases such as dementia and Alzheimer. Literature works propose the use of music tracks and videos to stimulate emotions, and cameras to recorder the evoked reactions in human beings. However, these approaches may not be effective in everyday life, due to camera obstructions and different types of stimulation which can be related also with the interaction with other human beings. In this work, we investigate the Electrocardiogram, the ElectroDermal Activity and the Brain Activity signals as main informative channels, acquired through a wireless wearable sensor network. An experimental methodology was built to induce three different emotional states through social interaction. Collected data were classified with three supervised machine learning approaches with different kernels (Support Vector Machine, Decision Tree and k-nearest neighbour) considering the valence dimension and a combination of valence and arousal dimension evoked during the interaction. 34 healthy young participants were involved in the study and a total of 239 instances were analyzed. The supervised algorithms achieve an accuracy of 0.877 in the best case.

Keywords Mood recognition · Machine learning · Social robotics

1 Introduction

The interest in affective computing is rising along with the interest in the development of a Social Assistive Robot (SAR). The reasons, underlying this interest, are several. Above all, there is the necessity to enable robot behaviour to better interact with users (Picard et al. 2001). Besides, the ability to understand human affective state would allow the robot to comprehend other human behaviours far more, such as health decision making (Ferrer and Mendes 2018) or assistance in assisted living settings (Glende et al. 2016).

This double information would exploit the robot's efficacy and its acceptability (Cavallo et al. 2018a). Furthermore, the subjects mood monitoring could be extremely useful in several clinical cases, such as people suffering from mild cognitive impairment (MCI), for which rapidly mood changes

might imply worsening of cognitive (Gallagher et al. 2018). It is worth to mention that the Quality of Life consists of multiple dimensions, including physical health, psychological state, level of independence, social relationships and their relationship with salient features of their environment.

In this framework, SAR should be considered not only in their simple "assistive soul" meaning. Their use goes beyond a prosthetic approach for supporting only in case of need. Indeed they should be conceived also as active supporters, that empower, promote and guide elderly persons in better managing their health status and well-being. In this sense, healthy coaching technologies have become of great importance in the last decade and several works in this sector have been presented at research and industrial level (Turchetti et al. 2011).

The new generation of service robots is expected to have a cognitive behaviour, biological inspired, to provide a greater integration into the human community. Human beings adapt their actions and behaviours merging stimuli, coming from internal perceptions and the environment. Scientific literature presents some attempts to model the robot's behaviour according to human cognitive structures (Karami et al. 2016; Rodić et al. 2017; Dağlarlı

✉ Filippo Cavallo
filippo.cavallo@santannapisa.it

¹ The BioRobotics Institute, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio 34, 56025 Pontedera, PI, Italy

² Department of Excellence in Robotics & AI, Piazza Martiri della Libertà 33, 56127 Pisa, PI, Italy

et al. 2017). All these proposed behavioural models have in common the multidisciplinary aspects of the human beings which is the result of the physical and cognitive status, clinical traits and environmental stimulus. Indeed, according to literature evidence (Doux 1996), emotion supports the decision-making process. Emotions are also linked with several other cognitive functions, such as memory. The close relationship between amygdala and hippocampus determine the consolidation of emotional aspects in our memories (McGaugh 2018). Through these emotional markers embedded in our memories, as well-known, emotions lead our behaviour, allowing a human to use heuristics in planning their actions, instead of deciding in a more rigorous and expensive, in terms of cognitive effort, way (Damasio 1996).

Emotions seem to be also linked to human attentive systems. Subjects with the phobia, or even anxiety disorder, seem to be more reactive in pre-attentive arousing process if compared to healthy controls (van den Hout et al. 2000). Emotions, generally considered as irrational and maladaptive, are closely embraced to human cognitive functions and work with them synergistically, allowing the emergence of mind from brain.

Although, the modelling of emotion is an arising topic of great interest; up to now, in literature, there are several theories which try to describe emotions (i.e., Russel 1980; Ekman and Friesen 1999). Particularly, according to Russell et al. human emotions can be categorized in a two-dimensional space, whose axes are called valence and arousal. They are meant as connotation attributed to experience and the intensity of felt sensation, respectively. This theory is extended by the Mehrabian's PAD model (Mehrabian 1996), which adds a third dimension called dominance, that is the capability of the person to be in control of his emotional output. Consequently, intelligent machines need to embed emotion modelling to enhance and improve human-machine interaction. The design of the behavioural model should comply with anatomic, operative and functional requirements. Indeed, the robot should be able to perform all the requested services with advanced human-robot interaction abilities and should be able to adapt its behaviour to the user. In this context, a recent review paper (Nocentini et al. 2019) on the development of behavioural models for SAR remarks the necessity to include emotions, among other inputs, in the architecture of the robot.

In this framework, this work aims at the developing and the testing of a physiological wearable sensors system able to recognize three different moods elicited by social interaction. Particularly, this system is composed of three different physiological sensors able to monitor the electrodermal activity (EDA), the electrocardiogram (ECG) and the electroencephalogram signals. Three supervised machine learning classifiers with different kernels, namely Support

Vector Machine (SVM), Decision Tree (Tree) and K-nearest Neighbors (KNN) were used to test and to compare if our proposed system can classify the moods based on perceived valence and arousal.

2 Related works

The literature concerning the developing of SAR is getting richer of contributions concerning the use of emotion channels so to improve human-robot interaction (HRI). The technological solutions used to exploit this channel are several. The overwhelming majority of the study concerning affective computing take into account the use of vision sensor, so to detect facial expressions with emotive value (Agrigoroaie and Tapus 2017). Although their spread use, this technology presents some issues, among which, users' privacy limitations, camera occlusion, and adequate room lighting (Cavallo et al. 2018b).

Other works try to overcome these limitations by using another type of technology, for example Pahl et al. used microphones for detecting affective haptic inputs (Pahl and Varadarajan 2015). On the other hand, physiological sensors are very useful for emotion recognition and can successfully manage this kind of issues (Betti et al. 2017). Although their use is not yet perfect and might be affected by some other issues, such as movement artefacts or a not adequate sensor adhesion (Chen et al. 2016).

Set out below, are reported some of the most outstanding works about emotion recognition through the use of physiological parameters (see Table 1). It is worth to underline that the studies included in this overview are not rather older than a couple of years. Since this research field is enlarging quite fast, and the intention was to report just some emblematic cases, so to depict the framework in which the research was conducted.

For example, Al Machot et al. (Al Machot et al. 2018a, b) analysed data from a public dataset for emotions analysis called "Multimodal Database for Affect Recognition and Implicit Tagging" (MAHNOB). The data encompassed in the dataset have been gathered using emotional video clips so to elicit the mood. The subjects' assessment provided also for the use of the Self-Assessment Manikins (SAM) questionnaire, through which subjects stated their level of valence and arousal. Another work, which takes into account emotion recognition based on physiological data, is the study carried out by Chen et al. (2017). In this work authors propose a three-stage decision method starting from physiological signals. Similarly to (Al Machot et al. 2018a, b) the authors used a database called "Database for Emotion Analysis using Physiological Signals" (DAEP). Furthermore, instead of using only information from electrodermal activity, Chen et al. analyzed also information from the EEG,

Table 1 Comparison of related works

Authors	Sensors	Parameters	Methods	Classifier	Accuracy
Subramanian et al. (2018)	EEG (1) ECG (2) GSR (3)	(1) Mean number of peaks, mean derivative of the inverse channel signal, average number of peaks in the inverse signal. (2) PSDs, IBI, HR and HRV. (3) Mean skin resistance and mean of derivative, mean differential for negative values. Proportion of negative derivative samples, number of local minima in the GSR signal, average rising time of the GSR signal	Video clips	Linear SVM Naive Bayes (NB).	SVM: ECG and GSR:0.63 NB: EGC and GSR: 0.65
Al Machot et al. (2018a, b)	EDA	SCL mean, standard deviation, maximum, minimum, root mean square, mean of the first derivation, mean of the second derivation, and mean of the negative slope. SCR, response latency of the first significant SCR, the sum of SCR amplitudes and the sum of SCR areas	Video clips	SVM-RBF SVM-KNN	SVM-RBF: 0.63 SVM-KNN: 0.65
Moaied et al. (2017)	ECG (1) EEG (2)	(1) RRI time series. (2) waves, δ , θ , α , low- β , high- β , γ .	Music tracks	SVM	Dependent subjects average accuracy of 95.7% Independent subjects average accuracy of 86.9%
Chen et al. (2017)	EEG (1) GSR (2) TMP (3) BVP (4) RSP (5) EMG (6) EOG (7)	(1) Mean, Var, peak-to-peak amplitude, Skewness. Difference of average PSD in θ , α , β , and γ bands for 14 channel pairs between right and left scalp. nonlinear features: Spectral Entropy, Shannon Entropy and C0 complexity Mean, (2) features such as Rising time, Decay time Sc-Mean, Sc-Median, Sc-Var, Sc-MinRatio. (3) Features: Mean, IDiff-Mean, Spectral power in the bands (0–0.1 Hz) and (0.1–0.2 Hz) (4) features: Hr-Mean, Hr-Var, Hr-Range Hrv-Mean. (5) features such as Mean, Var, Range, MaxRatio IDiff-Mean. (6) features such as Mean, Var, Total spectral power (7) features: Mean, Var, peak-to-peak amplitude, Energy	Video clips	Decision three	Arousal 77.57% Valence 43.57%
Zhuang et al. (2017)	EEG (1)	EEG signals were decomposed into IMFs by EMD. Three features were studied: (1) fluctuation phase, (2) fluctuation time series, and (3) normalized energy	Music videos	SVM	Arousal: IMF1:72.1% IMF2:66.6% IMF3: 64.6% IMF4: 64.0% IMF5: 62.4% IMF1–2: 70.5% IMF1–3: 70.1% IMF1–4: 69.6% IMF1–5: 69.0% Valence: IMF1: 70.4% IMF2: 63.5% IMF3: 61.5% IMF4:59.5% IMF5: 55.7% IMF1–2: 69.0% IMF1–3: 68.5% IMF1–4: 68.0% IMF1–5: 67.6%

Table 1 (continued)

Authors	Sensors	Parameters	Methods	Classifier	Accuracy
Mehmood et al. (2017)	EEG (1)	(1) EEG signals in the time- and frequency-domains	IAPS	Hjorth parameters based optimal feature selection, voting ensembles method includes top ranked classifiers	Arousal/Valence 76.6%

GSR, EMG, and others. Once again the dimensions utilized to classify emotional state have been valence and arousal. On the other hand, Subramanian et al. (2018) present their database, which encompassed data from ECG, EEG, GSR, and cameras. These authors gathered also information about personality traits of subjects. They decided to stimulate the emergence of the emotional state through the use of movies clips as the majority of the studies about emotion recognition. Differently, Moyade et al. (2017), used music tracks so to induce emotional state. Although the elicitation method was different, the parameters utilized were again valence and arousal, for what concern the mood, and ECG and EEG measures for concerning the physiological counterparty. Again the authors decided to used others database to run their analysis.

Interesting, Zhuang et al. (2017) decided to do not use music tracks or just video clips, but rather music videos. The subjects involved in their experimentation watched 40 music videos. Each music video (1 min long) has different emotional valence (score ranged from 1 to 9). The results show that three features were suitable for emotion recognition, revealing that component Intrinsic Mode Function 1 (IMI1) extracted from EEG through emotional decomposition analysis plays the most important role in emotion recognition. Another work investigated the use of EEG components to develop a classifier for emotion recognition is from Mehmood et al. (2017). In this work, the authors used the International Affective Picture System (IAPS) to elicit the emotional response. Authors applied deep learning techniques to analyse EEG signal and classifying the emotional response reaching an average accuracy equal to 76.6%.

Summarizing, research on emotion recognition often uses the same methods to elicit emotions, such as video clips, still picture or music tracks. Moreover, the majority of the studies about it take into account others database and do not personally carry the experimented trials out. In addition, for what concern the emotion dimensions, the articles analyzed take into account, mostly, only two dimensions: arousal and valence.

As recently underlined in a review paper by Nocentini et al. (2019) researchers effort aim to apply the emotion recognition technique to enhance the HRI in a social environment, where the SAR represent an assistant, or even better a companion of daily life. Nevertheless, it is worth to mention that a branch of robotics is addressing the Industry 4.0 challenge (Lu 2017). Indeed, in this context, the robot does not represent an assistant but rather a co-worker which should be endowed with the capability to capture and understand accurately and robustly human request (Maurtua et al. 2017). Non-verbal clues are critical even in such environment, because they can create an intuitive and direct human–robot communication, which will facilitate the collaboration (Sheikholeslami et al. 2017). Recently, Hu et al.

(2019) present the iRobot-Factory: an intelligent robot factory based on cognitive manufacturing and edge computing. The main goal of the authors is to develop an intelligent and emotion-aware robot factory. The proposed system acquired audio-visual, physiological, and emotional data resources from the user, so to recognize the emotional state of the employee. In this manner, the experience of the subject could be fully evaluated and combined with the robotic co-worker. In this context, our paper aims to go beyond the state of the art because the dyadic interaction to stimulate the emotion in participants monitored with physiological parameters. We adopted the idea of a social protocol, more specifically of human–human interaction, because it represents a more ecological (life-like) setting for emotion elicitation, better than commonly used movie-clips or picture. However, it is important also to notice that, during an interactive task, social robots are perceived like social actors, thus they evoke expectations and mental models known from human–human interaction (Horstmann and Krämer 2019).

3 Methods

3.1 Instrumentation

In this study, three physiological signals were monitored: the Electrocardiogram (ECG), the Galvanic Skin Response (GSR) and the Brain Activity (BA). We use three commercial wearable sensors which minimize the trade-off between the accuracy of the measure and the obtrusiveness (Table 2). Zephyr BioHarness is a Bluetooth chest able to monitor the ECG signal and to calculate parameters such as Heart Rate and R–R Intervals. The MindWave headset is a Bluetooth device able to capture the BA employing a single frontal lobe channel and to compute the meditation and attention indexes. The Shimmer GSR Module measures the GSR by means of two fingers electrodes (Fiorini et al. 2018).

A customized interface developed with Visual Studio™ 2017 (Microsoft®, Washington, USA) managed the device connection, the data acquisition and the data storage (Fig. 1).

3.2 Experimental protocol

The proposed protocol aimed at evoking a positive or a negative mood in the users through social interaction with the experimenters during the administration of a questionnaire composed of 55 multiple-choice questions as described in (Fiorini et al. 2018).

In the positive conditions, the questions selected had been prepared to result funny and strange to make laugh without making the subject feel ignorant if he did not know the correct answer, whereas for the negative condition the selected

questions had high level of difficulty which should evoke discomfort and shame in the subjects.

Two experimenters with different backgrounds administrated the test. Particularly, a psychologist led the interaction by administrating the questionnaire and an engineer provided the complementary interaction and was ready to intervene in case of necessity.

At the beginning the user was asked to wear the sensors and, after that, he was asked to rest on a chair without stimuli to record the baseline (5 min) and the relax phase (5 min). All the participants were unaware of the real purpose of the test for all the duration of the protocol.

Then, the interaction started to stimulate the desiderate mood for a total of 15 min. Both experiments showed different attitude in accordance with the positive or negative condition. In the positive condition, they were affable and gave signs of appreciation, whereas in the negative test they had a mean and rude attitude. In both cases, there were five interventions planned at a fixed moment of the protocol to reinforce the administered condition. Indeed, these five moments were intended to gradually increase the evoked mood.

At the end of the test, the Self-Assessment Mannikin (SAM) picture-oriented questionnaire was proposed to the subject (Bradley and Lang 1994) to quantify the valence, arousal and dominance felt during the experiment. Then, the real purpose of the experiment was revealed and the experimenters asked the participant to refer the SAM's scores to a moment coinciding (or close to) one of the five reinforcement. Indeed in this kind of experiment it is really important that subject do not understand beforehand the real purpose of the experiment (Harmon-Jones et al. 2007).

Beck Depression Inventory (BDI) (BECK et al. 1961) and Revisited Maudsley Obsessive Compulsive Questionnaire (MOCQ-R) (Sanavio et al. 1986), adapted from (Hodgson and Rachman 1977) to fit the Italian population, were administrated. The answers of the BDI and MOCQ-R were used to ensure the mental stability of the subjects. Subject with abnormal physical conditions were not included in the study.

3.3 Participants

34 healthy participants (19 males, 15 females) ranged from 20 to 35 years (average \pm standard deviation 23.62 ± 2.69) were recruited at Scuola Superiore Sant'Anna of Pisa among student and research/teaching staff. Particularly, a description of the experiment purpose was sent to the students' mailing lists (i.e., master student and PhD student) asking for volunteers. All the interested subjects were asked to contact the corresponding author of this research thus to assess the recruitment criteria before organize the test.

The recruitment criteria for this study were: (1) the subjects should have at least the high-school diploma; (2) the subjects should live and be educated in Italy; (3) the subjects should not present altered psychic or bodily abnormal conditions; (4) the absence of a career background dealing with the field of neuroscience or medicine. The subjects were randomly assigned to the positive or negative tests, with the constraint to maintain a balance between the assigned test. The two conditions have been not administered to the same subject because the mood evoked during the first interaction can influence the second test.

At the beginning of the experimental session, written informed consent was obtained from the participants.

3.4 Data analysis

3.4.1 Features extraction

The physiological signals were acquired during the tests and off-line analysed. Firstly the signals were segmented in a time-windows frame of 180 s with 50 s of overlapping between adjacent to properly handle the transition. Then, each physiological signals were analysed to extract the proper features in time (t) and frequency (f) domains as described below.

The physiological data acquired during the experimentation were analysed offline using Matlab® R2018a (MathWorks, Massachusetts USA) to investigate variations in physiological parameters that could be attributed to stress states.

The Shimmer GSR Sensor output galvanic resistance, which was converted into SC. The signal was filtered using a moving average filter. The algorithm for feature extraction was based on startle detection and was related to the scoring multiples response method of Boucsein (2012), which established a local baseline at the level of the onset of the second response and measured the distance from that baseline to the following peak. The detection algorithm identified all occurrences when the first derivative exceeded the threshold of 0.005 μ S. Furthermore, to avoid subsequent startles, a minimum distance was chosen, considering that a startle event is expected to last approximately 1–3 s. Once the response was detected, the zero-crossings of the derivative preceding and following the response were identified as the onset and the end of the startle (Sharma and Gedeon 2012). Finally, a set of nine parameters were calculated. The Zephyr BioHarness™3 output a raw ECG signal and IBI data that specified the temporal distance between a beat and the following one. The IBI signal was modified, identifying and correcting ectopic rhythm, which is an irregular heart rhythm due to a premature heartbeat; thus, a Normal-to-Normal (NN) interval sequence appropriate for HRV analysis was obtained. Since the NN interval sequence is irregularly

Table 2 Wearable sensors description

Sensors	Frequency (Hz)	Signals	Data transmission
Zephyr BioHarness	250	ECG	Bluetooth
Shimmer GSR module	51.2	GSR	Bluetooth
MindWave	512	BA	Bluetooth
	1	Attention, meditation	

The name of the sensors, the operative frequency used in this work, the acquired signals (i.e., ECG, GSR, BA and attention/meditation) and the data transmission were detailed

sampled in time, for spectral analysis it was converted to an equidistantly sampled sequence (Mali et al. 2014). After a smoothing of the signal, the sequence was resampled at 4 Hz. Finally, eight temporal parameters and ten frequency parameters were computed. The MindWave headset mobile device provides a raw EEG signal, a power spectrum in different frequency ranges (alpha1: 8–9 Hz, alpha2: 10–12 Hz, beta1: 13–17 Hz, beta2: 18–30 Hz, delta: 1–3 Hz, gamma1: 31–40 Hz, gamma2: 41–50 Hz, theta: 4–7 Hz), attention level and mediation level.¹ Table 3 summarizes all the features extracted, whereas a detailed description of feature extraction is reported in (Semeraro et al. 2018).

3.4.2 Feature selection

The computed features were then normalized to reduce the inter-subject variability. For each subject, the normalized features (F_n) was computed considering the data coming from the baseline acquisition, thus the values of each feature (F_e) was expressed as a percentage of the variation respect to this baseline value (F_b) (Eq. 1).

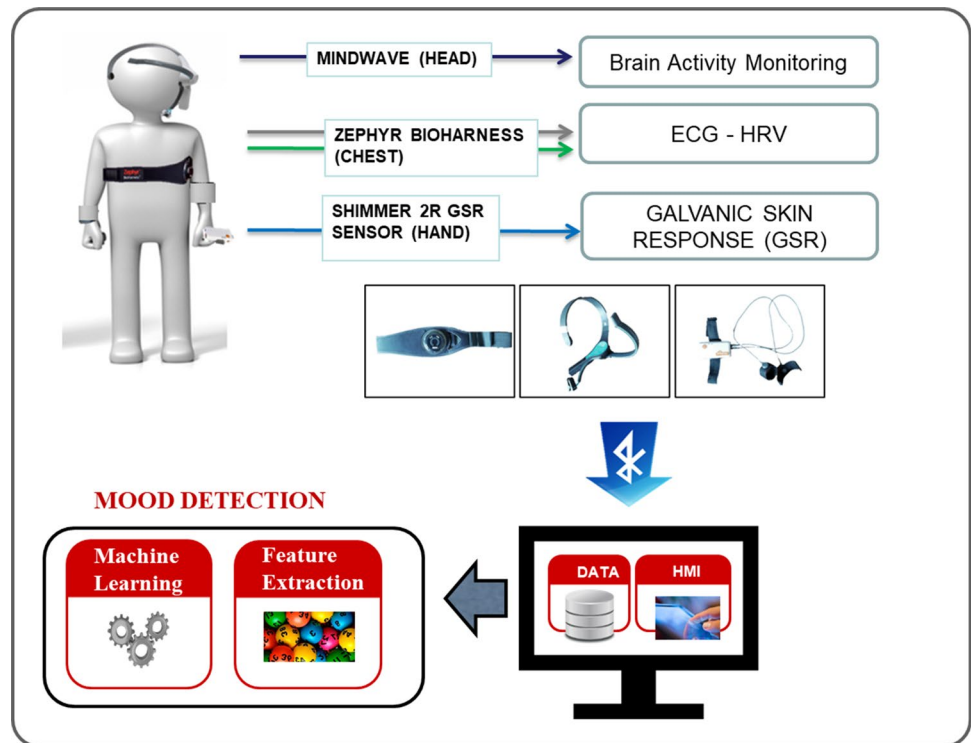
$$F_n = \frac{(F_e - F_b)}{F_b} \% \quad (1)$$

Then the dataset was manually labelled. The relaxed instances corresponding to the ones acquired during the relaxed phase. For what concern the elicited features, we consider only the features starting from the time-windows during which subjects declared to have felt the emotional state as reported in the SAM. The previous instances have been discarded because they do not carry information regarding the mood that we want to classify and are not in any way described by the SAM compiled by the user.

For assigning the label, we the Russels et al. dimensions: the (a) “Valence”, which means the intrinsic attractiveness,

¹ MindWave User Guide, Available: http://developer.neurosky.com/docs/lib/exe/fetch.php?media=mindwave_user_guide_en.pdf, p.14.

Fig. 1 System architecture. The Bioharness, Shimmer and Mindwave devices were connected through the Bluetooth to a computer. A customized interface managed data acquisition and storage. After the experimentation session, the sensor data were retrieved from the computer, analysed and classified using machine learning techniques



goodness or averseness, badness of an event, object or situation and the (b) “Arousal” which means the state of being awoken or of sense organs stimulated with respect to the point of perception (Frijda 1986). According to these domains, in this work, we assigned to these elicited phases two sets of labels.

The first set of label (“3 classes” option) assign the elicited features to the positive or negative labels according to valence scale. The valence scale between 0 and -4 were mapped to “negative” (low valence—LV) and between 0 and 4 were mapped as “positive” respectively (high valence—HV). Whereas in the second method we assigned the label according to the levels of the valence and the arousal assigned through SAM ratings as proposed in (Subramanian et al. 2018). The purpose of this second approach is to evaluate whether our proposed system is able to distinguish considering two different dimensions of the emotional paradigm. The valence scale was mapped as in the first method. For what concern arousal scale, values between 0 and -4 were mapped as “passive” and values between 0 and 4 as “active”, respectively. Indeed, the dataset was divided into four classes: (1) High Valence/High Arousal (HVHA), which includes positive emotions such as happy and excited. (2) High Valence/Low Arousal (HVLA), which includes emotions such as relaxed, calm and pleased. (3) Low Valence/Low Arousal (LVLA), which includes negative emotions such as sad and depressed. (4) Low Valence/High Arousal (LVHA), which includes emotions such as anger, fear and distressed. In the text, we refer to this set of label as “4 classes” method.

3.4.3 Feature reduction and machine learning analysis

Shapiro–Wilk test of normality was applied to each feature distribution. Since all features were not normally distributed, Kruskal–Wallis test was then applied to evaluate which features could be more useful in distinguishing between relaxed, positive and negative moods (HV and LV respectively). If $p > 0.05$, this specific feature was discarded from the analysis. Then, the Spearman correlation coefficient was computed. Those features with correlation coefficient higher than 0.85 (absolute values) and $p < 0.05$ were carefully discarded (Kao et al. 2015).

The reduced dataset was then classified with the Machine Learning Toolbox of Matlab 2018a. Particularly, Support Vector Machine (SVM), Decision Tree (Tree) and K-nearest Neighbors (KNN) were applied with different kernels to the dataset. Particularly, we considered the quadratic, the cubic and the medium Gaussian kernels for the SVM, for what concern the other parameters, in this work the preset selection of the Matlab toolbox was chosen. This toolbox is able to support the multiclass SVM, thus it finds the best hyperplane that separates all data points of one class from those of other classes. As concern the Tree we used the Gini’s diversity index and maximum number of split equal to 20. To train a k-nearest neighbours model, the Classification Learner app of Matlab was used. Particularly, the following specifications were used: we consider $k = 1$ and the Euclidean distance as distance metric for the fine model; $k = 10$ and Minkowsky distance metric

Table 3 Features from HRV, GSR and EEG signals in time (t) and frequency (f) domains

Feature name	Signal	Description
RR mean	HRV(t)	Mean of R-to-R inter-beat intervals belonging to the same time window
SDNN	HRV(t)	Standard deviation of normal RR intervals (also said as NN intervals)
HR mean	HRV(t)	Mean of heart rate
SD mean	HRV(t)	Heart rate standard deviation
RMSSD	HRV(t)	Square root of mean of squared differences between adjacent NN intervals
pNN50	HRV(t)	Percentage of differences between adjacent NN intervals > of 50 ms
VLF peak	HRV(f)	Frequency peak of heart activity VLF (0–0.04 Hz)
VLF power	HRV(f)	PSD area in VLF
%VLF	HRV(f)	Percentage ratio between PSD area in VLF and total one
LF peak	HRV(f)	Frequency peak of LF (0.04–0.15 Hz)
LF power	HRV(f)	PSD area in LF
%LF	HRV(f)	Percentage ratio between PSD area in LF and total one
HF peak	HRV(f)	Frequency peak of HF (0.15–0.40 Hz)
HF power	HRV(f)	PSD area in HF
%HF	HRV(f)	Percentage ratio between PSD area in HF and total one
LF/HF	HRV(f)	Ratio between LF and HF powers
# Startle	GSR(t)	Number of detected startles
Amplitude mean	GSR(t)	Mean value of startles peak amplitude (μS)
Amplitude std	GSR(t)	Standard deviation of startles peak amplitude (μS)
Sum rise time	GSR(t)	Sum of all detected startles rise time duration within the phasic signal portion analysed (s)
Sum fall time	GSR(t)	Sum of all detected startles fall time duration within the phasic signal portion analysed (s)
Rise rate mean	GSR(t)	Mean value of a startle rise time (s)
Rise rate std	GSR(t)	Standard deviation of startle rise time (s)
Decay rate mean	GSR(t)	Mean of a startle fall time (s)
Decay rate std	GSR(t)	Standard deviation of a startle fall time (s)
Phasic value mean	GSR(t)	Mean value of the phasic signal (μS)
Phasic value std	GSR(t)	Standard deviation of the phasic signal (μS)
Startle time mean	GSR(t)	Mean value of a startle duration (s)
Startle time std	GSR(t)	Standard deviation of a startle duration (s)
Startle RMS mean	GSR(t)	Mean Value of the Root Mean Square of the curve identifying a startle (μS)
Startle RMS std	GSR(t)	Standard deviation of the Root Mean Square of the curve identifying a startle (μS)
Startle RMS overall	GSR(t)	Value of the Root Mean Square of the whole phasic signal portion analysed (μS)
Alpha1	BA(f)	Signal power in frequency range 8–9 Hz
Alpha2	BA(f)	Signal power in frequency range 10–12 Hz
Beta1	BA(f)	Signal power in frequency range 13–17 Hz
Beta2	BA(f)	Signal power in frequency range 18–30 Hz
Delta	BA(f)	Signal power in frequency range 1–3 Hz
Gamma1	BA(f)	Signal power in frequency range 31–40 Hz
Gamma2	BA(f)	Signal power in frequency range 41–50 Hz
Theta	BA(f)	Signal power in frequency range 4–7 Hz
Attention	EEG(f)	NeuroSky index for user's level of mental "focus" or "concentration"
Meditation	EEG(f)	NeuroSky index for user's level of mental "calmness" and "relaxation"

for the cubic model and $k=10$ and Euclidean distance metric for the weighted KNN.

A 5-Cross Fold validation technique was used to train and test the classifiers. The results were organized in confusion matrix and accuracy, precision, recall, specificity and

F-Measure were then computed for each classes to estimate the overall performance of the classifiers (Lara and Labrador 2013):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F - \text{Measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where the True positive (TP) is the number of correct classification of positive instances and the True negative (TN) is the number of correct classification of negative instance. False Positives (FP) and False Negatives (FN) are the numbers of negative instances classified as positive and the number of positive instances classified as negative respectively.

All the analysis were computed on a windows machine with a 64-bit operative system and ×64-based processor [Intel® Core™ i7-6700HQ CPU@2.60 Ghz] with an installed RAM of 12 GB. The requested computational time was also recorded.

4 Results

According to the results, nobody had a severe score in the BDI questionnaire and nobody had a positive response to the MOCQ-R questionnaire. Thus all the recruited participants were included in the analysis. In the end, 17 subjects have been tested in the negative condition, while 17 subjects have been tested in the positive one.

A total of 239 instances were extracted from the acquired signals. Particularly, 72 instances are related to the relax phase. For what concern the other classes, according to the first method in which we consider only the valence dimension, 89 instances are related to the positive mood (HV) and 78 to the negative mood (LV). As concern the second method, 77 instances belongs to the HVHA groups, 12 instances to the HVLA groups and 78 instances to the LVHA groups. According to the SAM responses, the LVLA subgroup is empty (see Fig. 2).

Starting from a set of 42 features the results of the statistical analysis (Spearman Correlation Coefficient and Kruskal–Wallis) led to a subset of 16 features, 7 from ECG (IBI mean, SDHR, RMSSD, HF peak, %VLF percentage, %LF percentage, %HF percentage), 4 from GSR (# startle, Amplitude mean, Rise time std, Fall time std) and 5 from BA (Alpha1, Beta2, Gamma1, Gamma2, Delta).

Table 4 reports the features extracted from HRV, GSR and BA signals as mean values. At the end of each row,

the p-values, calculated with Kruskal–Wallis test for non-parametric data, are also shown. A *p* values index lower than 0.05 provided information on the presence of potentially significant differences between the investigated conditions. These values were calculated using the features extracted from signals recorded during the different phases of the protocol (relax, positive and negative). The result confirms that all the selected features were significantly different in distinguish among the three states, indeed they were included in the final analysis. Table 5 reports the computational time requested for each classifier. The fastest classifier is the KNN in both cases, among the KNN the fastest is the weighted KNN.

This paper aims to investigate whether is possible to classify different moods according to valence and to arousal dimension. As concern the “3 classes” analysis, among all the selected classifiers, the fine KNN classifier reaches the best overall performance in terms of accuracy and F-Measure (0.866 and 0.865 respectively), whereas the cubic KNN reaches the worst ones with an accuracy equal to 0.707 and F-measure equal to 0.699. All the other classifiers reach accuracy higher than 0.82, whereas the specificity value is higher than 0.90 for each classifier except the cubic KNN. The results of the classification into relax, HV (positive) and LV (negative) instances are summarized in Table 6.

As concern the “4 classes” analysis, similarly and the “3 classes” analysis, the fine KNN classifier reaches the best performance in terms of accuracy and F-measure (0.877 and 0.873 respectively). The worst classifier is the cubic KNN, which reports an accuracy value equal to 0.655. For what concern all the remaining classifiers, the accuracy is higher than 0.750, it is worth od mention that the SVM with cubic and quadratic kernel report similar performances. Table 7 reports the complete results for all classifiers.

Figure 3 compares the performance of the selected classifiers in terms of accuracy. From visual inspection, it is possible to notice that the two approaches reach comparable performances. Additionally, Cubic SVM, Quadratic SVM and Fine KNN classifiers reach higher performance with the “4 classes” approach whereas the remaining classifiers reach higher performance with the “3 classes” method. Since the fine KNN classifier reaches the best performance with both approaches, in the remaining part of this section, more details on its performance are reported. Figure 4 reports and compare the F-Measure values obtained in the proposed analysis with the fine KNN classifiers. Particularly, in the green box is reported the performance for the positive instances (high valence) and in the yellow box for the negative instances (low valence). Whereas, Fig. 5 depicts the confusion matrixes. It is worth to underline that the relax and LVHA (equal to HV group in our case) reach slight higher F-measure values with the “4 classes” approach rather than the other one (“3 classes”: Relax = 0.851, LV = 0.863; “4

Fig. 2 Result of the SAM assessment in the valence/arousal space performed at the end of the experimental session. The red dots represent the subjects which perform the positive test, whereas the blue dots represent the subjects which perform the negative test. The black dot represents the relax status, whereas the blue boxes represent the “3 classes” label assignment (i.e., positive—HV and negative—LV in addition to the relax phase) and the yellow boxes represent the “4 classes” label assignment (i.e., HVHA, HVLA, LVHA in addition to the relax phase)

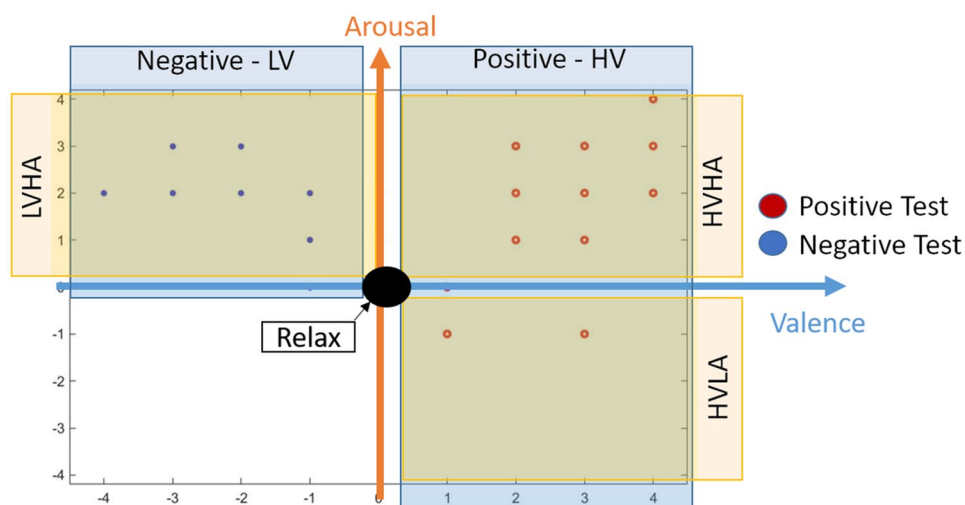


Table 4 Normalised mean values of the three states (Relax, Positive/HV and Negative/LV) and significance in distinguishing the three moods (p value) of the selected features

	Relax	HV	LV	p value
IBI mean	-0.0093	-0.14	-0.13	2.80e-23
SDHR	0.033	0.42	0.45	3.33e-08
RMSSD	-0.034	-0.20	-0.21	7.04e-06
HF peak	0.0043	-0.21	-0.20	7.26e-09
VLF percentage	0.037	0.59	0.52	2.58e-05
LF percentage	0.062	0.58	0.46	5.85e-05
HF percentage	-0.016	-0.15	-0.11	0.00062
# startle	0.43	0.92	2.86	1.58e-09
Amplitude mean	0.72	5.41	6.096	1.02e-11
Rise time std	0.17	0.35	2.95	0.042
Fall time mean	-0.048	0.29	0.013	4.32e-15
Alpha1	0.086	0.56	0.28	4.16e-08
Beta2	0.060	0.96	0.54	7.21e-15
Gamma1	0.27	1.85	1.31	1.76e-15
Gamma2	-0.13	1.01	0.052	1.64e-06
Delta	0.084	1.174	0.734	1.05e-10

classes”: Relax = 0.895, LVHA = 0.877). Conversely, the F-measure of the HVHA and HVLA groups are slightly lower than the HV group (HV = 0.880, HVHA = 0.863, HVLA = 0.857).

5 Discussion

The aim of this paper was to investigate whether the proposed system was able to recognize different moods according to the valence dimension (3 classes method) and to a combination of arousal and valence dimensions (4 classes method). The proposed test aims at the elicitation of positive

Table 5 The requested computational time expressed in s

	Computational time [s] “3 classes”	Computational time [s] “4 classes”
Cubic SVM	4.545	4.058
Quadratic SVM	4.627	4.146
Medium Gaussian SVM	2.537	0.429
Medium Tree	3.247	2.710
Fine KNN	1.381	1.453
Cubic KNN	0.982	1.08
Weighted KNN	0.907	0.429

and negative moods (valence) through dyadic interactions in healthy subjects. According to the SAM scores, the tests achieve the goal, since the methods it is able to elicit the desiderated moods. Indeed, as depicted in Fig. 2, the subjects who performed the positive elicitation reports the positive value of valence and the ones who performed the negative test reported negative values of valence (red dots and blue dots respectively). It is worth to underline that, as concern the arousal dimension, the negative test is able to evoke only high value of arousal (score values between 0 and 4), whereas the positive test is able to evoke both high and low level of arousal (between -4 and 0).

In this paper, we present the performance of seven different supervised classifiers, and we obtain high performance with both sets of labels (see Fig. 3). The best classifier is the fine KNN which represent a good trade-off between accuracy and the requested computational time. Indeed, fine KNN classifier reaches an accuracy equal to 0.866 and 0.877 with the “3 classes” and “4 classes” approaches respectively. These results are aligned with the results of our previous preliminary analysis that we obtained from a reduced dataset. Indeed, we obtained an accuracy equal to

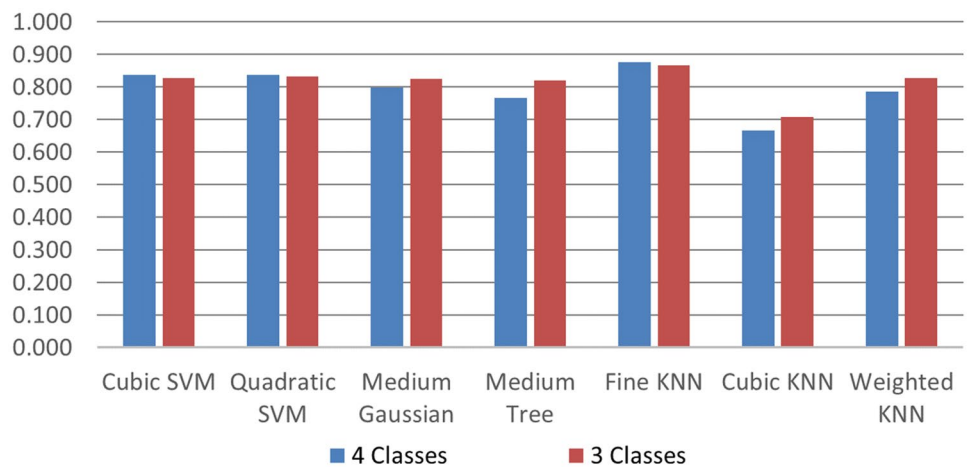
Table 6 Overall performance of the classifiers in distinguish the relax, HV and LV status (“3 classes” set of labels)

	Accuracy	F-measure	Precision	Recall	Specificity
Cubic SVM	0.828	0.805	0.830	0.829	0.907
Quadratic SVM	0.833	0.840	0.832	0.835	0.909
Medium Gaussian SVM	0.824	0.827	0.827	0.826	0.905
Medium Tree	0.821	0.861	0.822	0.820	0.901
Fine KNN	0.866	0.865	0.867	0.863	0.927
Cubic KNN	0.707	0.699	0.732	0.711	0.828
Weighted KNN	0.828	0.827	0.838	0.829	0.906

Table 7 Overall performance of the classifiers in distinguishes among the relax, HVHA, HVLA, and LVHA (“4 classes” set of labels)

	Accuracy	F-measure	Precision	Recall	Specificity
Cubic SVM	0.837	0.854	0.854	0.854	0.867
Quadratic SVM	0.837	0.839	0.827	0.855	0.864
Medium Gaussian SVM	0.799	0.786	0.849	0.754	0.855
Medium Tree	0.766	0.701	0.713	0.692	0.843
Fine KNN	0.877	0.873	0.854	0.904	0.884
Cubic KNN	0.665	0.511	0.517	0.529	0.805
Weighted KNN	0.787	0.768	0.813	0.746	0.857

Fig. 3 Comparison of the accuracy among the selected classifiers of the two set of labels, the three classes (Relax, HV, LV) and the four classes (Relax, HVHA, HVLA, LVHA)



89.67% for the SVM classifier with the polynomial kernel (Semeraro et al. 2018). As concern the comparison with literature (see Table 1), our proposed method reaches comparable performances with other works which use physiological signals but different ways to evoke emotions (i.e., music, videos, pictures). Recently, Al Machoth et al. (Al Machot et al. 2018a, b) and Subramanian et al. (2018) analyzing physiological features obtained lower performances compared with the ones obtained with our system. Albanie et al. (2018) obtained an average and comparable accuracy equal to 85% by using camera-based strategies.

As depicted in Fig. 4, the arousal domain slightly increases the recognition performance of the relax phase and the LV classes, whereas slightly decrease the performance of the positive instances. The highest accuracy performance

was achieved by the fine KNN in classifying the 4 classes. This could be explained by the nature of the signals recorded by the worn sensors. The Zephyr BioHarness and the Shimmer GSR Module record, in fact, measured the variations linked to the functioning of the autonomic nervous system. As well known modifications at this level are more related to the arousal than valence (Heller 1993).

In this work, we extract the features over a time-windows of 180 s to be sure to capture the variation of the physiological signals. Since we obtain high performance of the evaluation metrics, we could suppose that this time-window is appropriate to capture the variation of the physiological parameters.

Our approach proposed the dyadic interaction to evoke moods in the subject, whereas other literature works use other methods (i.e., pictures, video, sounds) that are far from

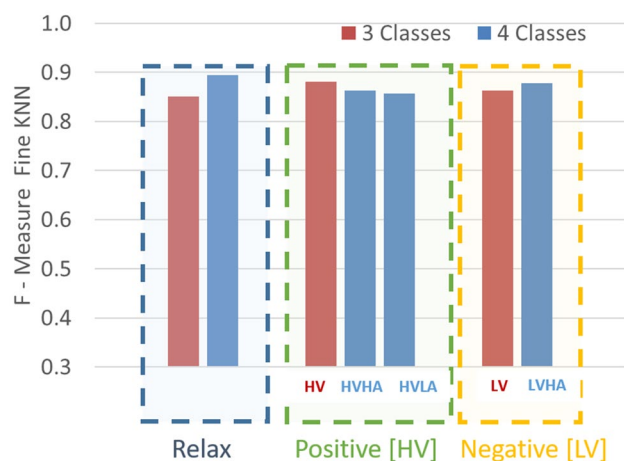


Fig. 4 F-measure values of the fine KNN classifiers over the classes: Relax, Positive (HV—HVHA, HVLA) and negative (LV—LVHA). The red bars refer to the “3 classes” model and the blue bar to the “4 classes” model

realistic applications where it is important to capture what happened during social interaction. It is worth to mention that physiological measures could be affected by high inter-subject variability during an elicited phase, consequently, there are high fluctuations of physiological features which can affect the performances of the classifiers. Additionally, the subjects were set free to move during the tests. Other literature works required that the subject perform reduced movements during the data acquisition to reduce the noise. In our work, subjects were free to move their body, and handle the interaction at their own willing, as long as they stayed sit down. Also, this aspect could introduce disturbances in acquiring the physiological sequences.

Additionally, the obtained results suggest that physiological signals could be used to monitor the emotional status of the subject as an alternative way to the most commonly used video-based analysis (Khan et al. 2013; Turabzadeh et al. 2018; Dong et al. 2018). Indeed, wearable sensors

solve part of the drawbacks of the camera system related to privacy, occlusion of the camera and light resource (Cavallo et al. 2018b). In fact, to correctly analyse video and achieve valuable performance, the subject should stay in the front of the camera with the correct brightness of the environment, which is not always feasible in real settings. Physiological signals do not require these constraints.

Furthermore, for what concern the wearability of the sensors for prolonged time, for instance in case you would like to monitor the emotional status of the user over the 24 h because of clinical reasons, ring-shaped devices have the potential to acquire physiological signals which can be monitored to investigate subject’ emotional status (Fiorini et al. 2019). On the other side, the performance of the system in terms of requested computational time could be enhanced exploiting cloud computing resources. In this sense, further research should be planned.

Further analysis and experimentation should overcome the limitations of this work which are mainly related to the dimension of the dataset, the off-line analysis and the limited mood evoked. Indeed further experimentation sessions should be planned to increase the number of subjects involved in the study trying to elicited and an higher number of moods/emotions. Additionally, in this study, we performed an off-line analysis, thus further development of the work should lead towards the implementation of a real-time system which can be used in affective computing applications such as the development of robot behavioural models and the monitoring of the health status since some cognitive disorders can be correlated with a variation of the moods.

6 Conclusion

The paper presents a wearable system able to identify different moods (i.e., relax, positive and negative). A total of seven supervised machine learning algorithms were applied to classify the physiological features over the valence and

Fig. 5 Confusion matrix of the fine KNN classifiers which result to be the best classifier. On the left, the confusion matrix obtained using the “3 classes” labels (i.e., relax, positive/HV, negative/LV). On the right, the confusion matrix obtained considering also the arousal domain as output in addition to the valence (i.e., relax, HVHA, HVLA, LVHA)

Relax	60	8	4
HV	3	81	5
LV	6	6	66
	Relax	HV	LV

Relax	62	2	8	
HVHA	6	57	14	
HVLA	1	3	7	
LVHA	8	5	65	
	Relax	HVHA	HVLA	LVHA

arousal dimensions. In the proposed approaches, the moods were elicited through social interaction, instead of using videos or pictures as proposed by related literature works. The results of the self-assessment questionnaire underlined that our proposed approach is able to elicit the desiderated moods. The final goal of our paper was to investigate if our system is able to recognize the selected moods, the achieved results were promising since the best classifier (fine KNN) reach an accuracy equal to 0.877 suggesting that it is able to classify the features along the valence and the arousal dimension.

Acknowledgements This work is financially supported by the CloudIA project (Regione Toscana POR CREO FESR 2014-2020, CUP 165.24052017.112000015) and ACCRA Project, founded by the European Community's Horizon 2020 Programme (H2020-SCI-PM14-2016)—Grant agreement no. 738251.

References

- Agrigoroaie R, Tapus A (2017) Contactless physiological data analysis for user quality of life improving by using a humanoid social robot. *Lect Notes Comput Sci* 10485:696–706. https://doi.org/10.1007/978-3-319-68548-9_63
- Al Machot F, Ali M, Ranasinghe S, Mosa AH, Kyandoghere K (2018a) Improving subject-independent human emotion recognition using electrodermal activity sensors for active and assisted living. *Nurs Homes* 29:38–39. <https://doi.org/10.1145/3197768.3201523>
- Al Machot F, Ali M, Ranasinghe S et al (2018b) Improving subject-independent human emotion recognition using electrodermal activity sensors for active and assisted living. In: *Proceedings of the 11th PErvasive technologies related to assistive environments conference*. ACM, pp 222–228
- Albanie S, Nagrani A, Vedaldi A, Zisserman A (2018) Emotion recognition in speech using cross-modal transfer in the wild. In: *Proceeding MM '18 Proceedings of the 26th ACM international conference on Multimedia*. ACM, New York, pp 292–301
- Beck AT, Ward CH, Mendelson M et al (1961) An inventory for measuring depression. *Arch Gen Psychiatry* 4:561–571
- Betti S, Molino Lova R, Rovini E et al (2017) Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers. *IEEE Trans Biomed Eng*. <https://doi.org/10.1109/TBME.2017.2764507>
- Boucsein W (2012) *Electrodermal activity*. Springer, Berlin
- Bradley M, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* 25:49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Cavallo F, Esposito R, Limosani R et al (2018a) Acceptance of Robot-Era system: results of robotic services in smart environments with older adults (Preprint). *J Med Internet Res*. <https://doi.org/10.2196/jmir.9460>
- Cavallo F, Semeraro F, Fiorini L et al (2018b) Emotion modelling for social robotics applications: a review. *J Bionic Eng*. <https://doi.org/10.1007/s42235-018-0015-y>
- Chen M, Ma Y, Song J et al (2016) Smart clothing: connecting human with clouds and big data for sustainable health monitoring. *Mob Netw Appl* 21:825–845. <https://doi.org/10.1007/s11036-016-0745-1>
- Chen J, Hu B, Wang Y et al (2017) Subject-independent emotion recognition based on physiological signals: a three-stage decision method. *BMC Med Inform Decis Mak*. <https://doi.org/10.1186/s12911-017-0562-x>
- Dağlarlı E, Dağlarlı SF, Günel GÖ, Köse H (2017) Improving human-robot interaction based on joint attention. *Appl Intell* 47:62–82. <https://doi.org/10.1007/s10489-016-0876-x>
- Damasio AR (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond Ser B Biol Sci* 351:1413–1420. <https://doi.org/10.1098/rstb.1996.0125>
- Dong Y, Hu H, Wen Y et al (2018) Personalized emotion-aware video streaming for the elderly. Springer, Cham, pp 372–382
- Doux J (1996) Emotional networks and motor control: a fearful view. *Prog Brain Res* 107:437–446
- Ekman P (1999) Basic emotions. In: Dalglish T, Power M (eds) *Handbook of cognitive emotion*. Wiley, Sussex, pp 45–60
- Ferrer RA, Mendes WB (2018) Emotion, health decision making, and health behaviour. *Psychol Health* 33:1–16. <https://doi.org/10.1080/08870446.2017.1385787>
- Fiorini L, Semeraro F, Mancioffi G et al (2018) Physiological sensor system for the detection of human moods towards internet of robotic things applications. *Front Artif Intell Appl* 303:967–980. <https://doi.org/10.3233/978-1-61499-900-3-967>
- Fiorini L, Cavallo F, Martinelli M, Rovini E (2019) Characterization of a PPG wearable sensor to be embedded into an innovative ring-shaped device for healthcare monitoring. In: *X Forum Italiano di Ambient Assisted Living (foritAAL)*. Ancona, pp 1–17
- Frijda NH (1986) *The emotions*. Cambridge University Press
- Gallagher D, Kiss A, Lancot K, Herrmann N (2018) Depression and risk of Alzheimer dementia: a longitudinal analysis to determine predictors of increased risk among older adults with depression. *Am J Geriatr Psychiatry* 26:819–827. <https://doi.org/10.1016/j.jagp.2018.05.002>
- Glende S, Conrad I, Krezdorn L et al (2016) Increasing the acceptance of assistive robots for older people through marketing strategies based on stakeholder needs. *Int J Soc Robot* 8:355–369. <https://doi.org/10.1007/s12369-015-0328-5>
- Harmon-Jones EM, Amodio DR et al (2007) Social psychological methods of emotion elicitation. *Handb Emot Elicitation Assess*. <https://doi.org/10.2224/sbp.2007.35.7.863>
- Heller W (1993) Neuropsychological mechanisms of individual differences in emotion, personality, and arousal. *Neuropsychology* 7:476–489. <https://doi.org/10.1037/0894-4105.7.4.476>
- Hodgson RJ, Rachman S (1977) Obsessional-compulsive complaints. *Behav Res Ther* 15:389–395. [https://doi.org/10.1016/0005-7967\(77\)90042-0](https://doi.org/10.1016/0005-7967(77)90042-0)
- Horstmann AC, Krämer NC (2019) Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Front Psychol* 10:939. <https://doi.org/10.3389/fpsyg.2019.00939>
- Hu L, Miao Y, Wu G et al (2019) iRobot-Factory: an intelligent robot factory based on cognitive manufacturing and edge computing. *Future Gener Comput Syst* 90:569–577. <https://doi.org/10.1016/J.FUTUR E.2018.08.006>
- Kao F-C, Wang SR, Chang Y-J (2015) Brainwaves analysis of positive and negative emotions. *WSEAS Trans Inf Sci Appl* 12:200–208
- Karami AB, Sehaba K, Encelle B (2016) Adaptive artificial companions learning from users' feedback. *Adapt Behav* 24:69–86. <https://doi.org/10.1177/10597123166634062>
- Khan RA, Meyer A, Konik H, Bouakaz S (2013) Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognit Lett*. <https://doi.org/10.1016/j.patrec.2013.03.022>
- Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15:1192–1209. <https://doi.org/10.1109/SURV.2012.110112.00192>
- Lu Y (2017) Industry 4.0: a survey on technologies, applications and open research issues. *J Ind Inf Integr* 6:1–10. <https://doi.org/10.1016/J.JII.2017.04.005>

- Mehmood RM, Du R, Lee HJ (2017) Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors. *IEEE Access* 5:14797–14806. <https://doi.org/10.1109/ACCESS.2017.2724555>
- Mali B, Zulj S, Magjarevic R et al (2014) Matlab-based tool for ECG and HRV analysis. *Biomed Signal Process Control* 10:108–116. <https://doi.org/10.1016/j.bspc.2014.01.011>
- Maurtua I, Fernández I, Tellaache A et al (2017) Natural multimodal communication for human–robot collaboration. *Int J Adv Robot Syst* 14:172988141771604. <https://doi.org/10.1177/1729881417716043>
- McGaugh JL (2018) Emotional arousal regulation of memory consolidation. *Curr Opin Behav Sci* 19:55–60. <https://doi.org/10.1016/j.COBEHA.2017.10.003>
- Mehrabian A (1996) Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in Temperament. *Curr Psychol* 14:261–292. <https://doi.org/10.1007/BF02686918>
- Moaiyed V, Firoozabadi M, Khezri M (2017) Recognition of music-induced emotions based on heart-brain connectivity. In: 24th National and 2nd International Iranian Conference on Biomedical Engineering, pp 330–333
- Nocentini O, Fiorini L, Acerbi G et al (2019) A survey of behavioural models for social robots. *Robotics*. <https://doi.org/10.20944/preprints201905.0251.v1>
- Pahl C, Varadarajan KM (2015) Emotional feedback for service robots using haptic-acoustic interface. In: 2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), pp 219–224. <https://doi.org/10.1109/robomech.2015.7359526>
- Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective/physiological state. *IEEE Trans Pattern Anal Mach Intell* 23:1175–1191. <https://doi.org/10.1109/34.954607>
- Rodić A, Urukalo D, Vujović M et al (2017) Embodiment of human personality with EI-robots by mapping behaviour traits from live-model. Springer, Cham, pp 438–448
- Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39:1161–1178. <https://doi.org/10.1037/h0077714>
- Sanavio E, Bertolotti G, Michielin P, Vidotto G, Zotti AM (1986) CBA-2.0 Scale Primarie: Manuale. Una batteria a largo spettro per l'assessment psicologico. Organizzazioni Speciali, Firenze. Seconda edizione ed ampliata, 1997
- Semeraro F, Fiorini L, Betti S et al (2018) Physiological wireless sensor network for the detection of human moods to enhance human-robot interaction. Springer, Cham
- Sharma N, Gedeon T (2012) Objective measures, sensors and computational techniques for stress recognition and classification: a survey. *Comput Methods Programs Biomed* 108:1287–1301. <https://doi.org/10.1016/j.cmpb.2012.07.003>
- Sheikholeslami S, Aj Moon, Croft EA (2017) Cooperative gestures for industry: exploring the efficacy of robot hand configurations in expression of instructional gestures for human–robot interaction. *Int J Rob Res* 36:699–720. <https://doi.org/10.1177/0278364917709941>
- Subramanian R, Wache J, Abadi MK et al (2018) Ascertain: emotion and personality recognition using commercial sensors. *IEEE Trans Affect Comput* 9:147–160. <https://doi.org/10.1109/TAFFC.2016.2625250>
- Turabzadeh S, Meng H, Swash R et al (2018) Facial expression emotion detection for real-time embedded systems. *Technologies* 6:17. <https://doi.org/10.3390/technologies6010017>
- Turchetti G, Micera S, Cavallo F et al (2011) Technology and innovative services. *IEEE Pulse* 2:27–35. <https://doi.org/10.1109/MPUL.2011.940428>
- van den Hout MA, De Jong P, Kindt M (2000) Masked fear words produce increased SCRs: an anomaly for Ohman's theory of pre-attentive processing in anxiety. *Psychophysiology* 37:283–288
- Zhuang N, Zeng Y, Tong L et al (2017) Emotion recognition from EEG signals using multidimensional information in EMD domain. *Biomed Res Int* 2017:1–9. <https://doi.org/10.1155/2017/8317357>

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