Physiological sensor system for the detection of human moods towards internet of robotic things applications

Physiological sensor system for the detection of human moods towards internet of robotic things applications

Laura FIORINI[a,1] Francesco SEMERARO[a], Gianmaria MANCIOPPI[a], Stefano BETTI[a], Luca SANTARELLI[a] and Filippo CAVALLO[a]
[a] *The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa - Italy*

Abstract. Internet of Robotic Things paradigm offers a concrete support to daily life. The pervasiveness of smart things, together with advances in cloud robotics, can help the smart systems to perceive and collect more information about the users and the environment. Often citizens have experienced "one-size-fits-all" approach, since the delivered service was not personalized, therefore resulting far from user's expectations. Hence, future smart agents, like robots, should produce personalized behaviours based on user emotions and moods in order to be more integrated into ordinary activities. In this work, we investigated the performances with unsupervised and supervised approaches to recognize three different moods elicited during a social interaction by means of a wearable system capable of measuring the Electrocardiogram, the ElectroDermal Activity and the Electroencephalographic signals. Particularly, the classification problem was analysed using three unsupervised (K-Mean, Self-Organizing Map and Hierarchical Clustering) and three supervised methods (Support Vector Machine, Decision Tree and k-nearest neighbour). The supervised algorithms reached an accuracy of 0.86 in the best case. The outcomes show that even in an unsupervised context the system is able to recognize the mood, reaching an accuracy equal to 0.76 in the best case.

Keywords. Mood Recognition, Wearable Physiological Sensors, Unsupervised Machine Learning

1. Introduction

Recently, literature findings [1] have demonstrated that service robotics and Information and Communication Technology (ICT) can represent a valid support for population. The integration of Robotics, Internet of Things and Artificial Intelligence, i.e. the "*Internet of Robotic Things*", enables the possibility to design and develop new frontiers in human-robot interaction and, more in general, in human-system interaction [2]. Unfortunately, in most of the cases end-users are left with a "one-size-fits-all" approach, which often leads to a frustrating and overall negative experience for the user himself because of the inadequacy of these devices, not specifically designed for such applications. Therefore, to exploit contemporary psychological theories of health-related behaviour changes and users-adaptation properly, a new more specific approach

[1] Corresponding Author, Laura Fiorini, The BioRobotics Institute, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio 34, 56025, Pontedera (Pisa) - Italy; E-mail: laura.fiorini@santannapisa.it.

design is needed.

According to literature evidences [3], emotions affect decisional processes in humans beings. Consequently, it is important for intelligent machines to embed emotion modelling to enhance and improve human-machine interaction.

Recently researches have been conducted in the field of emotion recognition [4]. Latest advancements achieved in this field are based on algorithms capable to discern among different types of emotions by inspecting different sensing modalities, e.g. vision or physiological activity. However, most of these works were not conceived for an application in daily-life situations, as they are mainly focused on the detection of costumers' satisfaction [5], or to investigate the interaction with children affected by autism [6]. In order to exploit emotion recognition for public utility service is fundamental to investigate measurement methodologies within different scenarios [7] in which emotions take place.

2. Related Work

The most famous theory regarding the representation of emotion was proposed by Ekman and Friesen [8]. Every emotive expression can be classified into one of six main categories: anger, disgust, fear, happiness, sadness, and surprise, plus a neutral one. Later, he added other emotional states (fun, contempt, contentment, embarrassment, excitement, guilt, pride, relief, satisfaction, pleasure, shame), called secondary emotions. The other school of thought is related to Russell's theory [9]. He asserted that emotions can be retrieved in empirical sections of a plane of two main coordinates: valence, which gives positive or negative connotation to the emotional state, and arousal, which measures the intensity of the feeling. However, the use of categorical emotions in the classification process is under revision since such approach has several limitations. It can create mistakes in interactional cases in which very different features represent the same emotion [10].

Most affective computing applications primarily use vision [4] as informative channel. Detection of facial expressions, in fact, is the most natural way to perceive emotions of people. However, it has recently become evident that the typical use of vision in affective computing applications may not be as feasible as in social robotics. In fact, the standard procedure to acquire information through vision is to make a person sit in front of a camera with properly placed lights [11]. In a real scenario, the person for most of the interaction period likely would not be standing in front of the robot at a fixed distance. Therefore, emotion perception must be achieved by exploiting different informative channels not directly controlled by humans, like physiological parameters. Many applications of affective computing have employed these signals (See Table 1). Nardelli et al. [12] used electrocardiogram (ECG) signals to classify emotions in subjects by means of auditory stimuli. Rattanyu et al. [13] used the same modality with the declared intent to subsequently embed this process into service robots. Khezri et al. [14] performed a detailed analysis by examining EEG, electrooculography (EOG), electromyography (EMG), blood volume pressure (BVP), galvanic skin response (GSR), and heart rate variability (HRV) simultaneously in subjects . Koelstra and Patras [11] included visual features together with EEG recordings to achieve better classification accuracy. In Henriques et al. [15] measurements of electrodermal activity (EDA) were performed within human-robot interaction scenarios

Table 1 In the table the presented work (Fiorini et al.) is compared to the related work about emotion/mood induction techniques (MIP), information channels (EEG, GSR, ECG/BVP, Facial Expression (FE) and Mood Self-Assessment Evaluation (ESE)) and usability of the proposed system in real life scenarios (CO). Our work aims to be innovative through the design of a social interaction (SI) stimulation and the integration of different information channels maintaining a high level of usability and user comfort during daily life activities.

	MIP					**Informative channel**					
Ref	Video	Picture	Sound	PC	SI	EEG	GSR	ECG BVP	FE	ESE	CO
[12]			X					X			X
[13]		X						X			X
[20]	X						X	X			X
[11]	X		X			X			X		
[15]					X		X				
[16]				X						X	
[17]	X		X			X					
Fiorini et al.					X	X	X	X		X	X

According to literature [4], different mood induction procedures (MIP), inherited from affective computing, can be successfully used to induce emotions in human subjects. In most of these cases subjects are asked to sit in front of a computer (PC) [16] and watch video clips [17] or images from International Affective Picture System (IAPS) dataset [18], or listen to music [12]. However, these experiments result unfit for social robotics not suiting the ecology of realistic situation since the social interaction, mainly in the form of a dyadic interaction [19], is not considered.

The proposed work is oriented to recognize, by using a wireless sensor network, the user's emotional state, better said mood, during a MIP involving an active social interaction (SI) as a vehicle for mood induction (See the comparison with similar work reported in Table 1). The first aim of this work is to preliminary investigate how to infer people's mood during social interactions by means of a wearable sensors system able to measure physiological parameters. Particularly, in this work, three emotional states were assessed: positive, negative and relaxed connotations of an experienced situation, without considering a particular emotional model. Statistical analysis were conducted to investigate which set of physiological features are able to distinguish among the moods. Moreover, this work aims also to investigate the use of unsupervised machine learning algorithms to recognize the selected moods. Unsupervised machine learning algorithms, in fact, do not require labelled data and can therefore be more adaptable to real applications of emotion recognition. Particularly, we analyse the performances of three unsupervised machine learning algorithms comparing them with three supervised ones.

3. Material and Methods

In order to achieve all the stated aims of this study, a structured methodology was applied. The first part of this section describes the wearable system and the data acquisition, whereas the second part is related to the data analysis.

3.1. Instruments

Three commercial wearable sensor devices were selected to measure physiological response to the stimuli. The reasonable trade-off between measure accuracy and unobtrusiveness led to the selection of Zephyr BioHarness™3 (Zephyr, Maryland, USA) Shimmer GSR Module (Shimmer, Ireland)and MindWave EEG headset (Neurosky, California USA). Zephyr BioHarness is a Bluetooth chest belt capable of monitoring cardiac activity by recording the electrocardiogram (ECG) signal so as to obtain the Heart Rate and the R-R Intervals. The ECG signal is sampled at 250 Hz. Inter-Beat-Interval data provided by the device has been used for data analysis and features extraction. The Shimmer GSR Module is a small-size, lightweight wearable sensor that streams one channel data related to ElectroDermal Activity (EDA) at sampling rate of 51.2 Hz. The Shimmer module is composed by two special finger electrodes and a main unit that streams data related to the galvanic skin response via Bluetooth connection. The MindWave headset is a Bluetooth device able to capture single frontal lobe channel EEG raw data at a sample rate of 512 Hz. Furthermore, the headset provided the indexes of attention and meditation of the user, related to the frequency power spectrum of the acquired signal, provided with a rate of 1 Hz.

3.2. Experimental Protocol

The proposed experimental protocol had three main different phases designed to put subjects respectively in neutral (baseline), relaxed and elicited (positive or negative) mood through social interaction. At the end of elicited phase a self-assessment survey of the felt mood was administered. Protocol was carried out by an engineer and a psychologist. The psychologist administrated all the phases with skilled social interaction attitudes. The engineer supervised from a technical point of view, particularly on the signal acquisition quality. In the following paragraphs, the phases are described in details.

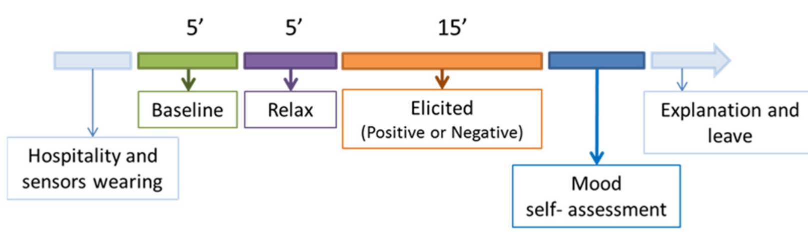

Figure 1. Experimental protocol phases.

Hospitality and sensors wearing - At the beginning, the experimenters gave to participant some explanation about what he/she was going to do, without revealing the final goal of the test, thus maintaining the user’s unawareness until the very end of the experiment [7]. The subject was asked to sign an informed consent form. After this brief introduction, the subject was asked to wear the set of sensors according to the instructions provided by the experimenters.

Baseline phase After the preliminary hospitality phase, the subject was asked to rest on a chair in a room for 5 minutes in absence of any sonorous stimulus.

Relax phase - The experimenters asked the subject to abandon himself in a state of relaxation. For the next 5 minutes the subject remained in a relaxing situation avoiding action or interaction with other object or people.

Elicited phase - It is impossible to administer sequentially both the positive and the negative test to the same subject because the mood achieved through the first interaction could affect the subject's mood during the second one. For this reason, only one of the two conditions has been assigned randomly to each subject, maintaining a balance in terms of percentage between the subjects excited with negative or positive mood induction. Below the different induction strategies used in this phase of the protocol are described:

Positive mood induction: This condition aimed at raising a positive experience in the participants. In this case, the psychologist was immediately very affable with the subject. He explained the tasks rules to the subject: he must answer 55 questions selecting one of the 4 possible answers, among which just one was correct. The questions provided for this particular condition had been prepared so as to result gratifying for the subject regardless him being able to answer or not, making him laugh and without letting him feeling ignorant in case of wrong given answer. There were no time restrictions for the response. Moreover, the experimenters had a kind attitude towards the subject. They never made him know whether a given answer was wrong or not, and each answer was followed by implicit signs of appreciation, like nodding. Furthermore, 5 specific interventions, set to be triggered at fixed moments of the protocol, were used to reinforce the administered condition: “keep going, you got a lot of correct answer”, “very well, you are doing great”, “you hit several of them”, “Thank goodness there are not prizes which may be won, otherwise you would have won that for sure!”, “I hope she/he does not bust our results”.

Negative mood induction: This condition was designed to raise a negative mood experience in the subject. The psychologist entered in the setting during the relaxation phase. He sat in front of the participant carelessly and without introducing himself. As soon as the relax acquisition finished, he unkindly explained to the subject that he was about to undergo a test in which each question had 4 possible answers, among which just one was correct. He required the participant to answer in less time as possible and to stay very focused, as questions and options could be repeated only once. 55 questions regarding general knowledge were purposely prepared by the experimenters , in such a way to induce the subject to feel ignorant in case of wrong answer. During the task execution both experimenters staged an unfair approach towards the subject. Particularly, they overly underlined the wrong answers. Each answer was followed by implicit signs of consternation, like head shake of disapproval. Moreover, along the procedure, there were 5 scripted interventions of experimenters planned at fixed moments of the protocol to reinforce the administered condition: “you should try harder, your score is not good", "sit up straight, or the recording will be useless", "are we recording? Because here is a mess", "if he/she keep going like this it is only a waste of time", "we can stop now, it is totally useless".

Mood self-assessment and Explanation - At the end of the elicited phase the Self-Assessment Manikin (SAM) questionnaire was provided to the subject [21]. The SAM is a picture-oriented instrument containing images for each of the three affective dimensions (pleasure, arousal and dominance). Such images were hence given a score by the participant in order to describe how he felt during the test in a quantitative manner. After that, the real aim of the experiment was revealed to the subject, explaining the importance of unawareness of the real condition in this type of emotion elicitation, and the subject is allowed to remove wearable sensors from himself.

Afterwards, the subject was asked to fill the Beck Depression Inventory [22] and the Maudsley Obsessive Compulsive Questionnaire [23] to assess the presence of depression and obsessive-compulsive behaviour in subjects. In fact, these are two mental diseases that significantly alter the self-perception of emotions [24]–[27] and their presence would render the subject ineligible for the purposes of the test.

Then, the experimenter asks the subject to indicate as accurately as possible the moment he felt the emotion he pointed out by filling the SAM. Thus, experimenters check if these moments are close to any of one of the 5 reinforcements used. Such information is important for the subsequent analysis (see Sect. 4.1.2). Finally, in the case negative condition was administered, the experimenters apologized for the pretended rude attitude.

Figure 2: (a) Experimental Set-up (b) a participant during the positive exited phase (c) a participant during the negative exited phase

3.3. Participants

Twenty-one voluntary healthy young subjects (9 male, 12 female) with a mean age of 24.0 years (standard deviation: 3.7, range: 20-35) participated on purpose in this study. Among tested subjects, 2 were occasional smokers whilst the remaining ones were no smokers. Participants completed the experimental session at the Scuola Superiore Sant'Anna (Pisa, Italy). Written informed consent was obtained from all the participants prior entering into the study. Only the results obtained in 17 out of 21 subjects resulted eligible for the purposes of the study. In fact, two subject's results were discarded since no significant emotion arousal emerged from the SAM test, one experienced BioHarness failure while recording, one understood the purposes of the test prior its administration. None of the subject resulted depressed or affected by obsessive-compulsive behavior.

4. Data Analysis

The physiological signals acquired during the experimental protocol were analyzed using Matlab 2016 (MathWorks, Massachusetts USA). The whole dataset, for each subject, was segmented in three distinct phases: baseline (5 minutes), relax (5 minutes) and elicitation (15 minutes).

The classification process between different emotional states, such as positive, negative and relaxed, in the prospective of a social robotics application should occur in a much smaller time than the entire duration of the experimental session. Therefore, the features extraction strategy discussed below was applied on signals samples belonging to 180s-long windows. The choosing of 180s as windows time allows to catch significant variations from the GSR signal in response to the external stimuli. The chosen time periods are affordable for the final purpose of this work, considering that emotional state require a certain physiological latency time to manifest. Moreover, the selected time period is one of most used ranges in physiological data analysis for emotion recognition [28]. A 50s-long overlapping of adjacent windows was implemented to correctly handle the transitions. Hence, this emotion recognition process would enable the system to recognize change in user's emotions each 130s. Two, two and six time windows were obtained from the baseline, relaxed and the exited phases for each subject, respectively. A pre-processing strategy and a features extraction algorithm were applied on each new dataset that included signals related to cardiac, electrodermal and brain activity. Each dataset window was now represented by a 42 features vector: 16 from HRV, 16 from GSR and 10 from brain activity. Their extraction and description is based on [29] and it is discussed below.

4.1.1. Signal processing and features extraction

The main informative channel in the field of emotion recognition related to the cardiac activity is the R-to-R signals. R-to-R samples were negatively affected by the presence of ectopic components of cardiac activity. Ectopic events are variations of the cardiac activity due to arrhythmia or fibrillation. The presence of ectopic beats gives deceivable results in HR analysis so they must be removed. A R-to-R sample was considered an ectopic interval if its percentage relative difference from the previous sample was¸ in absolute value, greater than 20% [30]. In this case, the current value is replaced with the average of the two antecedent and two following ones. After ectopy removal, 6 time-domain features were extracted for each time window (Table 2).

Table 2 Time-domain features from R-to-R values

Feature name	Description
RR mean	Mean of R-to-R inter-beat intervals belonging to the same time window
SDNN	Standard deviation of normal RR intervals (also said as NN intervals)
HR mean	Mean of heart rate
SD mean	Heart rate standard deviation
RMSSD	Square root of the mean of the squared differences between adjacent NN intervals
pNN50	Percentage of differences between adjacent NN intervals exceeding 50 ms

In order to extract frequency-domain features, the signal is firstly resampled at 4 Hz [31] and, secondly, smoothed through a quadratic fitting characterised by a sliding window with a width equal to 10% of samples amount. This fitting was subtracted from the signal, the latter was added with its mean beforehand. The obtained sequence was interpolated through a cubic spline interpolation and deprived of its mean value. Finally, after having been multiplied by a Hamming window of the same length, power spectral density (PSD) of the signal, with resolution 1/2048 Hz, was estimated through Burg method of 16-order [32]. A total of 10 frequency-domain features were extracted after PSD estimation (Table 3).

Table 3 Frequency domain features from R to R values

Feature name	Description
VLF peak	Frequency peak of heart activity VLF (0-0.04 Hz)
VLF power	PSD area in VLF
%VLF	Percentage ratio between PSD area in VLF and total one
LF peak	Frequency peak of LF (0.04-0.15 Hz)
LF power	PSD area in LF
%LF	Percentage ratio between PSD area in LF and total one
HF peak	Frequency peak of HF (0.15-0.40 Hz)
HF power	PSD area in HF
%HF	Percentage ratio between PSD area in HF and total one
LF/HF	Ratio between LF and HF powers

Table 4 GSR extracted features.

Feature name	Description
# Startle	Number of detected startles
Amplitude mean	Mean value of startles peak amplitude (μS)
Amplitude std	Standard deviation of startles peak amplitude (μS)
Sum rise time	Sum of all detected startles rise time duration within the phasic signal portion analysed (s)
Sum fall time	Sum of all detected startles fall time duration within the phasic signal portion analysed (s)
Rise rate mean	Mean value of a startle rise time (s)
Rise rate std	Standard deviation of startle rise time (s)
Decay rate mean	Mean of a startle fall time (s)
Decay rate std	Standard deviation of a startle fall time (s)
Phasic value mean	Mean value of the phasic signal (μS)
Phasic value std	Standard deviation of the phasic signal (μS)
Startle time mean	Mean value of a startle duration (s)
Startle time std	Standard deviation of a startle duration (s)
Startle RMS mean	Mean Value of the Root Mean Square of the curve identifying a startle (μS)
Startle RMS std	Standard deviation of the Root Mean Square of the curve identifying a startle (μS)
Startle RMS overall	Value of the Root Mean Square of the whole phasic signal portion analysed (μS)

The Shimmer^{TM} sensor provides as output galvanic resistance that has been converted into galvanic skin conductance (SC). The SC is characterised by a slow varying tonic activity and a relatively fast varying phasic activity, which is characterised by local peaks, so-called startles, lasting between 1 and 5 seconds [33]. Thus, the GSR is obtained by evaluating the features of these peaks; their frequency content is entirely located within 1 Hz [34]. On the contrary, the tonic phase is due to sweating activities, taking place over periods of time longer than 5 seconds, not related to the phasic component. The Shimmer frequency sampling is set to 51.2 Hz, and thus, the signal was filtered through a 4th-order Butterworth low-pass filter, with a cut-off frequency of 2 Hz. Thus,the tonic component was extracted through average process by means of a 5 seconds-long moving window. By subtracting tonic level from the filtered conductance signal, phasic response was obtained. Startles features are therefore extracted and the parameters reported in Table 4 are obtained.

Regarding features from brain activity, they are provided by the Mindwave itself. From this sensor, 10 features related to the PSD of the recorded brain activity were extracted (Table 5). Regarding parameters named “Attention” and “Meditation”, they are calculated by the device itself and the rationale behind them is protected by enterprise copyright. They both range from 0 to 100.

Table 5: Brain activity features extracted

Feature name	Description
Alpha1	EEG signal power in frequency range 8-9 Hz
Alpha2	EEG signal power in frequency range 10-12 Hz
Beta1	EEG signal power in frequency range 13-17 Hz
Beta2	EEG signal power in frequency range 18-30 Hz
Delta	EEG signal power in frequency range 1-3 Hz
Gamma1	EEG signal power in frequency range 31-40 Hz
Gamma2	EEG signal power in frequency range 41-50 Hz
Theta	EEG signal power in frequency range 4-7 Hz
Attention	NeuroSky index for user’s level of mental “focus” or “concentration”
Meditation	NeuroSky index for user’s level of mental “calmness” and “relaxation”

4.1.2. Label and Feature Reduction

For each subject a baseline value was calculated by computing the mean values from the features correspondent to the baseline phase. Then, the remained feature sets (relaxed and elicited) have been scaled as a percent variation compared to the baseline value in order to eliminate the interpersonal variability of physiological signals, thus to normalize the dataset.

It was important to understand which instances of the elicited phase could be used as carrier of reliable information about emotional states of a subject. Indeed, the minimal condition for this to be true was that the subject effectively had felt an emotion, better said a mood [35]. For each subject, the instances between the end of the relaxed phase and the moment in which subjects declared to have felt the emotion (see Sect. 3.2) were discarded. Therefore, only the remaining instances of the elicited phase were maintained and used for further analysis, Then, the dataset was manually labelled. The selected instances coming from positive condition were labelled as “Positive”, while those ones from negative condition as “Negative”. Instead, the features vectors coming from relaxed phase were labelled as “Relaxed”.

The normal distribution of the computed features was verified using the Shapiro-Wilk test of normality. As all raw scores were not normally distributed, the non-parametric Spearman (RHO) correlation coefficient between each feature was computed to keep in the analysis only the features with a coefficient greater than 0.85 (absolute values) and $p<0.05$ in order to reduce the noise due to the redundancy of data [36]. Mann-Whitney test was then applied to understand which set of uncorrelated features could be more useful in distinguishing between positive and negative moods. Consequently, the test was applied comparing feature values labelled as “Positive” with feature values labelled as “Negative”. If the p was lower than 0.05, the specific feature was discarded from the analysis.

First, the remaining features were normalized with Z-norm to avoid distortion and to have a zero mean and a unit standard deviation. Principal Component Analysis

(PCA) and Sammon's Map (SM) – applied sequentially – were used to identify how the subjects, in relation to the relaxed, positive and negative moods, would be grouped and consequently classified. As confirmed in [37] and [38], both approaches are efficiently applied to reduce the dimensionality of the dataset to a two-dimensional representation of data, which is more intuitive to understand, . As concern PCA, according to the Kaiser Rule [39], we consider only components with eigenvalues greater than 1.

4.1.3. Supervised and Unsupervised classification

The obtained dataset was then used to train supervised and unsupervised machine learning algorithms to compare the results. As concern the supervised method, a simple Decision Tree (DT) with a split criterion based on Gini's diversity index, Support Vector Machine (SVM) with linear kernel and fine k-nearest neighbour (k-NN) with 1 neighbour based on Euclidian distance were applied. The performance of these algorithms was assessed with 5-cross-fold validation technique.

Three unsupervised machine learning clustering techniques were used to group the instances into clusters. Particularly, in this work, the K-Mean (KM) algorithm based on Euclidian distance with three replicates to avoid local minima, Self-Organizing Map (SOM) with a hexagonal layer topology function and Hierarchical Clustering (HC) based on ward linkage were applied and compared. In particular, KM was applied considering the results obtained with the unsupervised and supervised machine learning algorithms were presented as a confusion matrix. Thus, the overall accuracy, F-measure, precision, and recall were computed to evaluate and compare the performance [40]. The Silhouette coefficient was used to study the separation distance between the resulting clusters. Silhouette coefficient value close to 1 highlights that the clusters are separated [41]. Additionally, the performance of the supervised analysis was used as a gold standard to compare the performance of the unsupervised approach.

5. Results

The aim of this work was to investigate with a wearable sensor system the relaxed status and two opposite moods evoked during a social interaction. At the end of the tests, participants were asked to compile a self-assessment questionnaire of the perceived mood. The results are shown in Table 6. The pleasure, as used in psychology, especially in discussing emotions, stands for the intrinsic attractiveness or averseness of an event, object, or situation. In this questionnaire, it ranges from -4 to 4 in the integer domain. The subjects expressed positive average value (3.15) during the positive stimulation and negative average value (-1.22) during the negative stimulation, which means that the proposed methodology is able to evoke in the subject the desired moods, even if the average dominance value is < 1 for both moods. Table 6 reports the average value, the standard deviations and the range of the SAM.

According to the methodology described in the previous section, a total of 96 instances were extracted from the acquired signals. Particularly, 34 instances are related to the relax phase, 35 are related to the positive mood and 27 to the negative mood. The final set of parameters was reduced with the Spearman correlation coefficient and the Mann-Whitney test from 42 to 12 features. The acquired signals contributed to the final dataset with the following features: HR mean, SDHR, LF power, %VLF, %HF from ECG, # startle, Rise time mean, RMS std from EDA and Beta2, Gamma2, Delta, Meditation from brain activity. According to Kaiser rule, the PCA reduced the dataset to 4 components, and SM finally reduce to 2D space. From a visual inspection, the three distinct groups were separated in the reduced feature space, as depicted in Figure 3. The results of the unsupervised analysis confirms the presence of three clusters with average silhouette values of 0.63, 0.62 and 0.60 for KM, SOM and HC respectively, which underline a quite good separation between the groups.

Table 6. Results of SAM Questionnaires.

		Valence	Arousal	Dominance
Positive Mood	Average	3,12	2,13	0,12
	SD	0,83	1,13	1,81
	Max	4	4	2
	Min	2	1	-2
Negative Mood	Average	-1,22	0,78	0,89
	SD	1,64	1,64	1,90
	Max	0 (2*)	2	3
	Min	-4	-3	-2

*a participant declared to enjoy the tests. It is considered as outlier in this case.

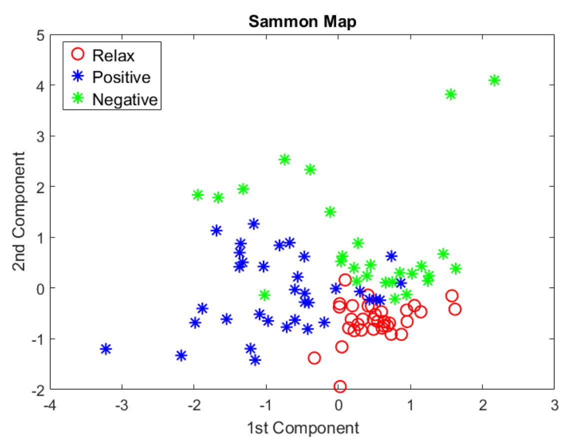

Figure 3. Data visualization in Sammon's Map space.

The three supervised techniques showed high results in terms of accuracy and F-measure. In particular, as reported in Table 7 (referring to supervised algorithms), the best results were obtained with the k-NN with an accuracy of 0.86 and a F-measure of 0.93. The DT algorithm showed a lower accuracy and F-measure (0.78 and 0.81 respectively) with respect to SVM which achieved an accuracy value of 0.81 and an F-measure of 0.81. The complete results of evaluation parameters are reported in Table 7. The three unsupervised algorithms show lower, still comparable results to the supervised ones, concerning accuracy and the F-measure. As regards precision and recall, the results are comparable for all the measures (average value equal to 0.80). Table 8 shows and compares the F-measure for all the machine learning used in this study in recognizing the three different moods. Particularly, the worst recognized status is the negative mood, which is often confused with relax status. It is worth underlying that, in terms of F-Measure, the KM and the SOM show higher results with respect to the DT and the SVM methods in recognizing the positive mood.

Table 7. Performance of the supervised and unsupervised methods.

	Accuracy	F-Measure	Precision	Recall	Specificity
DT	0,78	0,81	0,78	0,79	0,88
SVM	0,81	0,87	0,81	0,81	0,90
k-NN	0,86	0,93	0,86	0,86	0,93
KM	0,76	0,73	0,83	0,73	0,87
SOM	0,72	0,80	0,80	0,80	0,89
HC	0,65	0,54	0,67	0,61	0,78

Table 8. F-measure results of the supervised and unsupervised approaches in recognizing the three states (relaxed, positive and negative).

F-Measure	Relaxed	Positive	Negative
DT	0,81	0,73	0,81
SVM	0,87	0,79	0,77
k-NN	0,93	0,88	0,80
KM	0,75	0,88	0,58
SOM	0,72	0,86	0,46
HC	0,70	0,80	0,13

6. Discussion and Conclusion

The aim of the current paper was to develop and investigate the use of physiological wearable sensor system for mood detection during a social interaction. The proposed analysis achieves high results for both machine learning approaches (complete results are reported in Table 7), showing that our wearable system is able to distinguish among the different moods quantitatively. Furthermore, the data acquired have led to satisfactory qualitatively results in correctly grouping the features (PCA and SM), thus this work can be considered a promising pilot study in which the feasibility of the system is demonstrated.

The main innovation of this study consists in using a physiological wearable system, able to measure HRV, EDA and EEG, in order to obtain parameters the variations of which are highly correlated with changes in different mental/physical conditions. Additionally, another important innovation is the use of a protocol which involves a social interaction to induce different mood in the subject. As depicted in Figure 3, the negative group (green stars) is more scattered respect to the other two groups, whereas the relax instances are the most grouped (red circles). This different distribution of the groups in the space can be explained by the inter- and intra- subject variability of the reaction to the stimuli during the same induction protocol, as confirmed also by the standard deviations of the pleasure value of the SAM questionnaires (see Table 6). Importantly, human beings could react differently to the same social/external stimuli. For instance, during the test, a subject (LD20) was stimulated with the negative condition, but during the self-assessment phase he declared he enjoy the questionnaires giving high scores in the SAM (valence=2). Indeed, his physiological parameters reveal a status more similar to the positive mood group as depicted in Figure 3. Moreover, another subject (BF16) was tested for the positive moods but he/she reported the lower SAM value for the valence among the overall cohort. This result is aligned to what captured and analysed by the sensor system. Reasonably, the instances related to this subject are on the edge with the negative cohort (Figure 3). Although the data processing of this work was offline performed, some considerations about the concrete use in real situations, particularly with a real time approach, could be already discussed. It is worth to mention that SOM and KM algorithms show accuracy values > 0.70 comparable to the supervised methods. The unlabelled approach can help the recognition of different moods resulting hence suitable the exploitation of this system in the real world where data labelling is not always possible. Additionally, the window of 180s we chose is reasonably appropriate to distinguish among the moods.

This work, however, presents some limitations. We analysed only three state, one relaxed and two stimulated (positive and negative); future work should investigate also more specific emotion or human moods. This is a preliminary study to investigate the possibility to use innovative protocol and informative channels to induce and to recognize moods. Further studies should investigate which informative channels are more appropriate in distinguishing moods, thus to balance the trade-off between accuracy and obtrusiveness. Another limitation is the number of participants. Future studies will be performed to increment the number of participants thus to reinforce the results.

Acknowledgement

This work was supported by DAPHNE project (REGIONE TOSCANA PAR FAS 2007-2013, BANDO FAS SALUTE 2014, CUP J52I16000170002).

References

[1] L. Penteridis *et al.*, “Robotic and Sensor Technologies for Mobility in Older People,” *Rejuvenation Res.*, vol. 20, no. 5, pp. 401–410, Oct. 2017.

[2] P. P. Ray, “Internet of Robotic Things: Concept, Technologies, and Challenges,” *IEEE Access*, vol. 4, pp. 9489–9500, 2016.

[3] J. LeDoux, “Chapter 26 Emotional networks and motor control: a fearful view,” 1996, pp. 437–446.

[4] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, and P. Dario, “Emotion Modelling for Social Robotics Applications: A Review,” *J. Bionic Eng.*, vol. 15, no. 2, 2018.

[5] A. C. Granero, F. Fuentes-Hurtado, V. N. Ornedo, J. G. Provinciale, J. M. Ausín, and M. A. Raya, “A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents,” *Front. Comput. Neurosci.*, vol. 10, 2016.

[6] M. Leo *et al.*, “Automatic Emotion Recognition in Robot-Children Interaction for ASD Treatment,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2016.

[7] E. Harmon-Jones, D. M. Amodio, and L. R. Zinner, “Social Psychological Methods of Emotion Elicitation,” in *Handbook of Emotion Elicitation and Assessment*, 2007, pp. 91–105.

[8] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion.,” *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.

[9] J. A. Russell, “A circumplex model of affect.,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[10] M. Lewis and L. Canamero, “Are Discrete Emotions Useful in Human-Robot Interaction? Feedback from Motion Capture Analysis,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 97–102.

[11] S. Koelstra and I. Patras, “Fusion of facial expressions and EEG for implicit affective tagging,” *Image Vis. Comput.*, 2013.

[12] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo, “Arousal Recognition System based on Heartbeat Dynamics During Auditory Elicitation,” 2014.

[13] K. Rattanyu, M. Ohkura, and M. Mizukawa, “Emotion monitoring from physiological signals for service robots in the living space,” *Control Autom. Syst. (ICCAS), 2010 Int. Conf.*, pp. 580–583, 2010.

[14] M. Khezri, M. Firoozabadi, and A. R. Sharafat, “Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals,” *Comput. Methods Programs Biomed.*, 2015.

[15] R. Henriques, A. Paiva, and C. Antunes, “Accessing Emotion Patterns from Affective Interactions

Using Electrodermal Activity," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 43–48.
[16] C. N. Moridis and A. A. Economides, "Prediction of student's mood during an online test using formula-based and neural network-based method," *Comput. Educ.*, vol. 53, no. 3, pp. 644–652, 2009.
[17] D. Handayani, H. Yaacob, A. Wahab, and I. F. T. Alshaikli, "Statistical Approach for a Complex Emotion Recognition Based on EEG Features," *2015 4th Int. Conf. Adv. Comput. Sci. Appl. Technol.*, pp. 202–207, 2015.
[18] M. Greenwald *et al.*, "International Affective Picture System (IAPS): Technical Manual and Affective Ratings," *International Affective Picture System (IAPS)*. 1997.
[19] N. A. Roberts, J. L. Tsai, and J. A. Coan, "Emotion Elicitation Using Dyadic Interaction Tasks," in *Handbook of Emotion Elicitation and Assessment*, 2007, pp. 106–123.
[20] M. Khezri, M. Firoozabadi, and A. R. Sharafat, "Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals," *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 149–164, 2015.
[21] T. Bynion and M. T. Feldner, "Encyclopedia of Personality and Individual Differences," no. January, 2017.
[22] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, "An Inventory for Measuring Depression," *Arch. Gen. Psychiatry*, vol. 4, pp. 561–571, 1961.
[23] L. G. Sternberger and G. L. Burns, "Maudsley Obsessional-Compulsive Inventory: obsessions and compulsions in a nonclinical sample," *Behav. Res. Ther.*, vol. 28, no. 4, pp. 337–340, 1990.
[24] S. A. Surguladze, C. Senior, A. W. Young, G. Brébion, M. J. Travis, and M. L. Phillips, "Recognition Accuracy and Response Bias to Happy and Sad Facial Expressions in Patients with Major Depression," *Neuropsychology*, vol. 18, no. 2, pp. 212–218, 2004.
[25] D. C. Zuroff and S. A. Collusy, "Emotion Revognition in Schizophrenic and Depressed Inpatients," *J. Clin. Psychol.*, vol. 42, no. 3, pp. 411–417, 1984.
[26] R. Sprengelmeyer, M. Rausch, U. T. Eysel, and H. Przuntek, "Neural structures associated with recognition of facial expressions of basic emotions," *Proc. R. Soc. B Biol. Sci.*, vol. 265, no. 1409, pp. 1927–1931, 1998.
[27] R. Sprengelmeyer *et al.*, "Disgust implicated in obsessive-compulsive disorder," *Proc. R. Soc. B Biol. Sci.*, vol. 264, no. 1389, pp. 1767–1773, 1997.
[28] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.
[29] S. Betti *et al.*, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, 2017.
[30] N. Lippman, K. M. Stein, and B. B. Lerman, "Comparison of methods for removal of ectopy in measurement of heart rate variability," *Am. J. Physiol.*, vol. 267, no. 1, pp. H411–H418, 1994.
[31] B. Mali, S. Zulj, R. Magjarevic, D. Miklavcic, and T. Jarm, "Matlab-based tool for ECG and HRV analysis," *Biomed. Signal Process. Control*, vol. 10, no. 1, pp. 108–116, 2014.
[32] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, 2006.
[33] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, 2010.
[34] W. Boucsein, *Electrodermal activity*. Springer Science+Business Media, LLC, 2012.
[35] K. Lochner and M. Eid, "Successful emotions: How emotions drive cognitive performance," in *Successful Emotions: How Emotions Drive Cognitive Performance*, 2016, pp. 43–67.
[36] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.
[37] S. Patel *et al.*, "Analysis of the severity of dyskinesia in Patients with Parkinson's disease via wearable sensors," *Proc. - BSN 2006 Int. Work. Wearable Implant. Body Sens. Networks*, vol. 2006, no. May, pp. 123–126, 2006.
[38] L. Fiorini, F. Cavallo, P. Dario, A. Eavis, and P. Caleb-Solly, "Unsupervised Machine Learning for Developing Personalised Behaviour Models Using Activity Data," *Sensors*, vol. 17, no. 5, p. 1034, 2017.
[39] H. F. Kaiser, "The Application of Electronic Computers to Factor Analysis," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 141–151, Apr. 1960.
[40] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
[41] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987.