Physiological wireless sensor network for the detection of human moods to enhance human-robot interaction

Physiological Wireless Sensor Network for the Detection of Human Moods to Enhance Human-Robot Interaction

Francesco Semeraro, Laura Fiorini, Stefano Betti, Gianmaria Mancioppi, Luca Santarelli and Filippo Cavallo

Abstract Although it is already possible to issue utility services that use robots, these are still not perceived by society as capable of actually delivering them. One of the main motivations is the lack of a human-like behaviour in the interaction with the user. This is displayed both at physical and cognitive level. This work investigates the optimal sensor configuration in the recognition of three different moods, as it surely represents a crucial element in the enhancement of the human-robot interaction. Mainly focusing towards a future application in the field of assistive robotics, electrocardiogram, electrodermal activity and electroencephalographic signal were used as main informative channels, acquired through a wireless wearable sensor network. An experimental methodology was built to induce three different emotional states by means of social interaction. Collected data were classified with six supervised machine learning approaches, namely decision tree, induction rules and lazy, probabilistic and function-based classifiers. The results of this work revealed that the optimal configuration of sensors which maximizes the trade-off between accuracy and obtrusiveness is the one surveying cardiac and skin activities. This sensor configuration reached an accuracy of 87.07% in the best case.

Keywords Mood detection · Physiological sensors · MIP · Social interaction

F. Semeraro · L. Fiorini (✉) · S. Betti · G. Mancioppi · L. Santarelli · F. Cavallo
The BioRobotics Institute, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio 34, 56025 Pontedera (Pisa), Italy
e-mail: laura.fiorini@santannapisa.it

F. Semeraro
e-mail: francesco.semeraro@santannapisa.it

S. Betti
e-mail: stefano.betti@santannapisa.it

G. Mancioppi
e-mail: gianmaria.mancioppi@santannapisa.it

L. Santarelli
e-mail: luca.santarelli@santannapisa.it

F. Cavallo
e-mail: filippo.cavallo@santannapisa.it

A. Leone et al. (eds.), *Ambient Assisted Living*, Lecture Notes in Electrical Engineering 544, https://doi.org/10.1007/978-3-030-05921-7_30

1 Introduction

In a not remote future scenario, mobile robotic platforms will share environments with humans in most of their daily time, such as houses, offices and public spaces. The interactions between robots and humans would be led not only by the simple presence of humans in the environment, but also and especially by the current activity done by humans. Hence, it is crucial for a service robot acting in a social unstructured environment to be sensitive to emotions and properly react to them [1]. Additionally, robots should be able to engage a human-like interaction with users to efficiently cooperate with human beings.

It is demonstrated that several aspects influence the interaction between people, including emotions which can affect the decision-making process [2]. In this sense, emotions can be used to improve human-robot interaction [3], to efficiently handle relationships with humans or other autonomous physical agents by means of social behaviours and rules. Particularly, it is important, for an assistive robot, to monitor feelings of the user, not only for interactional purposes but also to assess his/her cognitive development, aspect usually ignored in the elderly care, especially in the case of Mild Cognitive Impairment (MCI) [4]. For this purpose, being aware of the emotional state of the user through his/her physiological activity could be extremely useful. Nonetheless, by issuing such sensing modality to this purpose, the robot will be continuously exposed to indicators of the physiological health of the user, enhancing this role of elderly monitoring.

2 Related Work

Most affective computing applications primarily use vision sensors [1], since the detection of facial expressions is the most natural and intuitive way to perceive emotions of people. However, it recently has become evident that the typical use of vision in social robotics is not feasible in real-life applications because of the occlusion of the camera, adequate room lighting and privacy issues [5]. It is also known that physiological systems are regulated by the autonomous nervous system, strictly related to emotions [6]. So, physiology surely constitutes a reliable informative channel to emotion recognition [1].

Besides privacy issues and problems of good conditions to acquire good-quality images, the user could sometimes hide the external expression of emotion, invalidating the recognition process. Moreover, if, in the context of an assistive robotics application, the camera is mounted on the robot, this last will engage the user prior to having clues about his/her emotional state, making this application pointless.

On the other side, physiology could partially solve these limitations. In fact, as people cannot consciously handle their own physiological activity, unless through proper training, emotional awareness of the robot could not be deceived. This is very helpful, mainly considering emotional monitoring in elderly care. Furthermore,

sensors would continuously stream data to the robot, enabling it to be aware of the emotional state of users before starting an interaction with him/her. Moreover, there are fewer problems related to needed conditions to have a good acquisition of this informative channel [7, 8].

Several works exploited physiological signals such as electrodermal activity (EDA), the electrocardiogram (ECG), and the electroencephalography (EEG) (Table 1). For instance, Nardelli et al. [9] and evoked the emotions in the subjects by means of the International Affective Digitalized Sounds (IADS) dataset and recorded the ECG signal, achieving a recognition accuracy of 84.26%. Instead, Henriques et al. [10] used an actor and a robot to provoke emotions in the subjects.

Khezri et al. [11] performed a detailed analysis by examining multiple parameters. Specifically, they recorded EEG, electrooculography (EOG), electromyography (EMG), blood volume pressure (BVP), galvanic skin response (GSR), and heart rate variability.

(HRV) from subjects simultaneously, while building an emotion recognition process based on a weighted linear fusion model.

As stated from literature [1], but also depicted in Table 1, subjects are elicited with different mood induction procedures (MIP). In most of these cases, subjects were asked to sit in front of a computer and watch films or static images or listen

Table 1 Comparison of related works (DT = Decision Tree, SVM = Support Vector Machine, HMM = Hidden Markov Model, MLP = Multilayer Perception, KNN = k-nearest neighbour, lazy = lazy learners, Bayesian = Bayesian learners)

References	Physiological source	Machine learning	Activity to evoke moods
Nardelli et al. [9]	ECG	Quadratic discriminant classifier	27 subjects were stimulated with sounds from the IADS dataset
Henriques et al. [10]	EDA	HMM, iBk	30 participants (two groups). Group one: emotion elicited by an actor's performance; Group two: elicited by NAO robot
Khezri et al. [11]	EEG, ECG, EDA, electrooculography, blood volume pressure, electromyography	SVM, k-nearest neighbour	The emotions in 50 subjects were elicited through videos
Our work	EDA, ECG, EEG	DT, SVM, MLP, KNN, Bayesian, lazy	20 participants. Three moods were elicited during a social interaction

to music [9, 11]. Even if these are effective MIPs, they do not address challenges in social robotics fields. Indeed, in these experiments, the dyadic interaction [12] is not considered.

3 Aim of the Work

This work aims to develop and test a physiological wearable system able to recognize three different moods, inside a purposely built MIP by means of social interaction: relaxed, positive and negative. Particularly, in this work, the authors investigated the optimal sensor configuration able to distinguish the three different moods with a good trade-off between the accuracy of the performance and the obtrusiveness of the system.

The proposed system is composed of three different physiological sensors able to monitor the EDA, the ECG and the EEG signals. Additionally, an RGD-B camera will be used to record all the experiment to support the analysis of the collected data and discuss the results, as common practice in emotional experiments [13]. Five supervised machine learning classifiers families, namely decision tree, induction rules, lazy, probabilistic learners and function-based classifiers were used to test and to compare the different sensor configurations.

4 Material and Method

The first part of this section describes the system and the data acquisition, whereas the second part is related to the data analysis.

4.1 Instruments

Three commercial wearable sensors were selected to measure physiological response to the emotional stimuli. The reasonable trade-off between accuracy in measurements and unobtrusiveness led to the selection of Zephyr™ BioHarness3™ (Zephyr™, Maryland, USA), Shimmer™ GSR Module (Shimmer™, Ireland) and The MindWave™ EEG headset (Neurosky®, California USA) (See Fig. 1).

Zephyr BioHarness is a Bluetooth chest belt capable of monitoring cardiac activity by recording the ECG signal and calculating parameters such as Heart Rate and R-R Intervals [14]. The ECG signal is sampled at 250 Hz. Inter-Beat-Interval data provided by the device were used for data analysis and features extraction (Fig. 1a).

The Shimmer GSR Module is a small-size, lightweight wearable sensor that streams one channel data related to EDA at sampling rate of 51.2 Hz [15]. The Shimmer module is composed of two special finger electrodes and the main unit

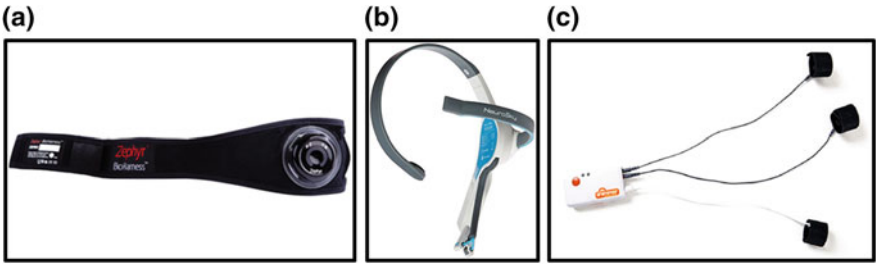

Fig. 1 **a** BioHarness. **b** MindWave headset. **c** Shimmer GSR module

that streams data related to the galvanic skin response using a Bluetooth connection (Fig. 1b).

The MindWave headset is a Bluetooth device able to capture single frontal lobe channel EEG raw data at a sample rate of 512 Hz [16]. In addition, the headset provided the indexes of attention and meditation of the user, related to the frequency power spectrum of the acquired signal, with a sample rate of 1 Hz (Fig. 1c).

A Kinect™ v1.80 camera (Microsoft®, Washington, USA), 30 Hz transmission of 640 × 480 images, was included in the system to record the experimentation thus to support the data analysis.

The overall set of sensors was handled through an interface implemented in Microsoft Visual Studio™ 2017 (Microsoft®, Washington, USA). It managed the connection of devices, the data acquisition, synchronization and storage. Furthermore, the experimenter could use this interface to take timed notes of events occurring during the experimentation.

4.2 Experimental Protocol

The protocol purposely designed to grow a positive or negative mood in the user through social interaction with the experimenters involved the oral administration of a questionnaire (Fig. 3). Before starting the 15 min phase dedicated to the MIP

Fig. 2 Process of the experimental methodology

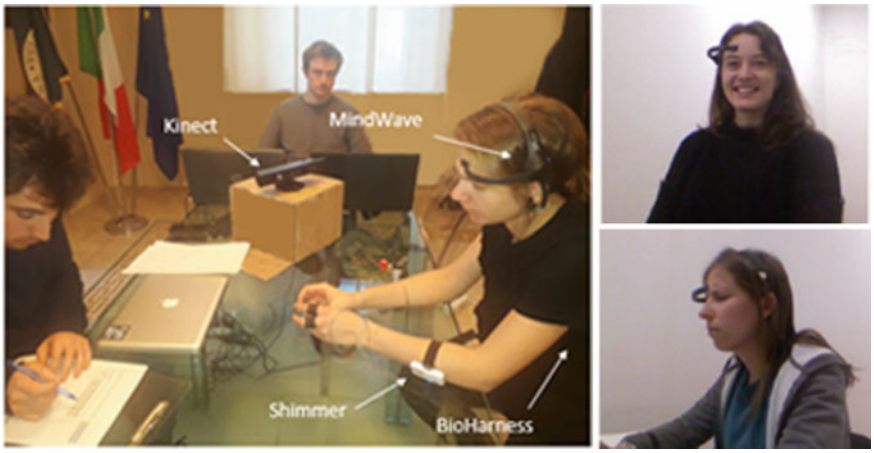

Fig. 3 Experimental set-up

(elicited phase), the subject, unaware of the real purpose of the test [17], was dressed in wearable sensors and accommodated on a chair inside a room with no specific visual and sound input. After that, the user remained for 10 min in the absence of stimuli, initially to record a baseline (baseline phase) of the physiological parameters and subsequently to achieve a state of relaxation (relax phase), closing his eyes and avoiding interactions with objects and people (Fig. 2).

After that, the interaction between the experimenters and the subject came into play to stimulate a specific mood. It was impossible to sequentially administer the positive and negative condition to the same subject, because the mood achieved during the first type of social interaction could have influenced the subject's mood during the second one. Only one between positive and negative conditions was administered to each subject, maintaining a balance in terms of percentage between the two cases.

In the positive condition, one of the experimenters, who was a psychologist (the other one, an engineer, was mainly charged to monitor the sensor interface, providing complementary interaction during each session), was immediately very affable with the subject, who had to answer orally to 55 questions selecting one of the 4 possible answers, among which just one was correct. The questions selected for this condition 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. Each answer was followed by implicit signs of appreciation, like nodding. However, there were 5 interventions experimenters planned at a fixed moment of the protocol to reinforce the administered condition. Indeed, these 5 moments were intended to gradually increase the complacency and gratification of the user.

The negative condition was dual to the positive one. In this case, instructions were provided unkindly by the second experimenter, which entered during the first recordings without introducing himself. The 55 questions had a level of difficulty that causes discomfort and shame in subjects, when not capable of answering them.

During the task execution, both experimenters performed a mean attitude towards the subject, through signs of contempt and 5 negative scheduled reinforcements, delivered with modality which was dual to the positive condition.

At the end of positive or negative induction phase, the Self-Assessment Mannikin (SAM) questionnaire was proposed to the subject [18]. It is a picture-oriented questionnaire, containing nine images for each of the three affective dimensions (pleasure, arousal and dominance). Filling the SAM, the participant described how and what they felt during the test in a quantitative manner.

Given the importance of the subject's unawareness on the real purpose of the test induction, the real aim of the experiment was revealed to the subject at the end of the protocol, after having assessed the persistence of unawareness through a post-experimental interview. Then, the experimenters asked the subject if they identified a moment coinciding or close to one of the 5 reinforcements after which the felt emotion mostly suited the selected SAM's scores. This information was used in the subsequent analysis (see Sect. 4.4.2). Finally, the subject filled the Beck Depression Inventory and the Maudsley Obsessive Compulsive Questionnaire to assess the presence of depression and obsessive-compulsive behaviour in subjects. In fact, these are two mental diseases that alter significantly the self-perception of emotions and their presence would represent an exclusion criterion. In the negative condition, the experimenters apologized for the pretended rude attitude.

4.3 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, two were occasional smokers, and the remaining 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 entry into the study.

Unfortunately, one subject was discarded during the post-processing analysis. The HRV signal was not usable probably because the BioHarness was not worn properly in this case. All the subjects were not depressed nor affected by obsessive-compulsive behaviour.

4.4 Data Analysis

The physiological signals obtained during the experimental protocol from the three sensors for each subject were segmented to subsequently obtain three different datasets, one for each of the three phases (baseline, relax, elicited). The dataset was analysed using Matlab 2016 (MathWorks, Massachusetts USA).

The features extraction strategy discussed below was applied to signals samples belonging to 180 s-long windows. In fact, emotion recognition is not a real-time application in common sense, as user's emotional state, by its nature, is not expected to change with a high-frequency. Moreover, the selected window allows analysing sufficiently long physiological signal ranges to be able to extract also the main frequency characteristics. Indeed, the selected time is one of the most used ranges in physiological data analysis for emotion recognition [6, 19]. A 50 s-long overlapping of adjacent windows was implemented to correctly handle the transitions.

4.4.1 Physiological Signals Analysis

The BioHarness device provided HRV data that specify the temporal distance between one heartbeat and the following one. Ectopic rhythm, which is an irregular heart rhythm due to a premature heartbeat, was identified and corrected. A R-to-R sample was considered an ectopic interval if its difference from the previous sample was, in absolute value, greater than 20% [20]. By correcting ectopic rhythm, replacing with the average of the two antecedent and two following ones, a Normal-to-Normal (NN) interval sequence, appropriate for HRV analysis, was obtained [21]. From NN signal, 6 time-domain features were extracted for each time window (Table 2).

Then, after a signal smooth detrending, the NN interval sequence was resampled at 4 Hz, as suggested by Mali et al. [22], to properly estimate Power Spectral Density (PSD), through a parametric autoregressive (AR) model of the order 16 with coefficients determined through the Burg method. A total of 10 frequency-domain features were extracted, high frequency (HF), low frequency (LF) and very low frequency (VLF) bandwidths (Table 2).

The Shimmer sensor provided output galvanic resistance that was converted into galvanic skin conductance (SC). The SC is characterised by startles, of 1–5 s of duration [23] and by a tonic component, related to other uncorrelated sweating activities taking place over a period longer than the one characterising the local peak events. The SC signal frequency content is entirely located within 0–1 Hz [7]. Therefore, the Shimmer sampling frequency was set to 51.2 Hz. The signal was filtered through a 4th-order Butterworth low-pass filter, with a cut-off frequency of 2 Hz. Then, the tonic component was extracted through average process by means of a 5 s-long moving window. By subtracting tonic level from the filtered conductance signal, phasic response was obtained. Finally, an ad hoc algorithm for detection of informative startles was implemented. 16 features were extracted from GSR signal (Table 2).

The MindWave EEG headset mobile device provided a one channel raw EEG signal and information about the PSD in different frequency ranges. Regarding parameters named "Attention" and "Meditation", they are calculated by the device itself starting from the PSD values. They both range from 0 to 100 (see Table 2).

Table 2 Features from HRV, GSR and EEG signals in time (t) and frequency (f) domain

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)

(continued)

Table 2 (continued)

Feature name	Signal	Description
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	EEG(f)	EEG signal power in frequency range 8–9 Hz
Alpha2	EEG(f)	EEG signal power in frequency range 10–12 Hz
Beta1	EEG(f)	EEG signal power in frequency range 13–17 Hz
Beta2	EEG(f)	EEG signal power in frequency range 18–30 Hz
Delta	EEG(f)	EEG signal power in frequency range 1–3 Hz
Gamma1	EEG(f)	EEG signal power in frequency range 31–40 Hz
Gamma2	EEG(f)	EEG signal power in frequency range 41–50 Hz
Theta	EEG(f)	EEG 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"

4.4.2 Feature Selection

Collected features were then normalized. The main motivation behind this procedure was the reduction of inter-subject variability of physiological activity [21, 23]. It was performed in each subject, calculating parameters by using its specific data coming from baseline acquisition. For these reasons, the features related to relaxed and elicited phases were scaled as a relative percentage variation from their correspondent mean baseline value.

Labels ("Relax", "Positive" and "Negative") were attributed to each feature windows to perform the feature selection process. Moreover, it was important to understand which instances could be used as a carrier of reliable information about elicited (positive and negative) states of the subject. The minimal condition for this to be true was that the subject effectively felt the emotional state, better said the mood [24]. Therefore, for the elicited phase, only features vectors coming from time windows during which subjects declared to have felt the emotional state reported in the SAM were withheld and used for further analysis. Indeed, at the end of the protocol, the experimenter asked the subject about the moment when he/she began to feel a positive or negative mood during the protocol, using the 5 fixed reinforcements as references (see Sect. 4.2).

The normal distribution of the computed features was disproved, using the Shapiro-Wilk test of normality. Consequently, the non-parametric Spearman (RHO) correlation coefficient between each feature was calculated. Those couples of features with correlation coefficient below 0.85 (absolute values) and $p < 0.05$ were carefully assessed to decide which ones to exclude for each case [25, 26].

As this work aimed at considering trade-offs between performance and feasibility of a solution in a daily-life application, the correlation analysis was performed, considering each possible combination of sensors used, in couples or singularly. This analysis generated 7 different sets of features to train the upcoming supervised learning processes. They namely were called: BMS (in which Bioharness, Shimmer and Mindwave were considered all together), BM, BS, MS, B, M and S.

4.4.3 Supervised Classification

From this point on, the 7 datasets were analysed in WEKA™ 3.8.2 (University of Waikato, Hamilton, New Zealand) environment. Five different families of algorithms were explored in their performance with each case considered: Decision Tree (C4.5 with default parameters) [27], Rules Induction (RIPPER with 10 folds) [28], Lazy Learners (iBk, with $k = 4$) [29], Probabilistic Learners (Bayesian classifier with default parameters [30]), and function-based classifiers, represented by a SVM (with training data standardization and polynomial kernel) [31] and Multilayer Perceptron (MLP) (7 neurons in the only hidden layer, learning rate of 0.9 and momentum of 0.1) [32].

Each classification process was tested through a Leave-One-Out cross-validation. and its overall accuracy, F measure and ROC area were calculated.

5 Results

The average, standard deviation, max value and min value of pleasure, arousal and dominance scores (integer values ranged from −4 to 4) were calculated for the three emotional components, separately for each condition:

- Positives' pleasure: 2.900, 0.876, 4, 2;
- Positives' arousal: 2.100, 0.994, 4, 1;
- Positives' dominance: 0, 1.764, 2, −2;
- Negatives' pleasure: −1.200, 1.549, 2, −4;
- Negatives' arousal: 1.000, 1.564, 2, −3;
- Negatives' dominance: −0.900, 1.792, 2, −3

Starting from a set of 42 features the results of the correlation analysis led to a subset of 29 features, 12 from HRV (SDNN, HR mean, SDHR, RMSSD, pNN50, LF peak, HF peak, VLF power, LF power, %VLF, %LF, %HF), 9 from GSR (# startle, Amplitude std, Rise rate mean, Rise rate std, Decay rate mean, Decay rate std, Phasic

value mean, Startle time std, Startle RMS std) and 8 from EEG (Alpha1, Alpha2, Beta2, Gamma1, Gamma2, Delta, Attention and Meditation). Since the analysis has showed significant correlation exclusively between parameters deriving from the same signal, to the performed analysis with different sensors combination it was enough to combine in a subset the uncorrelated features extracted from the selected sensors.

The results of the classification algorithms are summarized in Fig. 4. Concerning the classification performances, the function-based algorithms obtained the best results. Indeed, SVM provided the highest metric scores, while the MLP maintained the highest average performances. This suggests that the related family of learning algorithms could be suitable for multi-class problems in the recognition of human moods. On the other side, the Bayesian classifier is the algorithm with the worst performance in recognizing these three moods independently from the combination explored. The other algorithms showed oscillating results, even considering different evaluation metrics of the same case.

The complete configuration of sensors (BMS) obtained the best results in every evaluation metrics. Particularly, as concerns the accuracy, it reached the highest value (89.67%) with SVM. It is worth to underline that, for BMS case, SVM outperformed all other classifiers in each evaluation metric.

As concern the results of the combination of couples of sensors (BM, BS, MS), the BS and BM configurations reached better performance with respect to the MS configuration. Particularly, BM and BS performances were comparable one to each other, while the MS case performed slightly worse, except for the ROC area, in which all three cases scored approximately in the same range (Fig. 4). Mainly, regarding the BS configuration, the accuracy reached the highest value for the SVM (87.07%). The performances of SVM and MLP were very similarly in BMS and BS cases, 89.67% and 87.07% respectively. Table 3 reports the accuracy of the classifiers in recognizing the three moods with the BS configuration. The negative mood was the worst classified, whereas the relax and the positive mood were well classified. Particularly, the recognition of the positive mood reached the best accuracy with the SVM, whereas the recognition of the relaxed status reached his highest performance with the Rules classifier.

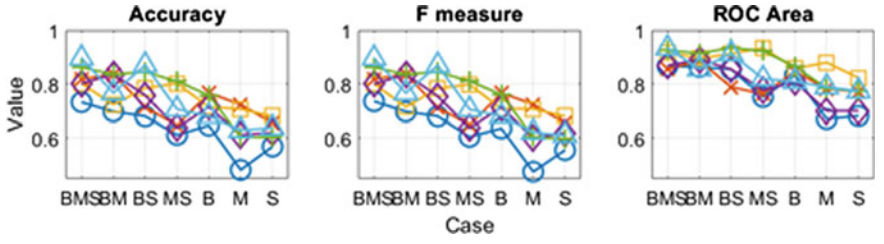

Fig. 4 Accuracy values for each classifies and combination of sensors. In the figures ○ = Bayes, x = Tree, □ = Lazy, ◊ = Rules, + = MLP and Δ = SVM

Table 3 Accuracy (%) of BS configuration in recognizing the three moods

Accuracy	Relax	Positive	Negative
Bayes	77.50	65.20	60.00
Tree	80.00	71.70	60.00
Lazy	85.00	76.10	73.30
Rules	87.50	80.40	53.30
MLP	82.50	89.10	80.00
SVM	80.00	91.30	90.00

Regarding cases with the use of only one sensor (B, S, M), the B case was the most performing one considering all the evaluation metrics, which were always greater than the best ones of M and S cases, for each metric considered. The DT algorithm applied to B features showed the higher accuracy level of single sensor configurations (76.72%).

6 Discussion and Conclusion

The aim of this work was to assess the optimal sensors' configuration to recognize three different moods evoked during a social interaction.

The results obtained from SAM regarding the pleasure average rate (2.90 for positive and −1.20 for negative condition) show that our proposed protocol can evoke in the subject the desired moods. The performances obtained by the classifiers in the case of using only the features extracted from the HRV show that cardiac activity might represent a good standalone informative channel in recognizing mood in real-life applications (76.72% accuracy with DT). However, the present work highlighted that the exploitation of cardiac and skin activities (BS configuration) should be pursued for the final application of recognizing moods of the user inside social environments. According to the results, BS is the configuration which should be selected as the best one to optimize the trade-off between accuracy and obtrusiveness of sensors. Indeed, the accuracy of 87.07% obtained with the SVM applied on HRV and GSR features are significantly higher than the best single sensor case discussed above but does not present significant disadvantages respect to use of the whole system (89.67%).

This leads to think that in the final application the use of only one device able to acquire both HRV and GSR signals would be feasible and performing at the same time. In fact, it is important to state that the ECG waves would not be needed to obtain HRV signal. For instance, photoplethysmography (PPG) would come in handy for measuring the HRV [33, 34]. PPG is a measurement methodology that is applied at the level of fingers. Therefore, as GSR is measured on the fingers, thinking about a device that acquires both the sensing modalities from the same site is a plausible and affordable hypothesis [35].

The obtained results in terms of accuracy are in line with the state of the art (see Sect. 2). But in this case, the result is quite remarkable for the conditions in which it was obtained. In fact, such outcome was achieved monitoring a social interaction, without the usual ideal conditions of emotional experiments. According to the experimental protocol, people could freely move, with no requirement to stay still during the experimental phases, thus introducing noise sources. The positive outcome of this work opens the path to further developments in the way of using physiological activity to monitor affection of people in social environments, which can constitute useful information for a social robotics application, as well.

This work can undergo further developments in each of its aspects. First, other sensing modalities could be integrated to the explored other informative channels and compared with the vision sensor. Secondly, the so-built MIP through social interactions could become increasingly more complex and develop itself in different versions capable of eliciting a higher number of emotional states and covering a wider variety of social contexts in which emotions take place. Third, unsupervised learning methods could be assessed, to overcome the constraints of having instances that must be labelled in advance, main concept behind supervised classifiers. At this point, the behavioural architecture of an intelligent machine, or robot, able to efficiently use the achieved result, should be implemented. The intelligent machine should be finally validated in an experimental framework that will incorporate its presence inside it.

Acknowledgements This work was supported by the ACCRA Project, founded by the European Commission—Horizon 2020 Founding Programme (H2020-SCI-PM14-2016) and National Institute of Information and Communications Technology (NICT) of Japan under grant agreement No. 738251.

References

1. Cavallo F, Semeraro F, Fiorini L, Magyar G, Sinčák P, Dario P (2018) Emotion modelling for social robotics applications: a review. J Bionic Eng 15
2. Isen AM (2000) Some perspectives on positive affect and self-regulation 11:184–187
3. 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
4. Bruno B, Mastrogiovanni F, Sgorbissa A (2013) Functional requirements and design issues for a socially assistive robot for elderly people with mild cognitive impairments. In: 22nd IEEE international symposium on robot and human interactive communication, pp 768–773
5. Koelstra S, Patras I (2013) Fusion of facial expressions and EEG for implicit affective tagging. Image Vis Comput
6. Betti S 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
7. Schumm J, Bächlin M, Setz C, Arnrich B, Roggen D, Tröster G (2008) Effect of movements on the electrodermal response after a startle event. In: Proceedings of 2nd international conference on pervasive computing technologies for healthcare, PervasiveHealth, 2008, pp 315–318
8. Chen M, Ma Y, Song J, Lai C-F, Hu B (2016) Smart clothing: connecting human with clouds and big data for sustainable health monitoring. Mob Netw Appl 21:825–845

9. Nardelli M, Valenza G, Greco A, Lanata A, Scilingo EP (2014) Arousal recognition system based on heartbeat dynamics during auditory elicitation
10. Henriques R, Paiva A, Antunes C (2013) Accessing emotion patterns from affective interactions using electrodermal activity. In: 2013 humaine association conference on affective computing and intelligent interaction, pp 43–48. IEEE
11. Khezri M, Firoozabadi M, Sharafat AR (2015) Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. Comput Methods Programs Biomed
12. Roberts NA, Tsai JL, Coan JA (2007) Emotion elicitation using dyadic interaction tasks. Handbook of emotion elicitation and assessment, pp 106–123
13. Koelstra S, Uhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput 3:18–31
14. Kim J-H, Roberge R, Powell JB, Shafer AB, Williams WJ (2013) Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr BioHarness ™. Int J Sports Med 34:497–501
15. Burns A, Greene BR, McGrath MJ, O'Shea TJ, Kuris B, Ayer SM, Stroiescu F, Cionca V (2010) SHIMMER™—a wireless sensor platform for noninvasive biomedical research. IEEE Sens J 10:1527–1534
16. Salabun W (2014) Processing and spectral analysis of the raw EEG signal from the MindWave. Prz Elektrotechniczny 169–173
17. Harmon-Jones E, Amodio DM, Zinner LR (2007) Social psychological methods of emotion elicitation. Handbook of emotion elicitation and assessment, pp 91–105
18. Bradley M, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25:49–59
19. Kreibig S (2010) Autonomic nervous system activity in emotion: a review. Biol Psychol 3:394–421
20. Lippman N, Stein KM, Lerman BB (1994) Comparison of methods for removal of ectopy in measurement of heart rate variability. Am J Physiol 267:H411–H418
21. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS (2006) Heart rate variability: a review. Med Biol Eng Comput 44:1031–1051
22. Mali B, Zulj S, Magjarevic R, Miklavcic D, Jarm T (2014) Matlab-based tool for ECG and HRV analysis. Biomed Signal Process Control 10:108–116
23. Boucsein W (2012) Electrodermal activity. Springer Science + Business Media, LLC
24. Lochner K, Eid M (2016) Successful emotions: how emotions drive cognitive performance. Successful emotions: how emotions drive cognitive performance, pp 43–67
25. Wang H-M, Huang S-C (2012) SDNN/RMSSD as a surrogate for LF/HF: a revised investigation. Model Simul Eng 2012:1–8
26. Kao F-C, Wang SR, Chang Y-J (2015) Brainwaves Analysis of Positive and Negative Emotions. WSEAS Trans. Inf. Sci. Appl. 12:200–208
27. Webb GI (1999) Decision tree grafting from the all-tests-but-one partition. IJCAI Int J Conf Artif Intell 2:702–707
28. Cohen WW (1995) Fast effective rule induction. In: Machine learning: proceedings of twelth international conference
29. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn 6:37–66
30. John GH, Langley P (1995) Estimating continuous distributions in Bayesian classifiers. In: Eleventh conference on uncertainty in artificial intelligence, pp 338–345
31. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK (2001) Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput 13:637–649
32. White BW, Rosenblatt F (1963) Principles of neurodynamics: perceptrons and the theory of brain mechanisms. Am J Psychol 76:705
33. Tamura T, Maeda Y, Sekine M, Yoshida M (2014) Wearable photoplethysmographic sensors—-past and present. Electronics 3:282–302

34. Lee J, Matsumura K, Yamakoshi KI, Rolfe P, Tanaka S, Yamakoshi T (2013) Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. In: Annual international conference of the IEEE engineering in medicine and biology society EMBS, pp 1724–1727
35. Esposito D, Cavallo F (2015) Preliminary design issues for inertial rings in ambient assisted living applications. In: 2015 IEEE international instrumentation and measurement technology conference (I2MTC) proceedings, pp 250–255. IEEE