Mood evaluation by Remote-PPG and facial expression recognition

Andrea Manni [1], Andrea Caroppo [1], Pietro Siciliano[1] and Alessandro Leone [1]

[1] *National Research Council of Italy, Institute for Microelectronics and Microsystems, Via per Monteroni c/o Campus Universitario Palazzina A3, Lecce, Italy*

Abstract

Psychological health monitoring plays an important role in mood evaluation, especially of ageing subjects within the home environment. For this purpose, the development of innovative and easy-to use platforms based on the use of contact or contactless smart sensor is spreading widely. This paper presents the design and the implementation of a novel framework able to evaluate the mood combining vital signs and facial expressions. For this purpose, a low-cost and commercial vision sensor is used to allow a wider diffusion of the proposed solution and with the aim of increasing the acceptability of the proposed solution. Specifically, a heart rate estimation algorithm and a facial expression recognition module are combined to evaluate the end user's mood. This result has been achieved through use of deep learning and transfer learning algorithms that work in real time also on embedded hardware platform not equipped with GPUs, consequently increasing its usability. The first added value of the proposed framework consists in the possibility of detecting facial expressions "in the wild" independently from the selected vision sensor and from face orientation. Another important added value lies in the implementation of rule-based expert system which combines data acquired from the same smart sensor but whose operation is also maintained using information from heterogeneous sensors that provide the same type of discrete input values. Due to COVID-19 restrictions, the overall system is currently being tested first in a controlled environment and then in a real environment to achieve the final goal. The findings of the preliminary experiments show promising results for heart rate and facial expression monitoring with a low average error expressed in terms of Root Mean Square Error for HR estimation and high accuracy regarding facial expression recognition.

Keywords

Mood Evaluation, Contactless Sensors, Ageing Subjects, Deep Learning, Transfer Learning, Heart Rate Monitoring, Facial Expression Recognition.

1. Introduction

Mood generally reflects a person's mental state but can also have a significant relationship with the physical health [1]. Consequently, his assessment could have a very important impact on daily life and work. For example, it is well known that a negative mood is a key factor that influences human health and different studies demonstrated that a negative mood over a long period of time can contribute to various health problems such as depression or heart disease [2].

Generally, there are two ways to evaluate the mood. The first one is based on the estimation of emotional behavior of the person/patient through the analysis of facial expressions patterns. Moreover, the mood can be evaluated analyzing the physiological signals of the observed subject, such as heart rate (HR), HR variability, electrocardiogram (ECG), and electroencephalogram (EEG).

Italian Workshop on Artificial Intelligence for an Ageing Society (AIxAS 2021), November 29th, 2021

EMAIL: andrea.manni@le.imm.cnr.it (A. 1); andrea.caroppo@cnr.it (A. 2); pietro.siciliano@le.imm.cnr.it (A. 3); pietro.siciliano@le.imm.cnr.it (A. 3)

CEUR Workshop Proceedings (CEUR-WS.org)

The evaluation of the mood turns out to be an important information especially for monitoring the health status of ageing subjects and/or frailty subjects. In this context, an enormous advantage is obtained by implementing HW/SW platforms that are implemented through devices (possibly commercial) such as to increase the degree of acceptability of the proposed solution. Various experiments have shown that wearable devices are poorly accepted by elderly subjects for monitoring, for example, vital signs. Furthermore, another disadvantage arises from the fact that such devices can be unworn due to an oversight, compromising any long-term analysis. Consequently, many scientific studies have turned towards the extraction of features for the evaluation of the mood using non-contact, minimally invasive sensors, increasing the degree of acceptability of the proposed solution.

From the analysis of the most recent scientific literature in the sector it is evident how facial expressions are the most widely employed modality for the evaluation of mood. For example, in [3] a stationary wavelet transform is used to extract features for facial expression recognition and the selected features are then passed into a feed-forward neural network that is trained through a back-propagation algorithm. Moreover, in [4] a hybrid feature descriptor-based method is proposed to recognize human emotions from their facial expressions.

Previous research showed that even HR is a good indicator for the evaluation of mood since it was demonstrated that HR fluctuates with mood changes. In [5] an experiment showed that physiological signals have unique responses to different emotions. For example, HR increased significantly when people were angry or fearful, but decreased substantially during disgust. In [6] it is demonstrated that HR during a positive mood was lower than during a neutral mood. The authors of [7] in their study showed that the effects of relaxation and fear on HR were significantly different, and the average HR during happiness was lower than in a sad state.

As described above, both facial expressions and HR help in mood assessment. In recent years, with the more advanced process of fusing information from different sources, it has become possible to merge the features of reference emotional states. To automatically recognize emotions, many works have proposed fusion with audiovisual information, e.g., combining speech with facial expressions. In [8] a database (emoF-BVP) is presented consisting of various audio and video recordings of actors expressing various intensities of different emotional expressions. Then, four deep belief network (DBN) models are presented which allow the generation of robust multimodal features for emotion classification in an unsupervised manner.

Some studies have investigated the combination of EEG and physiological signals. For instance, in [9] the database "Dataset for Emotion Analysis using Physiological signals" is used for the classification of emotions. The aim is to determine which of the physiological and EEG signals are most relevant for emotion recognition.

In this paper, an algorithmic pipeline that fuse HR values and facial expressions for the evaluation of mood is presented. The mood is evaluated by combining HR and facial expressions using a rule-based expert system. This system automatically detects the mood of the observed subject integrating a low-cost and commercial camera, a face detection algorithmic module based on Deep Learning (DL) and a Facial Expression Recognition (FER) module based on the concept of Transfer Learning (TL) algorithms. Moreover, the pipeline integrates a contactless HR detection module. A first important added value of the proposed pipeline is to be found in its running in real-time on a non-GPU embedded hardware platform. Another important advantage is that the algorithmic pipeline is independent from the image capture device and from the face orientation of the observed subject, achieving more accurate prediction of HR and facial expression of the end-user in a typical Ambient Assisted Living (AAL) context.

The remainder of this paper is organized as follows. Section 2 explains our proposed algorithmic pipeline and provides an overview of the methodology by detailing the algorithmic step implemented for HR, facial expression estimation and mood evaluation. The results obtained are reported in Section 3. Finally, Section 4 shows our conclusions and discussions on some ideas for future work.

2. Method

An overview of the proposed framework via a block diagram representation is depicted in Figure 1.

The algorithmic pipeline has as input an image acquired by a commercial vision sensor in the RGB color space. It consists of four main blocks: 1) a pre-processing step integrating a face-detection module and a series of algorithmic steps useful to format the data for the following steps, 2) a HR estimation module based on a specific Region of Interest (ROI) extraction, filtering, detrending and Fast Fourier Transform (FFT) of the signals obtained, 3) a FER module based on a pre-trained deep-learning model, 4) a final software module for the evaluation of the end-user mood. In the last main block, the estimated value of HR and the corresponding facial expression are sent as input to an expert system that returns the patient's mood using established rules. Each component of the pipeline is detailed in the following.

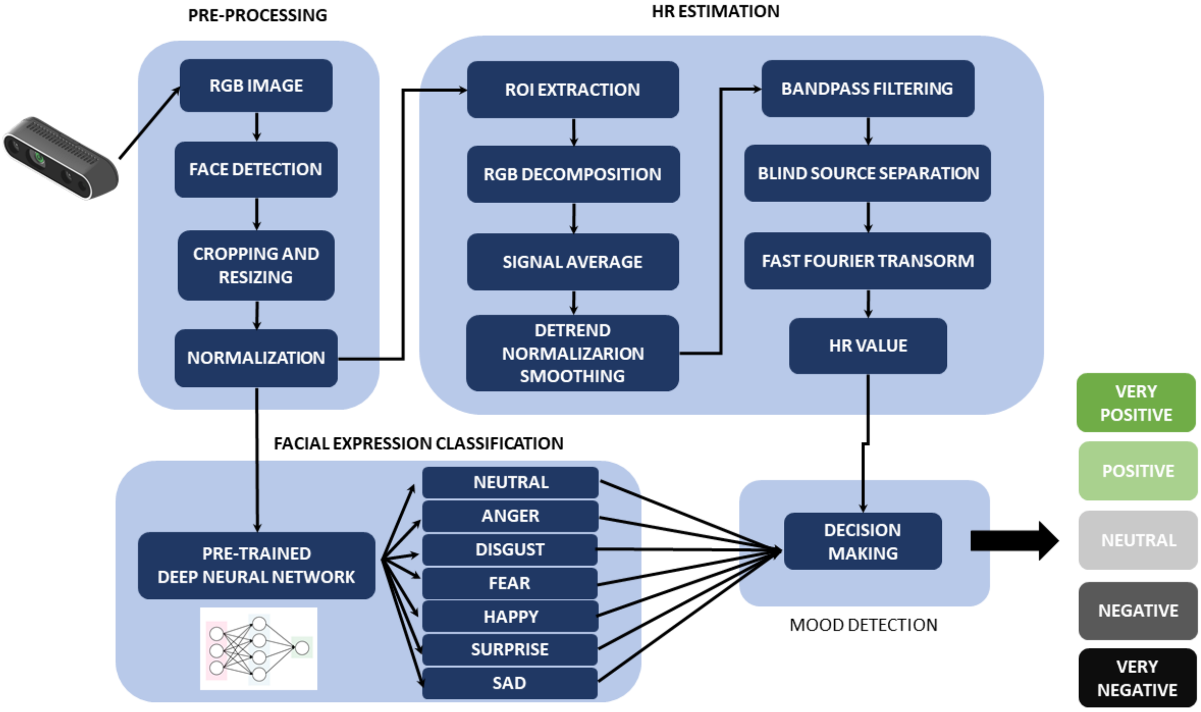


Figure 1: Block diagram of the proposed algorithmic pipeline designed and implemented for mood evaluation.

2.1. Pre-Processing

The first main block of the pipeline implements the pre-processing algorithms on the initially acquired image. One of the most important algorithmic step involves the detection of the facial region in the streaming video captured by the commercial vision sensor. It is important to underline that this block is in common within the pipeline both for the estimation of the heartbeat and for the recognition of the facial expression of the observed subject.

To obtain an accurate real-time face detection, the latest version of OpenCV library is used, where a deep neural network (DNN) architecture is included. This module allows face detection "in the wild" in real-time on a PC without GPU. Moreover, this approach allows face detection even in less-than-ideal lighting conditions. The module uses a reduced ResNet-10 model [10] and its output is the bounding-box coordinates of the facial region together with a confidence index. After face detection, to crop only the facial region, a software procedure is implemented extracting only the coordinates of the upper left corner, the height, and the width of the face, thus removing all the information not related to the face. In addition, both a down-sampling step and an increasing resolution step are added, depending on the resolution of the facial image.

Specifically, a simple linear interpolation was used for down-sampling, while a nearest-neighbor interpolation was implemented to increase the size of the facial images. At this point, a "normalization" step is added to stabilize the contrast and brightness of the image. Here, normalization was performed through the application of "contrast-limited adaptive histogram equalization" (CLAHE) [11].

2.2. Heart Rate Estimation

The gold standard techniques for measuring HR such as ECG and photoplethysmography (PPG) require skin contact and can inevitably cause discomfort, especially in the current pandemic period. Recently, remote photoplethysmography (rPPG) has obtained an increasing attention because it allows measurement of HR in a contactless way.

In this section, heart rate estimation block is explained. After the application of pre-processing steps, given the face identified by the landmarks, regions of interest (ROIs) corresponding to a region with a strong blood modulation transition are identified. As described in [12] and other HR estimation research, the forehead and the cheeks resulted to be the most suitable regions for the purpose since the strength of PPG signals differs between different regions of the face, with the cheek and forehead regions tending to produce the strongest PPG signals [13]. Due to forehead occlusion depending on the hair style, in the present version of the framework the validation of the HR measures was carried out considering only the cheeks. To identify the ROIs, a shape predictor (with 68 landmarks) that includes the face is used (Figure 2a). Then, to identify cheeks, the areas delimited by landmarks 1-29-34-4 (right cheek) and 17-29-34-14 (left cheek) are considered (Figure 2b).

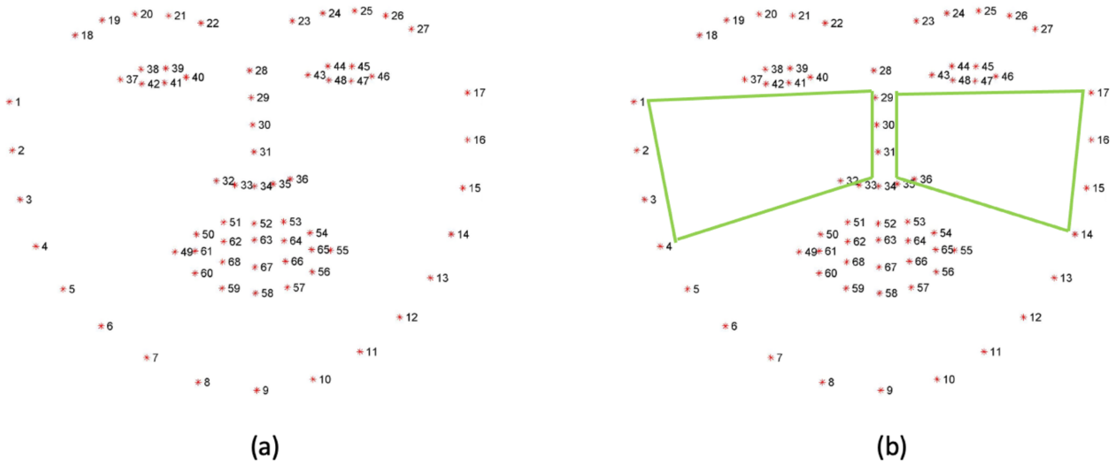


Figure 2: ROIs identification: (a) shape predictor with 68 landmarks and (b) left and right cheeks identification.

The next step is to filter the obtained raw RGB signals to remove frequencies that are not realistic for a human heart. Accordingly, a band-pass filter with ideal behavior is applied to remove high- and low-frequency noise. The filter removes components occurring outside the frequency band [0.65, 3] Hz, which has been commonly used in the literature and corresponds to the HR between 40 and 180 bpm.

Identification of cyclic components in signals is achieved using the power spectrum, but sufficient signal quality and length are required. However, all parameters - including the periodic component of the signal - vary and the HR signal is limited to a specific time interval. Furthermore, as the noise spectrum is similar to the spectrum of the signal to be recovered, filtering to increase the signal-to-noise ratio (S/N) is critical. In addition, the three raw RGB signals obtained can be decomposed into basic source signals using Blind Source Separation (BSS) algorithms. Since raw RGB signals contain HR information in mixed components, in order to extract the source signals from these mixed signals, independent component analysis (ICA) is used. ICA is a BSS since it calculates a linear sum W of the available data sets y (raw RGB color channels) with weights w in order to maximize one independent source at a time. The data sources x must not have a Gaussian distribution and there may only be linear mixes M of these unknown sources. Thus, the mixed data sets and the original independent components can be expressed by Equations 1 and 2:

$$y(t) = \boldsymbol{M}x(t) \tag{1}$$

$$x(t) = \boldsymbol{W}y(t) \tag{2}$$

where $\boldsymbol{M}$ is the unknown mixing matrix, and its inverse $\boldsymbol{W}$ is the de-mixing matrix found by the ICA. Several ICA algorithms are available. In this proposed system, FastICA [14] method is used to analyze the RGB signals and to reveal the original source signals removing noise artifacts. A fourth order moment (Kurtosis) is used to identify the independent components (three components in the proposed approach). Although there is no ordering of the ICA components, the second component typically contained a strong plethysmography signal and consequently, for the sake of simplicity and automation, is selected here as the desired source signal. Finally, Fast Fourier Transform (FFT) is applied on the selected component to obtain the power spectrum inside the frequency band [0.65, 3] Hz matching to [40, 180] bpm. The peak of the power spectrum in the given range represents the pulse frequency (Figure 3).

Then, to improve the performance of the system, HR values are collected in a time window, fixed in the actual version of the module in 30 seconds of length. Then, in order to reject possible artifacts, the outliers are removed, and a single value is obtained by calculating the median value of the residual components in the time window.

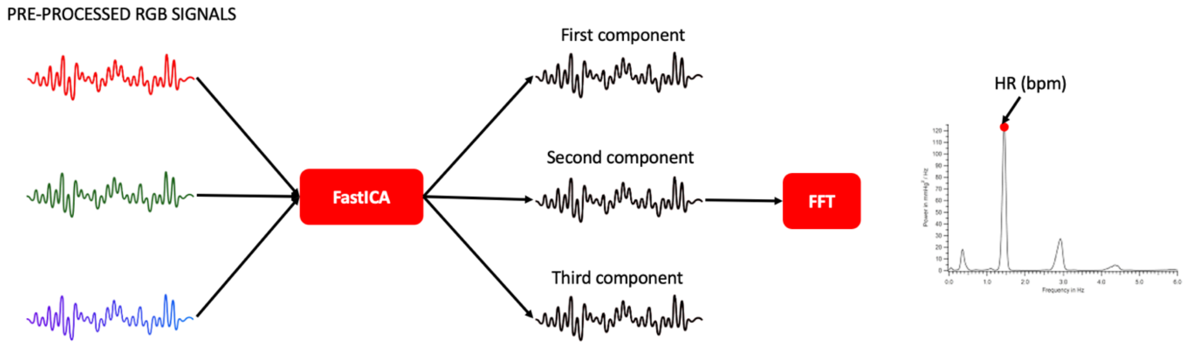


Figure 3: Feature extraction for contactless HR measurements. Three pre-processed RGB signals are extracted from ROI and subsequently filtered. FastICA is applied on the normalized, de-trended and smoothed RGB signals to recover three independent source signals. Finally, FFT is applied to the second component and the highest power of the spectrum is selected as the estimated HR.

2.3. Facial Expression Recognition

The FER theory consists of extracting a small number of basic emotions such as happiness, sadness, neutrality, surprise, anger, disgust, and fear. In addition to these expressions, neutral facial expression is also estimated in some studies.

The performance of a FER method is based on the use of the most discriminating features. There are two main categories of features methods in this research area: hand-crafted features extraction and automatic feature extraction. The first ones are frequently used as geometric or appearance descriptors (such as Scale-Invariant Feature Transform, Local Binary Pattern, …). The second category is more recent and focuses on features generated automatically by a DL architecture. From the state-of-the-art algorithms, CNNs [15] have worked very well for the FER problem in unconstrained scenarios. CNNs allow to process data having a grid pattern, such as images, by learning spatial hierarchies of features (from low to high-level patterns) automatically. The most important problem regarding the use of CNN for FER is the availability of facial expression datasets with a very high number of labelled images, since training DL architectures with a limited number of images can lead to the problem of overfitting. To address this problem, one of the solutions used is to evaluate the concept of TL [16]. TL is based on training a specific network on a small dataset. The network is first subjected to a pre-training phase on an extremely large dataset and then applied to the given task of interest.

In this work, TL is used for the FER task. The famous architecture VGG16 [17] is pre-trained using the Facial Emotion Recognition 2013 (FER-2013) dataset. FER-2013 was introduced in the ICML 2013 workshop's facial expression recognition challenge. The dataset is quite challenging, since faces greatly vary in age, pose and occlusion conditions [18].

Generally, VGG16 architecture is structured into two main and well separated sections: feature extraction and classification. In our work, considering that the feature extraction section is used for extracting new dataset features, the classification section originally structured with 3-FC layers (named FC6, FC7 and FC8) is replaced with a novel FC layer useful for tuning the desired output, i.e., the number of facial expression classes (seven classes in our case, six different facial expressions plus the neutral expression). The steps to design the proposed new architecture are shown in Figure 4.

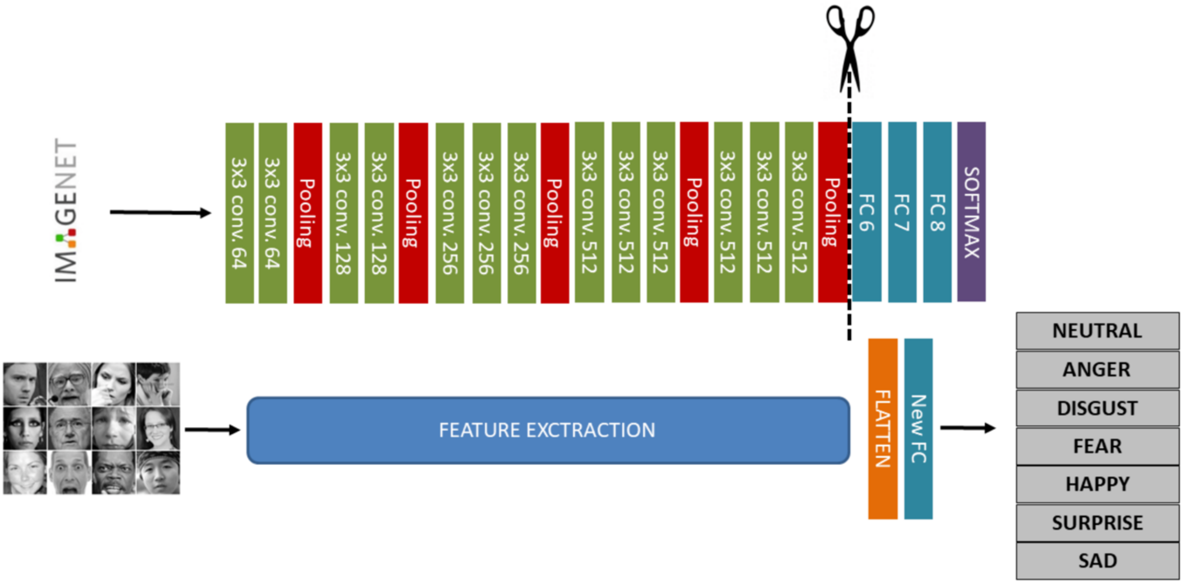


Figure 4: VGG-16 Deep Convolutional Neural Network (DCNN) architecture trained on ImageNet database. In the top row the network has 16 layers and can classify images into 1000 object categories; in the bottom row transfer-learning application for FER replacing classification layers of original VGG16 architecture is shown.

From the above schematic representation, it can be seen that the feature extraction part is the same as the origin of the VGG16 architecture, but a new FC layer is added to adjust the number of outputs with the number of new classes in the dataset, i.e., 7. In addition, there is a flatten layer between the feature extraction and the new FC layer whose function is to change the size of the input tensor of the previous layer and ensure that the size of the output is a 1 × 1 tensor with a length corresponding to the input tensor volume.

Then, VGG16 is trained with the stochastic gradient descent algorithm [19], estimating the error gradient for the current state of the model using examples from the training dataset and then updating the model weights using the backpropagation of error algorithm. The classifier layer of the transfer-learned model can classify seven classes of expressions: "Anger", "Disgust", "Fear", "Happy", "Sad", "Surprise", and "Neutral". In the present work, the deep CNN features were evaluated on three different machine learning classifiers that have shown promising results in previous FER studies, such as Support Vector Machine (SVM), Logistic Regression (LR), and k-nearest neighbors (kNN).

SVM separates categorical data in a high dimension space finding a hyperplane with the maximum possible margin between the same hyperplane and the cases [20]. LR is a predictive analysis tool and describes the relationship between one dependent binary variable and one/multiple independent variables. In LR, the dependent variable is binary, in contrast with linear regression having continuous dependent variable. In this work, for LR we set only the parameter C (that is, the inverse of regularization strength λ) to 0.01 [21]. At last, kNN calculates, in a non-parametric way, the distances between the nearest k training cases and an unclassified case, and classifies the latter to the highest of the nearest k training cases. Different distance metrics can be applied in an experimental stage, with the most widely used that are the Euclidean Distance and the Manhattan distance. Here, k value was set to 2 and the Manhattan distance was used as a distance function [22].

This module returns a facial expression label with a sampling time of one second. Consequently, to make the whole system suitable for the analysis of video sequences, a decision strategy based on the temporal consistency of the FER results is introduced. Facial expression is taken by analyzing a time window of the same size of the HR estimation and checking which facial expression associated with a confidence index greater than 0.8 is most prevalent in the window.

2.4. Mood Evaluation

In this module, a decision strategy based on production rules, usually used as a simple expert system in artificial intelligence, is implemented. Specifically, HR values and detected facial expression are combined for mood evaluation. A production rule is composed of an IF part and a THEN part, turning out to be:

$$P_i: IF\ \mathrm{X}\ THEN\ \mathrm{Y}, \tag{3}$$

where P_i represents the rule i, X is the antecedent of the rule i, and Y is the consequent. Here, X is composed of (x_1, x_2) where x_1 represents the HR value and x_2 represents the detected facial expression. Rules are activated when their conditions are satisfied.

Table 1 reports a sample of used rules. For instance, if the HR value is less than 70 bpm and the facial expression is “happy”, then the output of mood evaluation module is: VERY POSITIVE (Rule 1 in Table 1).

Table 1
Extract of the rules.

Rule No.	Antecedent		Consequent
	HR	FER	
1	<= 70	Happy	Very positive
2	> 70 and <= 90	Happy	Positive
3	< 90	Sad	Neutral
…	…	…	…
16	>= 90	Anger	Very negative

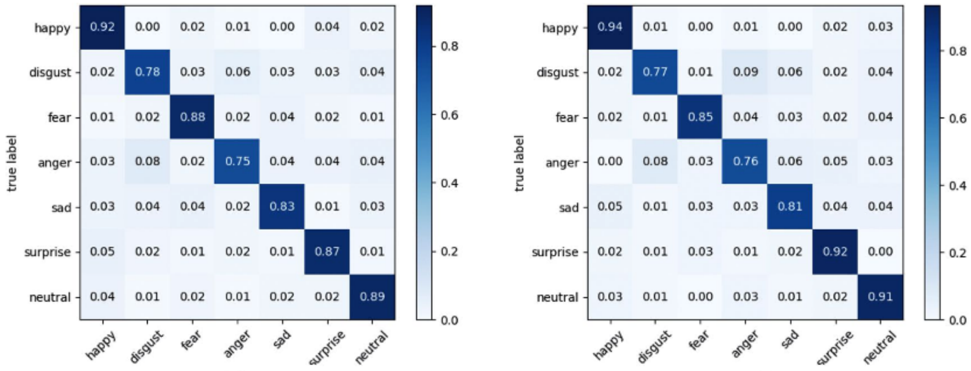
3. Results and discussion

Currently, due to COVID-19 restrictions, only the HR estimation module and FER pipeline blocks have been tested. The validation was conducted in the laboratory of the Institute of Microelectronics and Microsystems (IMM) in Lecce, Italy. The experimental setup consisted of an embedded PC with Intel core i7 and 8GB of RAM, using Python (3.7) language with OpenCV for algorithm development. The Intel RealSense^{TM} D435 camera was used for image streaming acquisition. A total of 15 participants (nine males and six females) with ages ranging from 35 to 69 years were included in this study after giving their voluntary consent. For HR estimation, the root mean squared error (RMSE) is proposed for evaluating the accuracy of HR measurements, considering a commercial pulse oximeter as ground truth. The experiments were run by varying head poses, lighting conditions, and distance from the vision sensor (from 0.5m to 2m). For the sake of brevity, Table reports the RMSE obtained at varying of head poses (ranging from -40° to +40° for yaw angle and -20° and +20° for pitch angle) and lighting conditions (in the range 30-100 lumens) at a fixed distance from the vision sensor (0.5 mt.).

Table 2
RMSE at varying of light conditions (30 and 100 Lux), yaw angle (between -40° and +40°) and pitch angle (between -20° and +20°).

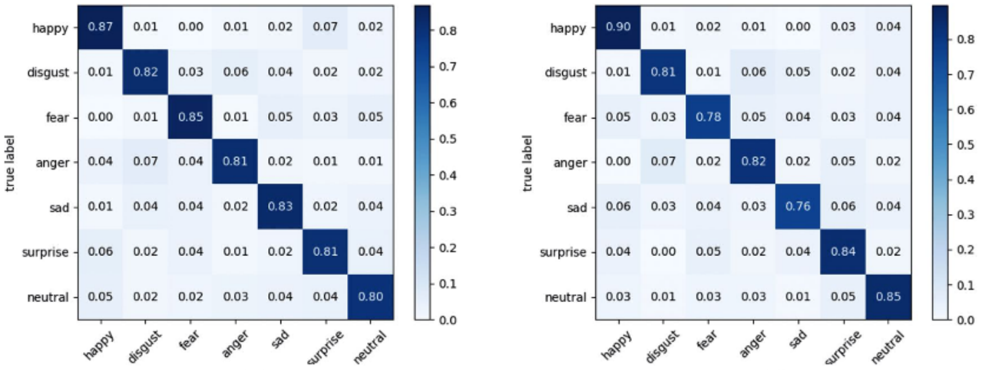
Lx	30					100				
Yaw Angle	-40	-20	0	+20	+40	-40	-20	0	+20	+40
RMSE mean	5.72	4.49	2.41	4.36	5.80	4.86	2.45	1.97	2.78	4.35
Pitch Angle		-20	0	-20			-20	0	-20	
RMSE mean	/	2.56	1.87	2.43	/	/	2.35	1.57	2.25	/

These results show that the implemented approach allows effective HR classification even in the presence of significant changes in head pose and lighting conditions, with the RMSE increasing slightly as lighting intensity decreases. For FER module, each user involved in the trial simulated in sequence the classical six facial expressions plus the neutral expression. The performance of FER module was evaluated using accuracy as metric. Accuracy (Acc) is the overall classification in term of True Positive (TP) and True Negative (TN) of the proposed method. In Figure 5, Figure ***6***, Figure ***7*** the confusion matrices of the average accuracies obtained using the considered classifiers (i.e. SVM, LR and kNN) are reported. The accuracies were calculated by averaging the accuracies obtained by varying lighting conditions and face orientation.



(a) **(b)**

Figure 5: Confusion matrices for seven classes of facial expressions using SVM as classifier and at varying (a) yaw angles and (b) pitch angle.



(a) **(b)**

Figure 6: Confusion matrices for seven classes of facial expressions using LR as classifier and at varying (a) yaw angles and (b) pitch angle.

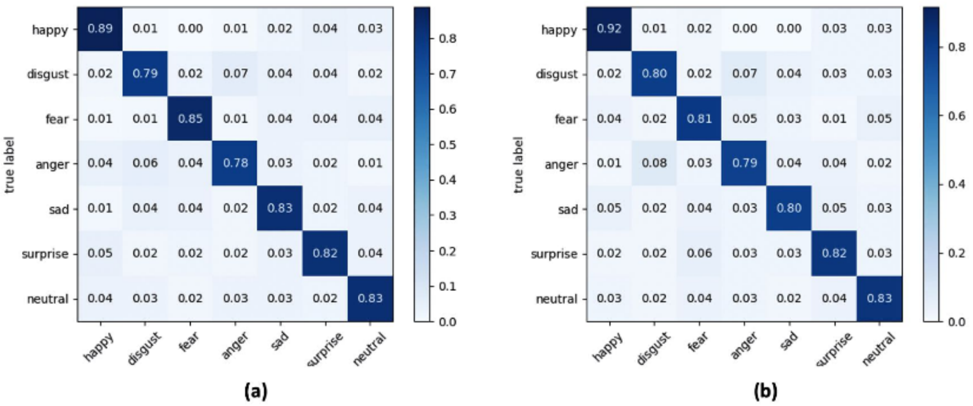


Figure 7: Confusion matrices for seven classes of facial expressions using kNN as classifier and at varying (a) yaw angles and (b) pitch angle.

From the confusion matrices, it can be seen that the SVM classifier achieves the best performance in the efficient recognition of the considered facial expressions varying light conditions the pitch and yaw angle. More specifically, SVM obtains an improvement in the classification of about 5.8% compared to the LR classifier and of about 4.6% compared to the kNN classifier with lux = 100, while the performance in terms of accuracy with lux = 500 is less evident (+2.3% compared to LR and +2% compared to kNN).

4. Conclusion

In this work, a novel algorithmic pipeline for mood evaluation starting from HR estimation and FER was proposed. The hardware platform returns the HR and facial expression of the subject in real time using the same input information, i.e., the facial region. Moreover, the platform implements a software module capable of evaluating the "mood" through temporal combinations of the information previously extracted. An added value lies in the fact that to capture the substantial visual features of HR and facial expressions from the face, a series of algorithmic steps were designed and implemented considering various head poses, distances from the sensor, and variations in lighting conditions. Thanks to these algorithmic steps, the entire pipeline allows greater usability of the proposed solution, integrating perfectly into a AAL environment where generally fragile or ageing subjects are present.

From the performance point of view, the algorithmic pipeline achieved satisfactory results in terms of RMSE for HR estimation and accuracy for FER in the wild.

The next development of the proposed work will be the test of the introduced production rules with the purpose to distinguish at least positive, negative, and neutral mood of the observed subject. A further development of this work will involve the extraction of the breathing rate from the same facial region, and the combination within the rules of this information, to provide an output mood that is as close to reality as possible.

5. Acknowledgements

This work has been carried out within the project PON Si-Robotics funded by MUR-Italian Ministry for University and Research.

6. References

[1] K. Gouizi, C. Maaoui, F. B. Reguig, Negative emotion detection using EMG signal, in: Proceedings of the 2014 International Conference on Control, Decision and Information Technologies (CoDIT), pp. 690-695, Metz, France (3-5 November 2014).

[2] L. D. Kubzansky and I. Kawachi. “Going to the heart of the matter: do negative emotions cause coronary heart disease?”, J Psychosom Res., A, 48 (4-5) (2000).

[3] H. Qayyum, M. Majid, S. M. Anwar and B. Khan. “Facial Expression Recognition Using Stationary Wavelet Transform Features”, Mathematical Problems in Engineering, vol. 2017 (2017).

[4] T. Kalsum, S. M. Anwar, M. Majid, B. Khan and S. M. Ali. “Emotion recognition from facial expressions using hybrid feature descriptors”, IET Image Process., 12, pp. 1004-1012 (2018).

[5] P. Ekman. “An argument for basic emotions”, Cognition and Emotion, 6(3-4), pp. 169-200 (1992).

[6] A. Britton, M. Shipley, M. Malik, K. Hnatkova, H. Hemingway and M. Marmot. “Changes in Heart Rate and Heart Rate Variability Over Time in Middle-Aged Men and Women in the General Population (from the Whitehall II Cohort Study)”, The American Journal of Cardiology, 100(3), pp. 524-527 (2007).

[7] M. T. Valderas, J. Bolea, P. Laguna, M. Vallverdú and R. Bailón, Human emotion recognition using heart rate variability analysis with spectral bands based on respiration, in: Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6134-6137, Milan, Italy (25-29 August 2015).

[8] H. Ranganathan, S. Chakraborty and S. Panchanathan, Multimodal emotion recognition using deep learning architectures, in: Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-9, Lake Placid, NY, USA (7-10 March 2016).

[9] C. A. Torres-Valencia, H. F. García-Arias, M. A. Álvarez López and A. A. Orozco-Gutiérrez, Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models, in: Proceedings of the 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, pp. 1-5, Armenia, Colombia, (2014).

[10] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[11] A. M. Reza. “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement”, Journal of VLSI signal processing systems for signal, image and video technology, 38(1), pp. 35-44 (2004).

[12] A. Challoner. “Photoelectric plethysmography for estimating cutaneous blood flow”, Non-Invasive Physiol. Meas. 1979, 1, pp. 125-151

[13] M. Kumar, A. Veeraraghavan and A. Sabharwal. “DistancePPG: Robust non-contact vital signs monitoring using a camera”, Biomedical optics express, 6(5), 1565-1588. (2015)

[14] A. Hyvarinen. “Fast and robust fixed-point algorithms for independent component analysis”, IEEE transactions on Neural Networks, 10(3), 626-634 (1999).

[15] D. H. Hubel and T. N. Wiesel. “Receptive fields and functional architecture of monkey striate cortex”, J. Physiol. 1968, 195, pp. 215-243.

[16] K. Weiss, T. M. Khoshgoftaar and D. Wang. “A survey of transfer learning”, J. Big Data 2016, 3, pp. 1-40.

[17] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556* (2014).

[18] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner and Y. Zhou, Challenges in representation learning: A report on three machine learning contests. in: International Conference on Neural Information Processing, Springer, Heidelberg, 2013, pp. 117–124.

[19] L. Bottou. ”Stochastic gradient descent tricks”, Neural networks: Tricks of the trade, Springer, Berlin, Heidelberg, 421-436. (2012)

[20] J. A. Suykens and J. Vandewalle. “Least squares support vector machine classifiers”, Neural Process. Lett. 1999, 9, pp. 293-300.

[21] D. W. Hosmer, S. Lemeshow Jr. and R. X. Sturdivant. “Applied Logistic Regression”, John Wiley & Sons: Hoboken, NJ, USA, 2013, Volume 398.

[22] S. A. Dudani. “The distance-weighted k-nearest-neighbor rule”, IEEE Transactions on Systems, Man, and Cybernetics, IEEE: New York, NY, USA, 1976, pp. 325-327.