2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Emotion recognition from multimodal physiological measurements based on an interpretable feature selection method

Edoardo Maria Polo[1], Maximiliano Mollura[2], Marta Lenatti[3], Marco Zanet[3], Alessia Paglialonga[3] and Riccardo Barbieri[2]

Abstract— **Many studies in literature successfully use classification algorithms to classify emotions by means of physiological signals. However, there are still important limitations in interpretability of the results, i.e. lack of feature specific characterizations for each emotional state. To this extent, our study proposes a feature selection method that allows to determine the most informative subset of features extracted from physiological signals by maintaining their original dimensional space. Results show that features from the galvanic skin response are confirmed to be relevant in separating the arousal dimension, especially fear from happiness and relaxation. Furthermore, the average and the median value of the galvanic skin response signal together with the ratio between SD1 and SD2 from the Poincarè analysis of the electrocardiogram signal, were found to be the most important features for the discrimination along the valence dimension. A Linear Discriminant Analysis model using the first ten features sorted by importance, as defined by their ability to discriminate emotions with a bivariate approach, led to a three-class test accuracy in discriminating happiness, relaxation and fear equal to 72%, 67% and 89% respectively.**

Clinical relevance **This study demonstrates the ability of physiological signals to assess the emotional state of different subjects, by providing a fast and efficient method to select most important indexes from the autonomic nervous system. The approach has high clinical relevance as it could be extended to assess other emotional states (e.g. stress and pain) characterizing pathological states such as post traumatic stress disorder and depression.**

I. INTRODUCTION

Emotions are very important in the life of human beings, as they ensure survival and reproduction through adaptation to the environment [1]. Emotions are a very complex network of neuronal and hormonal interactions which strongly influence decision-making processes [2].

Among the different systems of classification of emotions, two are the most common: the categorical system and the dimensional system. In the categorical system, emotions are classified as discrete entities, independent of each other and easily distinguishable. Commonly, six fundamental emotions are identified (anger, fear, disgust, joy, surprise, sadness) and of these emotions both facial expressions and their respective interpretations are found to be universal throughout humanity [3]. The poor congruence of the data coming from the researches and the considerable difficulties of researchers and clinicians in identifying and describing emotions as categorical [4], have led several authors to develop a less rigid model of emotions, far from a categorical taxonomic system. The human being, in fact, does not express emotions as specific and discrete entities, but rather as ambiguous and blurred experiences, often overlapping with each other. Emotions are not clearly distinguishable and separable from each other. In the last decade, many studies have adopted a bidimensional approach to quantify emotions on a numerical scale [5][6] thus facilitating their identification and characterization. In this model, each emotion can be explained as the linear combination of two specific and independent dimensions, valence and arousal, by defining the circumflex model of emotions. Valence (positive or negative) is able to classify emotions along a continuum of pleasantness-unpleasantness. Arousal, on the other hand, indicates the intensity of the emotion in terms of physiological activation (low or high). According to the dimensional model, the combination of these two dimensions, associated with the physiological response resulting from stimulation and cognitive attribution, is hypothesized to give rise to the subjective emotional sensation.

Many studies use dimensionality reduction/transformation methods, such as principal components analysis (PCA) and Fisher projection, in combination with a classifier [7][8]. Consequently, a new subset of transformed features is then used for the classification rather then original features. However, despite a high model performance can be reached, the importance and meaning of physiological values resulting from using the original feature set is lost and the resulting models loose employability and explanability in characterizing physiological variations due to different emotions.

In this study, we propose a methodological and interpretable approach to select the most relevant features from the original subset by maintaining their original values and we provide a classification algorithm to recognize happiness, relaxation and fear. Specifically, we use the electrocardiogram (ECG), the blood volume pulse (BVP) and the galvanic skin response (GSR) to create the classification algorithm for emotion recognition.

II. METHODS

A. Study Design and Data

From the Continuosly Annotated Signals of Emotion (CASE) database [9], we considered a subset of physiological signals: ECG, BVP, and GSR, recorded from 30

[1]Edoardo Maria Polo is with the Department of Automatic Control, Bioengineering and Operations Research, Università La Sapienza, Rome, Italy polo@diag.uniroma1.it

[2]Maximiliano Mollura and Riccardo Barbieri are with Department of Electronic, Information and Bioengineering, Politecnico di Milano, Milan, Italy

[3]Marta Lenatti, Marco Zanet, and Alessia Paglialonga are with the National Research Council of Italy (CNR), Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Milan, Italy

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

subjects subjected to video-stimuli. Annotations as well as subjective ratings of the videos are provided together with the dataset for each video. From the initial four emotions (happiness, relaxation, boredom, fear) provided in the dataset, we discarded *boredom*, since 70% of the subjective ratings of videos annotated as boring were associated to the other emotions, especially to *relaxation*. The three different emotions were elicited by showing subjects 6 different videos, 2 for each emotion, randomized in the order of display and separated each by a 2-minute blue screen visualization. Blue screens were used to separate each emotional video from the following one in order to bring signals to baseline condition.

Since all video durations were different and around 120 seconds, only the last 100 seconds were considered for each video for the sake of feature extraction. By looking at the videos, we noticed, indeed, how the last part of the videos was more relevant for emotion elicitations. The video lengths were then uniformed also in order to compute all features for the classification as differences between each feature computed on each video and the same feature computed on the preceding blue screen.

B. Data Processing and Feature Extraction

ECG and BVP signals were downsampled at 250 Hz with a previous anti-aliasing filtering stage with a 4th order zero-phase low-pass Butterworth filter with cut-off frequencies at 125 Hz and 25 Hz, respectively. The time locations of each R peak on the ECG were extracted through a Pan-Tompkins based algorithm, the location and corresponding values of systolic and diastolic events from the BVP signal were extracted and synchronized with R-events by looking at maximum and minimum points between two successive R-events, respectively. The considered fiducial points allowed for the extraction of the following time series: tachogram from the R events, the systolic time series as well as the pulse pressure (PP), estimated as the difference between corresponding systolic and diastolic BVP values, and pulse arrival time (PAT) as the time interval between the pulse onset and the corresponding R-peak. Time and frequency domain heart rate variability features were extracted from the tachogram, time and frequency blood pressure variability measures were extracted from systolic time series, and the first two statistical moments were extracted from PP and PAT. Frequency domain features were computed with an autoregressive model with the Yule-Walker approach by choosing the optimal order (in the range of 7-15) as the lowest order that guarantees white residuals and/or minimized the Akaike information criterion. The extracted features were:

ECG features: average and standard deviation of NN intervals (*AVNN, SDNN*), power spectral density of RR in very low (*RR_VLF*), low (*RR_LF*), high (*RR_HF*) frequencies and *LF/HF*, normalized power spectral density of RR in low (*RR_LFn*) and high (*RR_HFn*) frequency ranges, first and second standard deviations and their ratio from the Poincarè analysis (*SD1, SD2, SD_ratio*).

BVP features: average pulse pressure (*AVPP*), average and standard deviation of systolic values (*AVSYS, SDSYS*), diastolic values (*AVDIA, SDDIA*), power spectral density of systolic values in low (*SYS_ LF*), high (*SYS_HF*) frequencies and normalized power spectral density of SYS in low (*SYS_LFn*) and high (*SYS_HFn*) frequencies.

ECG-BVP features: average pulse arrival time (*AVPAT*) and power spectral density of PAT in very low (*PAT_ VLF*), low (*PAT_LF*) and high (*PAT_HF*) frequency ranges.

The GSR signal was pre-filtered at 2 Hz with a zero-phase low pass Butterworth of 4th order and downsampled at 5 Hz. In order to find the phasic component, a median filter was applied by computing the median GSR of the surrounding samples in an interval of +/- 4 seconds centered on the current sample [10]. The median GSR was then subtracted from the raw signal to obtain the phasic component. After finding peak onsets (amplitude > 0.01 μS) and offsets (0 μS $<$ amplitude) on the phasic signal, a GSR peak was identified on the raw signal between each onset and offset occurrences.

GSR features: The list of extracted features includes the number of GSR peaks (*GSR_n_peaks*), the average rise and recovery time of each GSR peak (*AVGSR_rise_time*, *AVGSR_recovery_time*), GSR kurtosis (*GSR_kurt*), GSR skewness (*GSR_skew*), GSR median (*GSR_med*), GSR average (*AVGSR*), GSR standard deviation (*SDGSR*). In addition, from the raw GSR signal filtered with a Butterworth band-pass filter of the 2th order from 0.5 to 1 Hz we obtained the slope (*GSR_slope*), maximum signed amplitude between two consecutive maxima (*MAXGSR_sign_amp*), as well as the average, standard deviation and maximum of the first derivative (*AVGSR_der, SDGSR_der, MAXGSR_der*).

C. Feature Selection and Classification

As classification labels we used the annotations provided by the database corresponding to happiness, relaxation and fear. Of note, these annotations match with subjective ratings, thus providing a reasonable balance in defining the three classes. The dataset was randomly split into training (70% of the sample, 126 records) and test (30% of the sample, 54 records) datasets. Stratification was applied in order to maintain the same proportion of each emotion in the training and test partitions.

The choice of the most relevant features was performed only considering the training dataset in order not to generate bias in the prediction of test observations. Our method relies on an analytical approach which allows for feature sorting by importance in separating the emotions. Specifically, in order to understand features that produce a more effective separation among the three emotions, 2D boxplots were created between selected couples of features. For each feature pair, considering all the possible combinations, the average of each feature +/- 95% confidence intervals (CI) were computed. Consequently, 2D boxplots were built by centering the rectangles around the coordinates of the corresponding average values, where the length of the sides of the rectangles were set equal to the CI of the two averages. A specific example is shown in Fig. 1.

Only those feature pairs where rectangles of different emotions did not intersect among each other were considered.

Feature importance was then defined based on the area of the triangular region enclosed by the average values of each emotion. A larger area corresponds to a greater separation of the three emotions in the space created by the pairs of features and therefore to higher importance of the features.

Then, we computed Pearson's correlation between all the feature pairs and we chose the first ten features which showed highest triangular areas and were not highly correlated (>0.65 or <-0.65) with more than one feature belonging to a more relevant pair.

The selected features were fed to a Linear Discriminant Analysis (LDA) model in order to classify the three emotions. Given the small size of the dataset (180 records = 6 videos*30 subjects), classification results on the test set were compared with the average result obtained on a 10-fold cross-validation of the training dataset. In order to address over-fitting of the model, accuracy was computed for the training cross-validation and testing procedure. The model was built by using the singular value decomposition as solver and features were normalized between their minimum and maximum.

III. RESULTS

By computing all the possible combinations of feature pairs, 435 pairs were found. Of all these combinations, 63 showed no intersection among rectangles created with 2D boxplots.

Fig. 1 shows the 2D boxplot with the greatest area of the triangle formed by the feature averages. As we can notice, the most relevant feature pair able to separate the three emotions includes *AVGSR* and *SD_ratio* from the Poincarè analysis. The figure shows also the capability of these two features in dividing low from high arousal by the yellow dotted line. Fear and happiness belong indeed to high arousal while relaxation is linked to low values of arousal. The separation in the valence dimension is shown by the black dotted line. Fear belongs to low valence while happiness and relaxation belong to high valence, since they are more pleasant emotions.

From the feature selection analysis, the pairs *SD_ratio/AVGGSR* and *SD_ratio/GSR_med* resulted to be the ones with the highest triangular area. In particular, *AVGGSR* and *GSR_med* appeared to be part of the highest number of pairs which did not show any intersection among the three rectangles (both features appeared in 29 pairs out of 63). It means that in most of the cases, the presence of *AVGGSR* and *GSR_med* in a pair was necessary in order to prevent an overlap among the three emotions. From all the other features, whose appearances in the pairs with no intersection ranged from 2 to 5, we discarded those which were highly correlated with more than one feature associated to pairs linked to higher triangular areas. In this regard, from the first twelve important features we discarded *SDNN* and *SD2* which were found to be highly correlated between theirselves and with *SD_ratio* and *SYS_LF*.

At last, the ten most relevant features from the ECG identification (*SD_ratio*, *RR_HFn*), from the BVP (*SYS_LF*),

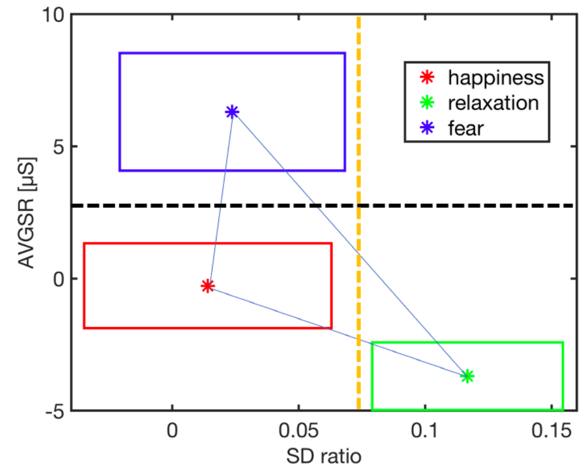

Fig. 1. 2D boxplots are represented. * represent the averages of *SD_ratio* on x axis and *AVGSR* on y axis. All rectangles are linked to the average features +/- 95% confidence intervals for the average estimations. The triangle joining the average values of the features was used as a selection criterion for feature importance. The highest the area of the triangle the more important the feature pair was considered. The yellow dotted line shows how emotions are separated in terms of arousal (fear and happiness: high arousal/ relaxation: low arousal). The black dotted line shows emotion separation in terms of valence (happiness and relaxation: high valence/ fear: low valence)

from the combined identification between ECG and BVP (*AVPAT*) and from the GSR (*AVGSR, GSR_med, SDGSR, MAXGSR_sign_amp, GSR_n_peaks, AVGSR_rise_time*) were used.

Results from the model computation are reported in Tab. 1, dividing the model performance with and without the 10 fold cross-validation. In the upper part of the table results from LDA model with 10 fold cross-validation are presented in terms of average training accuracy and average validation accuracy for each emotion. Test accuracy is also reported divided by emotions. Overall, by joining the three emotions, we reported the average accuracy for training, validation and test sets.

IV. DISCUSSION

The purpose of the study was to create a classification algorithm for the recognition of three emotions (happiness,

TABLE I
LDA MODEL RESULTS OBTAINED FOR EACH EMOTION ON TRAINING, VALIDATION AND TEST SETS.

Training and Validation Accuracy (%)			
Data Set	Happy	Relax	Fear
Training set	57	70	72
Validation set	56	60	65
Test Accuracy (%)			
Test Set	72	67	89
Averaged Accuracy among emotions (%)			
Data Set	Accuracy		
Training Set	66		
Validation Set	60		
Test Set	76		

relaxation and fear) by means of features extracted from three physiological signals: ECG, BVP and GSR.

While many studies try to increase the accuracy of models using space reduction methods, creating a new feature space, we maintained the original physiological features in order to identify the most effective in separating the three emotions.

The feature selection method that we used was easy to implement and it succeeded in sorting and combining features preserving their original values unlike, transforming the space by Fisher projections and PCA, as it is usually performed for this aim. The results obtained with the classification model are in line with similar studies where a three-emotion user independent classifier is used, whose performances range from an average test accuracy of 62% in [11] to 78% in [12].

Our feature selection analysis shows that most of the relevant features found are extracted from the GSR signal. It is well known that this signal is highly related to the arousal dimension and it is already a common measure used to quantify emotional arousal in a control scenario [13]. In this regard, *AVGSR* and *GSR_med* were found to be the most relevant features since in almost all feature pairs which did not show any overlap among the three emotions, one of these two features was present. Specifically, GSR features helped separating *fear*, the emotion with the highest arousal values with an accuracy of 89% by the LDA model. Here, the sympathetic branch of the autonomic nervous system is highly excited and sweat gland activity increases. Indeed as it can be noted in Fig.1, *AVGSR* has higher values for *fear* and the rectangle related to this emotion is well separated from the others. On the other hand, *happiness* and *relaxation* are characterized also by negative values of *AVGSR* since all the features were computed as differences between each emotional video and the preceding blue screen. This means that, in some cases for the two emotions the gland activity is lower than when subjects looked at the blue screen only.

Cardiovascular features were also found to be important in separating the three emotions. Of all the features considered, the one that has received the most importance is *SD_ratio*. SD features are relatively important as a marker of symphathovagal activation, with *SD1* reflecting parasympathetic activity and *SD2* sympathetic modulation [14]. The main reason we did not find significance in the most commonly used frequency domain indices might be due to the short intervals (100s) that we had to consider. From Fig. 1 we can see how among the three emotions, the values of *SD_ratio* are greater for relaxation, where parasympathetic activity prevails. Although *happiness* is a high arousal emotion, it does not elicit as much arousal as *fear*. Therefore, it is reasonable as confirmed by our analysis that there is more difficult to separate it from *relaxation*. Some test observations were indeed confused between *happiness* and *relaxation*, as revealed by our analysis.

V. CONCLUSION

We here present an original characterization of emotional states through physiological measures extracted from ECG, BVP and GSR time series. In particular, we focus on a feature selection method aimed at preserving the original physiological values of the features without incurring in space reduction methods.

In general, our results validate the importance of GSR features in describing the arousal dimension. Overall, the average and median values of the GSR signal, and *SD_ratio* from the Poincarè analysis on the ECG were found to be the optimal measures that separate *happiness, relaxation* and *fear*. To further validate our assessment along both arousal and valence dimension, we used the first ten important features extracted from the three signals to develop a LDA model able to identify the different emotions with accuracy as high as 76%.

REFERENCES

[1] J. Sorinas, J. M. Ferrández, and E. Fernandez, “Brain and Body Emotional Responses: Multimodal Approximation for Valence Classification,” Sensors, vol. 20, no. 1, p. 313, Jan. 2020, doi: 10.3390/s20010313.

[2] P. R. Kleinginna Jr. and A. M. Kleinginna, “A categorized list of motivation definitions, with a suggestion for a consensual definition,” Motivation and Emotion, vol. 5, no. 3, pp. 263–291, Sep. 1981, doi: 10.1007/bf00993889.

[3] A. A. Varghese, J. P. Cherian, and J. J. Kizhakkethottam, “Overview on emotion recognition system,” presented at the 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Feb. 2015, doi: 10.1109/icsns.2015.7292443.

[4] C. Saarni. ”The development of emotional competence”, Guilford press, 1999.

[5] J. POSNER, J. A. RUSSELL, and B. S. PETERSON, “The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology,” Develop. Psychopathol., vol. 17, no. 03, Sep. 2005, doi: 10.1017/s0954579405050340.

[6] J. A. Russell, “Core affect and the psychological construction of emotion.,” Psychological Review, vol. 110, no. 1, pp. 145–172, 2003, doi: 10.1037/0033-295x.110.1.145.

[7] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, “EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation,” The Scientific World Journal, vol. 2014, pp. 1–10, 2014, doi: 10.1155/2014/627892.

[8] P. Chettupuzhakkaran and N. Sindhu, “Emotion Recognition from Physiological Signals Using Time-Frequency Analysis Methods,” presented at the 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), Jul. 2018, doi: 10.1109/icetietr.2018.8529145.

[9] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, “A dataset of continuous affect annotations and physiological signals for emotion analysis,” Sci Data, vol. 6, no. 1, Oct. 2019, doi: 10.1038/s41597-019-0209-0.

[10] J. Bakker, M. Pechenizkiy, and N. Sidorova, “What’s Your Current Stress Level? Detection of Stress Patterns from GSR Sensor Data,” presented at the 2011 IEEE International Conference on Data Mining Workshops (ICDMW), Dec. 2011, doi: 10.1109/icdmw.2011.178.

[11] G. Rigas, C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, “A User Independent, Biosignal Based, Emotion Recognition Method,” in User Modeling 2007, Springer Berlin Heidelberg, pp. 314–318.

[12] K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” Med. Biol. Eng. Comput., vol. 42, no. 3, pp. 419–427, May 2004, doi: 10.1007/bf02344719.

[13] . M. Lempert and E. A. Phelps, “Neuroeconomics of Emotion and Decision Making,” in Neuroeconomics, Elsevier, 2014, pp. 219–236.

[14] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo, “Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability,” IEEE Trans. Affective Comput., vol. 6, no. 4, pp. 385–394, Oct. 2015, doi: 10.1109/taffc.2015.2432810.