ECG-Based Human Emotion Recognition Across Multiple Subjects

Desislava Nikolova[1], Petia Mihaylova[1], Agata Manolova[1](✉), and Petia Georgieva[2]

[1] Technical University of Sofia, Sofia, Bulgaria
desislava.v.nikolova@gmail.com, mihaylova_p@yahoo.com, amanolova@tu-sofia.bg

[2] DETI/IEETA, University of Aveiro, Aveiro, Portugal
petia@ua.pt

Abstract. Electrocardiogram (ECG) based affective computing is a new research field that aims to find correlates between human emotions and the registered ECG signals. Typically, emotion recognition systems are personalized, i.e. the discrimination models are subject-dependent. Building subject-independent models is a harder problem due to the high ECG variability between individuals. In this paper, we study the potential of two machine learning methods (Logistic Regression and Artificial Neural Network) to discriminate human emotional states across multiple subjects. The users were exposed to movies with different emotional content (neutral, fear, disgust) and their ECG activity was registered. Based on extracted features from the ECG recordings, the three emotional states were partially discriminated.

Keywords: ECG · Affective computing · Human emotion recognition · Machine learning · Artificial Neural Networks · Logistic Regression

1 Introduction

Emotions are part of any natural communication involving humans. Given the strong interface between affect and cognition on the one hand, and given the increasing versatility of computer agents on the other hand, the attempt to enable our computer tools to acknowledge affective phenomena rather than to remain blind to them appears desirable. They can be expressed through several channels and modalities. Facial expressions, gestures, postures, speech and intonation of voice are certainly those that are the most obvious. However, emotional information can also be found in many other modalities. For instance, it has been shown that there are different physiological states of the body corresponding to different emotions. Examples of such states are paralysis of muscles in case of fear, increase of heart rate for aroused emotions. They are generally

V. Poulkov (Ed.): FABULOUS 2019, LNICST 283, pp. 25–36, 2019.
https://doi.org/10.1007/978-3-030-23976-3_3

less perceivable by people unless an observer is close enough to the person that feels the emotion. However, these reactions could be easily recorded using specific sensors. Some of those physiological changes can also be directly perceived such as a sharp increase in blood pressure that would lead to a blush of the cheeks.

Currently different human computer interface (HCI) systems use physiological signals for classifying the human emotional state such as: electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EoG), skin conductive resistance (SCR), skin temperature (ST), and respiration rate (RR). Among these, ECG and EMG are the most popular choices for developing portable, non-intrusive, reliable, and computationally efficient emotion recognition systems [1]. One of the advantages of recognizing emotions and feelings using ECG signal is that this is unconscious response, basic biological necessity of the human body, and therefore it is very difficult to falsify or conceal. The use for ECG in HCI for human emotion recognition would revolutionize applications in medicine, entertainment, educ0ation, safety, etc. Nevertheless there are many theoretical and practical challenges with regard to ECG-based emotion recognition methodology. For example the heart rate can increase when the person is feeling fear or excitement or arousal. Another challenge presented in this paper is how correct is the choice between subject-dependent or subject-independent classification procedure in the case of ECG emotion recognition.

The goal of this research work is to investigate the usability of the physiological ECG signal in affective computing. The rest of this paper is organized as follows. Section 2 describes introduces the recent advances in research on emotion recognition based on ECG signals. Section 3 provides a detailed overview of the proposed methodology. Section 4 presents the experimental results based on two methods: Logistic regression and Artificial Neural Networks. The last section discusses some of the challenges and opportunities in this field and identifies potential future directions.

The paper is organized as follow: In Sect. 2 the most common feature extraction methods used in ECG signal processing for emotions recognition are reviewed. In Sect. 3 is described the proposed research methodology. In Sect. 4 are presented the experimental results and finally in Sect. 5 the conclusions are drawn.

2 Related Work

The ECG signal processing to extract relevant features can be performed either in time or frequency domain. However, the combination of features from both time and frequency domains provides often better insight of the underlying characteristics of the ECG signal.

In [2–4] different binary classifiers are compared to recognize Joy and Sadness emotional states based on the frequency domain features Continuously Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT).

In [5] Local Pattern Description (LPD) methods combined with k-Nearest Neighbour (kNN) classifier are applied to distinguish between three emotional states (Joy, Anger and Sadness).

The authors of [6] aim to assess five human emotions (happiness, disgust, fear, sadness, and neutral) using Hearth Rate Variability (HRV) features derived from the ECG. The emotions were induced via video clips on 20 healthy (23 years old) students. The ECG signals were acquired using 3 electrodes and were preprocessed using a 3rd order Butterworth filter to remove the noise and

Table 1. Overview of methods of ECG-based emotions recognition

Features	Feature selection method	Classifier
Continuously Wavelet	Binary Particle Swarm	k-Nearest Neighbor (kNN)
	Transform Optimization	
	Hybrid Particle Swarm Optimization	
	Genetic Algorithms	Fisher classifier
Discrete Wavelet Transform (DWT)	Tabu Search Algorithm (TS)	kNN Linear Discriminant Analysis (LDA)
		Fisher classifier
Local Pattern Description	Local Binary Pattern	kNN
	Local Ternary Pattern	
Hurst	Rescaled Range Statistics	Bayesian Classifier
	Finite Variance Scaling	Regression Trees
	Higher Order Statistics	kNN
		Fuzzy kNN
Fast Fourier Transform (FFT)	Tabu search (TS)	Fisher Classifier
Statistical features		Adaptive Neuro-Fuzzy
		Inference System
		Support Vector Machine
Raw signal		Deep Learning
		(Convolutional & Recurrent Neural Networks)
Discrete Cosine Transform	Principal Component Analysis (PCA)	Probabilistic Neural Network
	LDA	
	Kernel PCA	

baseline wander. DWT was used to extract statistical features from the HRV signals. The kNN and Linear Discriminant Analysis (LDA) were applied to map the statistical features into corresponding emotions.

The affective computing system proposed in [7] distinguishes six emotions (happiness, sadness, fear, disgust, surprise and neutral) induced by audio visual stimuli. The Hurst features are computed based on Re-scaled Range Statistics (RRS), Finite Variance Scaling (FVS) methods and Higher Order statistics (HOS). Bayesian Classier, Regression Tree, kNN and Fuzzy kNN are comparatively studied classifiers for this task. The results demonstrate that RRS and FVS methods have similar classification accuracy, however the FVS and HOS combined features performed better for the classification of the six emotional states.

Recently [8] proposed a deep neural network (DNN) to decode human emotions directly from row ECG data. The motivation behind the deep learning architectures is that they are able to automatically extract relevant features that may be overlooked by human experts.

Table 1 summarizes common feature extraction methods used for ECG-based emotions recognition.

Based on these recent papers, we can conclude that there is a big diversity in feature selection methods and none is predominating over the others with better recognition rates or accuracy.

3 Proposed Methodology

The process for ECG-based emotion recognition consists of four steps - data collection, feature extraction, feature normalization (if necessary) and classification.

3.1 Data Collection

Data used in this study are provided by the Psychology department of University of Aveiro. 25 volunteers (10 males, 15 females) took part in the experiments. Electrocardiogram signals were recorded while each participant watched three different movies in distinct days. The movie contents were carefully selected in order to induce the following emotions:

- movie with Neutral emotional content
- movie with Fear emotional content
- movie with Disgust emotional content

75 ECG time series were collected corresponding to 3 movies with different emotional content for each of the 25 participants. The ECG signal was divided in four segments corresponding to the baseline period (before the movie start), pre-video, mid-video and last-video periods reflecting the assumption that the emotional intensity varies over the movie duration. The average duration of each segment is as follows:

- Baseline (4 min of preparation before the movie start) - 240 000 samples;
- pre-Video (first 5 min of the movie) - 300 000 samples;
- mid-Video (next 15 min of the movie) - 900 000 samples;
- last-Video (last left x min of the movie) - varying number of samples.

The statistical analysis has shown that the ECG signal collected during the pre-video and the mid-video periods have the highest discrimination capacity, therefore the emotion recognition was focused into these signal segments.

3.2 Feature Extraction

Typical ECG signal is illustrated in Fig. 1. Quantitative information, such as amplitude and latency, regarding the P-wave, T-wave and QRS-complex wave, are the main ECG characteristics based on which most emotion recognition systems are built.

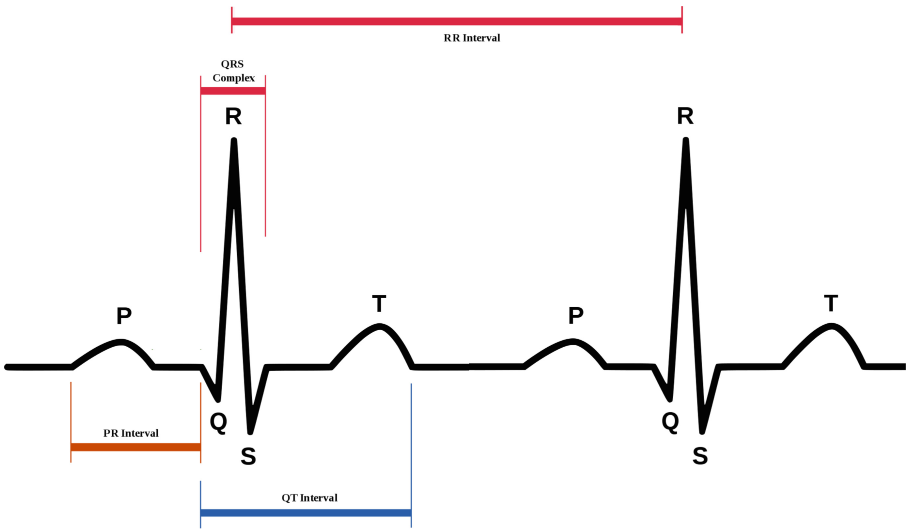

Fig. 1. ECG signal

In the present study the R peaks amplitude and the RR intervals (the number of samples between two R peaks) were extracted as primary quantities from the raw ECG signal. Based on them eight statistical features were computed, two of them related with the R peak amplitude, and six with the length of the RR intervals. The Length of the RR interval is the interval between successive R peaks. In Fig. 2 is illustrated the extraction of the R peaks from the complete ECG signal recorded during one movie. In Fig. 3 is presented a fragment of the same signal for a better visualization.

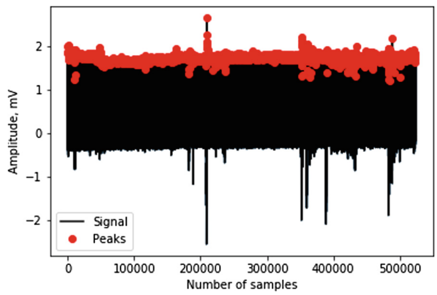

Fig. 2. ECG R peaks (Complete signal)

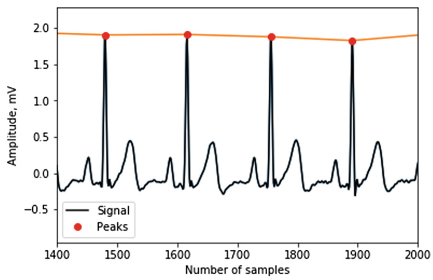

Fig. 3. ECG R peaks (Zoomed signal)

Features Based on R Peak Amplitude

The maximum amplitude of the R wave (measured in mV) is called "R peak amplitude", or just "R peak".

– *Average of R peak amplitude* ($\overline{A}$):

$$\overline{A} = \frac{\sum_{i=1}^{N} A_i}{N}, \tag{1}$$

where A_i is the current R peak amplitude, N is the number of R peaks collected during one video segment.

– *Standard Deviation of R peak amplitude (StdA)* quantifies the amount of dispersion of the R peak amplitude:

$$StdA = \sqrt[2]{\frac{\sum_{i=1}^{N} (A_i - \overline{A})^2}{N}}. \tag{2}$$

Features Based on Length of the RR Intervals

– *Mean value of RR Intervals* ($\overline{I}$):

$$\overline{I} = \frac{\sum_{i=1}^{N} I_i}{N}, \tag{3}$$

where I_i is the current length of the RR interval, N is the number of RR intervals collected during one video segment.

– *Beats per minute (BPM)*

Heart rate is the speed of the heartbeat measured by the number of contractions of the heart per minute. A normal resting heart rate for adults ranges from 60 to 100 beats a minute. The sum of the RR intervals along the axis divided by the number of intervals. Every minute contains 60 000 ms.

$$BPM = \frac{60000}{\overline{I}}. \tag{4}$$

– *Mean value of the RR Intervals absolute difference* ($\overline{I}_{abs}$):

$$\overline{I}_{abs} = \frac{\sum_{i=2}^{N} |I_i - I_{i-1}|}{N-1}. \tag{5}$$

– *Square Root Mean Value of the RR intervals absolute difference* ($\overline{I}_{sqrt}$): The root mean square successive difference in heart period series is a time domain measure of heart period variability.

$$\overline{I}_{sqrt} = \frac{\sum_{i=2}^{N} \sqrt[2]{|I_i - I_{i-1}|}}{N-1}. \tag{6}$$

– *Standard Deviation of RR intervals (StdR)*:

$$StdR = \sqrt[2]{\frac{\sum_{i=1}^{N} (I_i - \overline{I})^2}{N}}. \tag{7}$$

– *Standard Deviation of the R intervals absolute difference (StdRdif)*:

$$StdRdif = \sqrt[2]{\frac{\sum_{i=2}^{N} (I_i - I_{i-1})^2}{N-1}}. \tag{8}$$

3.3 Normalization

Though the extracted features do not vary in significantly different ranges, we studied the classification performance both with normalized and not-normalized data sets. The following normalization was applied:

$$x_{norm} = \frac{x_{original} - mean(x_{vector})}{max(x_{vector}) - min(x_{vector})}. \tag{9}$$

The distribution of the eight normalized features with respect to the three classes (disgust, fear, neutral) and during the four video segments - baseline (4 min of preparation before the movie start), pre-video (first 5 min of the movie), mid-video (next 15 min of the movie), last-video (last 5 min of the movie) are illustrated in Fig. 4.

3.4 Classification

At the classification step, two classifiers - Logistic Regression (LR) and Artificial Neural Networks (ANN) - were compared. The results are discussed in the next section.

4 Experimental Results

The statistical analysis has shown that the pre-Video and mid-Video ECG temporal segments have the highest discrimination capacity, therefore the classifiers were provided with features extracted during these segments.

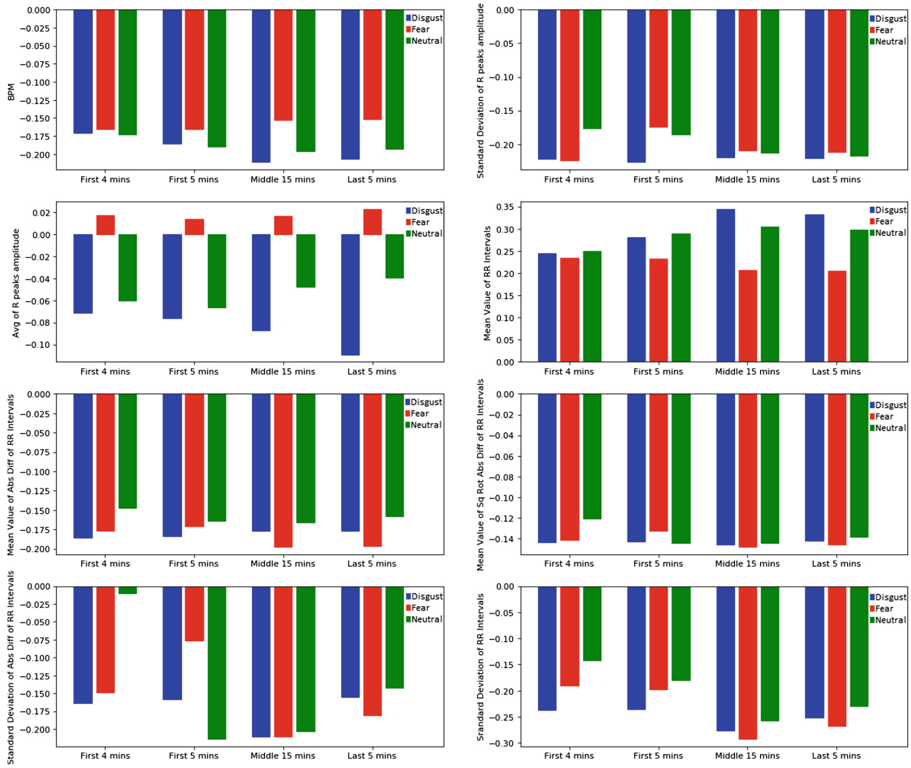

Fig. 4. Distribution of normalized features over classes (Disgust (in blue), Fear (in red), Neutral (in green)) and video segments (baseline, pre-, mid-, last-video) (Color figure online)

4.1 Logistic Regression

The performance of the Logistic Regression (LR) classifier in terms of accuracy on training and testing data was studied. LR behaves significantly better with non-normalized data, therefore only these results are shown on the next figures.

First, the convergence properties of the classifier were studied and the results are depicted in Figs. 5 and 6. For both time segments the accuracy converges after 40–50 iterations, however the testing accuracy with the pre-video ECG features is higher and closely follows the training accuracy.

Next, the optimal number of input features is assessed and the results are summarized in Figs. 7 and 8. The features are ordered based on their importance determined by the recursive feature elimination (RFE) method. The feature rank of importance is the the following: (0) Standard Deviation of R peaks amplitude, (1) Average of R peaks amplitude, (2) Beats per minute, (3) Standard Deviation of the absolute difference of RR intervals, (4) Standard Deviation of RR intervals, (5) Mean value of abs difference of RR Intervals, (6) Mean value of RR Intervals, (7) Mean Value of Square Root Abs difference of RR intervals. Note, that the

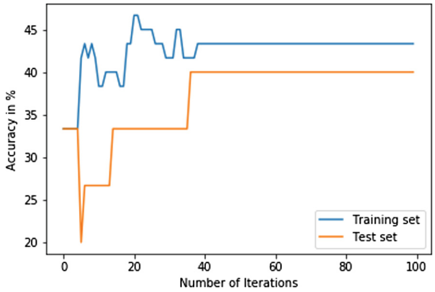

Fig. 5. LR performance vs. # of iterations for the pre-Video segment.

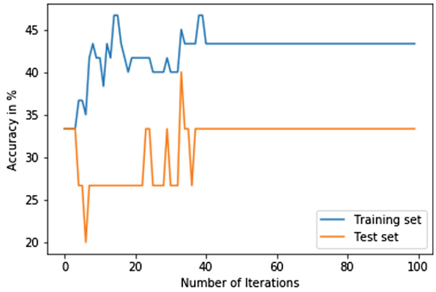

Fig. 6. LR performance vs.# of iterations for the mid-Video segment.

first five features (for the pre-video segment) and the first six features (for the mid-video segment) are sufficient to achieve the maximum test accuracy, which is the ultimate goal of the recognition model. Similarly to the previous study, the pre-video ECG features have better generalization properties (train and test performance are close enough).

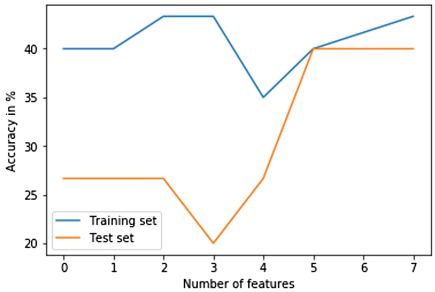

Fig. 7. LR performance vs. # of features for the pre-Video segment.

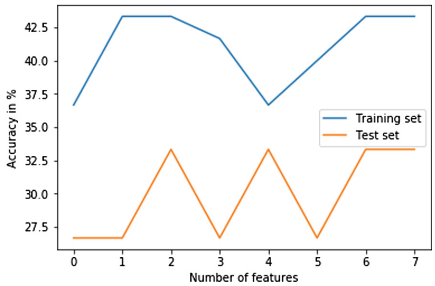

Fig. 8. LR performance vs. # of features for the mid-Video segment.

4.2 Artificial Neural Networks

Now the performance of the Artificial Neural Network (ANN) classifier in terms of accuracy on training and testing data was assessed. ANN architecture with only one hidden layer was considered as sufficient for this problem. In contrast to LR, ANN is more successful with normalized data, therefore only these results are shown on the next figures.

First, the convergence properties of the ANN classifier were studied for different choice of the activation functions (*ReLu*, *Sigmoid*, *Tanh*). As can be seen in Figs. 9 and 10, the ANN model largely outperforms the LR model (90 % training accuracy), however it is paid by a huge number of iterations (i.e. epochs), about 10000 iterations were necessary for the learning process to converge. For both time segments, *Tanh* activation function seems to be the most suitable to maximize the training accuracy.

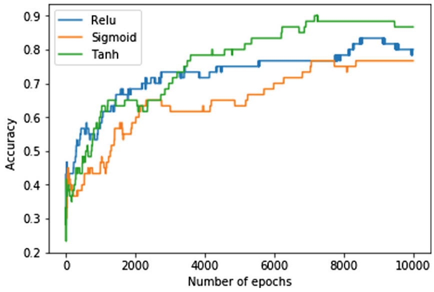

Fig. 9. ANN performance vs. # of iterations for the pre-Video segment

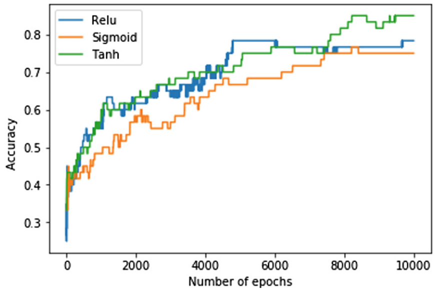

Fig. 10. ANN performance vs. # of iterations for the mid-Video segment

The search for the optimal number of nodes is summarized in Figs. 11 and 12. The final ANN architecture was fixed with 7 *Tanh* nodes.

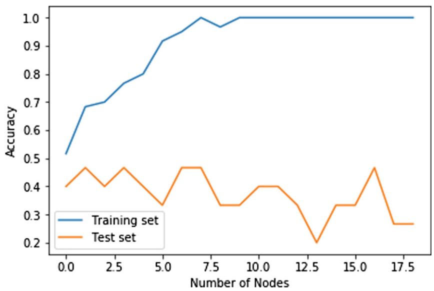

Fig. 11. ANN performance vs. # of nodes for the pre-Video segment

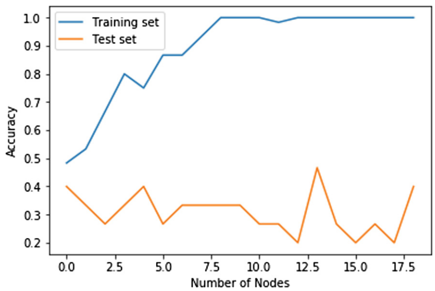

Fig. 12. ANN performance vs. # of nodes for the mid-Video segment

Note, the notorious overfitting ANN problem with the increasing number of hidden layer nodes. This problem was addressed by adding a regularization term in the cost function. The optimal value of the regularization parameter l2 (l2 = 3) was obtained after a grid search as shown in Figs. 13 and 14.

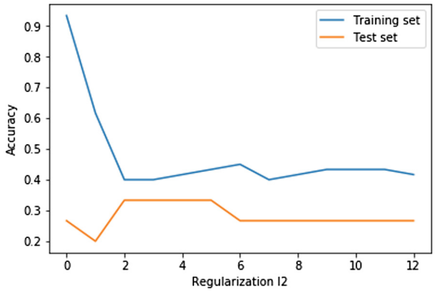

Fig. 13. ANN performance vs. regularization parameter l2 for the pre-Video segment

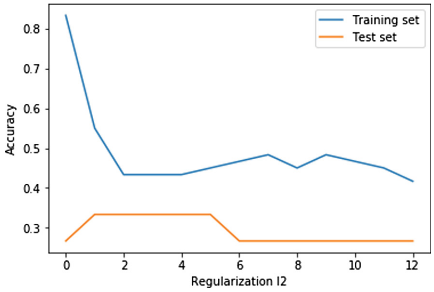

Fig. 14. ANN performance vs. regularization parameter l2 for the mid-Video segment

5 Conclusion

The goal of this paper was to built an ECG-based human emotion recognition system of three specific emotions (fear, disgust, neutral) across multiple subjects (25 volunteers, in particular). The system was trained with data from some of the participants (20 subjects) and then tested with data from the rest of the participants (5 subjects). This is a very challenging scenario, taking into account the significant ECG variability not only between subjects but also within different sessions with the same subject. Nevertheless, based on eight statistical features extracted from two major ECG characteristics - R peak amplitude and RR intervals we have obtained 40% testing accuracy with the LR classifier and 35% testing accuracy with the ANN classifier. These promising results suggest that the ECG modality may potentially be useful in affective computing if combined with other modalities such as physiological signals, facial expression, voice analysis.

Acknowledgements. This work was supported by European Regional Development Fund and the Operational Program “Science and Education for Smart Growth” under contract UNITe BG05M2OP001-1.001-0004-01 (2018–2023).

References

1. Konar, A., Chakraborty, A.: Emotion Recognition: A Pattern Analysis Approach. Wiley, Hoboken (2014)
2. Xu, Y.: A method of emotion recognition based on ECG signal. In: 2009 International Conference on Computational Intelligence and Natural Computing, pp. 202–205 (2009)
3. Min, H.: Analysis of affective ECG signals toward emotion I. Introduction ECG data collection III. ECG feature extraction. J. Electron. (China) 9–14 (2010)
4. Cai, J.: The research on emotion recognition from ECG signal. In: Proceedings of 2009 International Conference on Information Technology and Computer Science, ITCS 2009, pp. 497–500 (2009)

5. Tivatansakul, S.: Emotion recognition using ECG signals with local pattern description methods. Int. J. Affect. Eng. **15**, 51–61 (2016)
6. Murugappan, M.: Frequency band analysis of electrocardiogram (ECG) signals for human emotional state classification using discrete wavelet transform (DWT). J. Phys. Ther. Sci. **25**, 753–759 (2013)
7. Selvaraj, J.: Classification of emotional states from electrocardiogram signals: a non-linear approach based on hurst. BioMed. Eng. **12**, 44 (2013)
8. Keren, G.: End-to-end learning for dimensional emotion recognition from physiological signals. In: Proceedings of IEEE International Conference on Multimedia and Expo, pp. 985–990 (2017)