Emotion Classification Based on Bio-Signals Using Machine Learning Algorithms

Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Myung-Ae Chung
IT Convergence Technology Research Laboratory & Creative Future Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
{cleta4u, bj_park, shk1028, machung}@etri.re.kr

Yeongji Eum, Jin-Hun Sohn
Department of Psychology, Brain Research Institute
Chungnam National University
Daejeon, Republic of Korea
petitaudrey@hanmail.net, jhsohn@cnu.ac.kr

***Abstract*—In human-computer interaction researches, one of the most interesting topics in the field of emotion recognition is to recognize human's feeling using bio-signals. According to previous researches, it is known that there is strong correlation between human emotion state and physiological reaction. Bio-signals takes noticed lately because those can be simply acquired with some sensors and are less sensitive in social and cultural difference. We have applied four algorithms, linear discriminant analysis, Naïve Bayes, decision tree and support vector machine to classify emotions, happiness, anger, surprise and stress based on bio-signals. In this study, audio-visual film clips were used to evoke each emotion and bio-signals (electrocardiograph, electrodermal activity, photoplethysmo-graph, and skin temperature) as emotional responses were measured and the features were extracted from them. For emotion recognition, the used algorithms are evaluated by only training, 10-fold cross-validation and repeated random sub-sampling validation. We have obtained very low recognition accuracy from 28.0 to 38.4% for testing. This means that it needs to apply various methodologies for the accuracy improvement of emotion recognition in the future analysis. Nevertheless, this can be helpful to provide the basis for the emotion recognition technique in human-machine interaction as well as contribute to the standardization in emotion-specific autonomic nervous system responses.**

Keywords-emotion classification; bio-signal; feature extraction; machine learning algorithm

I. INTRODUCTION

Recently, one of the most interesting fields in Human Computer Interaction (HCI) is to understand human's feeling and to categorize emotions. Some engineers and psychologists have tried to analyze facial expressions, voices, gestures and bio-signals in an attempt to recognize emotions [1][2]. In particular, studies to recognize human's feeling using various bio-signals have gradually increased, because signal acquisition by non-invasive sensors is relatively simple and physiological responses by emotion are less sensitive in social and cultural difference. Emotional states are recognized by some signals reflect physiological responses such as heartbeat, respiration, skin temperature and, so on. For example, Electrocardiograph (ECG) is a signal to detect the electrical activity of the heart through electrodes attached to the outer surface of the skin and reflects emotional states such as tension or stress. Also, Electrodermal Activity (EDA) is a physiological signal that can characterizes changes in the electrical properties of the skin owing to the activity of the sweat glands and a good indicator of arousal level due to external sensory and cognitive stimuli [3][4]. Skin Temperature (SKT) is an important and effective indicator of emotion states and reflects Autonomic Nervous System (ANS) activity. Photoplethysmograph (PPG) is also a signal that indicates pulsation of chest wall and great arteries followed by heartbeat, and measures activities of the sympathetic and para-sympathetic nervous system. Many previous studies have already examined that there is a strong relation between physiological responses and human's emotional states [5]. For example, in research reviewed 134 studies about ANS activity [6], anger is related to ANS responses such as a modal response pattern of reciprocal sympathetic activation and increased respiratory activity, particularly faster breathing, and ANS responses of fear point to broad sympathetic activation including cardiac acceleration, increased myocardial contractility, vasoconstriction, and electrodermal activity. Emotion recognition using bio-signals has been mostly performed by machine learning algorithms (e.g., Fisher's Linear Discriminant (FLD) [7], Support Vector Machine (SVM) [8] and so on). For example, Picard and colleagues at MIT Lab [1] have conducted a recognition accuracy of over 80% on average which seems to be acceptable for realistic applications using linear pattern recognition method. In this paper, we introduce the analysis processes such as signal processing, features extraction and deduction for classification of emotions (happiness, anger, surprise and stress) based on bio-signals and results of emotion classification using some machine learning algorithms. To induce four basic emotions, ten emotional stimuli sets which have been verified their appropriateness and effectiveness by replicate experiments were used in experiment. ECG, EDA, SKT and PPG as bio-signals are acquired by MP100 Biopac system Inc. (USA) [9] and analyzed to extract features for emotional pattern dataset. To classify four emotions, four machine learning algorithms, which are Linear Discriminant Analysis (LDA) [7], Classification And Regression Tree (CART) [10], Naïve Bayes [7] and Support Vector Machine (SVM) [8] are used. The results will offer information about the emotion recognizer with feature selections using bio-signlas induce by four emotions.

 ISBN: 978-1-61208-340-7

II. EXPERIMENTAL METHODS

Twelve subjects (males: 20.8 years±1.26, females: 21.2 years±2.70) participated in this study. They are normal persons who did not report any history of medical illness or psychotropic medication. A written consent was obtained before the beginning of the experiment.

A. Emotion Induction Experiment

To effectively induce the four emotions (happiness, anger, surprise and stress), forty emotional stimuli, which consist of 10 sets for the four emotions, are used in the experiments. Those are constituted 2~4 min long audio-visual film clips which are captured originally from movies, documentary and TV shows. The stimuli for happiness have included scenes such as victory, wedding or laughing and scenes such as massacre, beating, or attack for anger induction, scenes of a sudden or unexpected scream etc., as the stimuli for surprise, and audio/visual noise on screen for stress. Audio-visual film clips take advantage that these have the desirable properties of being readily standardized, involving no deception, and being dynamic rather than static. They also have a relatively high degree of ecological validity, in so far as emotions are often evoked by dynamic visual and auditory stimuli that are external to the individual [11].

In preliminary study, to verify whether each emotional stimulus induce real emotion or not, we had examined the appropriateness and effectiveness of them. The appropriateness means consistency (%) between target emotion designed by experimenter and label of subjects' experienced emotion. The effectiveness is the emotion intensity by subjects' rating on a 1 to 11 point Likert-type scale (e.g., 1 being "least happy" or "not happy" and 11 being "most happy"). Twenty-two college students, that are different group from participants in the experiment, categorize their experienced emotion and rate intensity of their categorized emotion on emotional assessment scale after presentation of each emotion stimuli.

The emotional stimuli have the appropriateness of 92.25% and the effectiveness of 9.3 point on average of 10 sets as shown in the results (Table 1). The appropriateness of each stimulus is ranged from 75 to 100% and the effectiveness comes out from 8.0 to 10.3 point as shown in results. This means that the selected stimuli can provoke each emotion, suitably and effectively.

The procedures for main experiment are as like figure 1. Subjects have introduced to experiment procedures and have an adaptation time in the laboratory setting. Then they are attached electrodes on their wrist, finger, and ankle for bio-signals measurement. Bio-signals have recorded for 60 sec prior to the stimulus presentation (baseline) and for 2 to 4 min during the stimulus presentation as emotional state, then for 60 sec after presentation of stimulus for debriefing. Subjects have rated the emotion that they experienced during presentation of the stimulus on the emotion assessment scale. This procedure is conducted on each of the four emotions for 10 times.

TABLE I. THE APPROPRIATENESS AND EFFECTIVENESS OF EMOTIONAL STIMULI

	HAP	ANG	SUR	STR	M
1	100% (8.4)	75% (9.7)	75% (9.3)	92% (9.3)	83% (9.5)
2	100% (8.9)	75% (9.9)	92% (9.7)	100% (9.1)	94% (9.6)
3	100% (8.8)	75% (9.7)	100% (9.7)	100% (8.8)	93% (9.3)
4	100% (9.6)	75% (9.5)	100% (9.9)	100% (8.9)	95% (9.7)
5	100% (9.6)	92% (9.8)	83% (9.6)	100% (9.3)	94% (9.6)
6	100% (9.3)	92% (9.4)	83% (9.6)	100% (8.8)	95% (9.5)
7	100% (9.3)	92% (8.9)	100% (9.5)	92% (9.3)	92% (9.3)
8	92% (8.0)	83% (9.2)	83% (9.4)	100% (9.3)	92% (9.2)
9	100% (9.7)	92% (9.5)	83% (8.6)	100% (9.1)	96% (9.4)
10	92% (8.8)	92% (9.7)	75% (10.3)	100% (9.3)	91% (9.5)
M	98% (9.1)	84% (9.5)	89% (9.5)	98% (9.1)	92% (9.3)

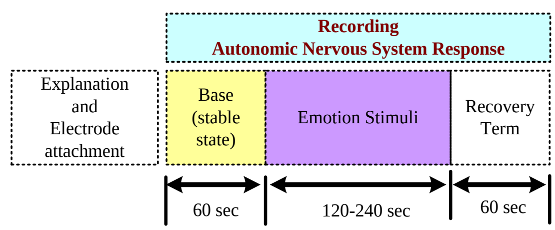

Figure 1. Experimental protocol for emotion induction

B. Measurement of Bio-signals

The MP100WS and AcqKnowledge (version 3.8.1) (Biopac Systems Inc., USA) were used to acquire the data and analyse them. The sampling rate was fixed at 250 Hz for all the channels and appropriate amplification and band-pass filtering were performed. Figure 2 shows an example of the obtained signals from device. The ECG was measured from both wrists with the two-electrode method its basis from the left ankle (Lead I) as reference. The respiration sensor measured expansion and contraction of the chest cavity using a Hall effect sensor attached around the chest with a Velcro band. EDA was measured from two Ag/AgCl electrodes attached to the index and middle fingers of the left hand. PPG and SKT were measured from the little finger and the ring finger of the left hand, respectively. PPG allows non-invasive recording of arterial-blood-volume pulses at the finger.

Data for 30 sec from the baseline and another 30 sec from each emotional state were used in the analysis. The emotional states are determined by the result of subject's self-report (scene that emotion is most strongly expressed during presentation of each stimulus). Total 360 signal data except for severe artefact effect by movements, noises, etc. are used for analysis.

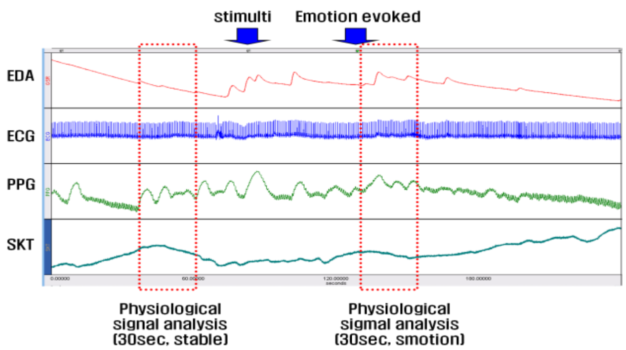

Figure 2. The example of bio-signals measures

C. *Feature Extraction*

27 features extracted from the bio-signals and used to analysis (Table 2). The features extracted from EDA, which was down-sampled to 100Hz after low-pass filtering the signal, are skin conductance level (SCL), average of skin conductance response (mean SCR) and number of skin conductance response (NSCR). Figure 3 shows the example of EDA signal. These are features which show statistically significant change between baseline and emotion state by paired t-test (using SPSS ver.18.0).

TABLE II. THE FEATURES EXTRACTED FROM BIO-SIGNALS

Bio-signal			**Features**
EDA			SCL, NSCR, meanSCR
SKT			meanSKT, maxSKT
PPG			meanPPG
ECG	Time domain	Statistical parameter	meanRRI, stdRR, meanHR, RMSSD, NN50, pNN50
		Geometric parameter	SD1, SD2, CSI, CVI, triangular index, TINN
	Frequency domain	FFT	apLF, apHF, nLF, nHF, LF/HF ratio
		AR	apLF, apHF, nLF, nHF, LF/HF ratio

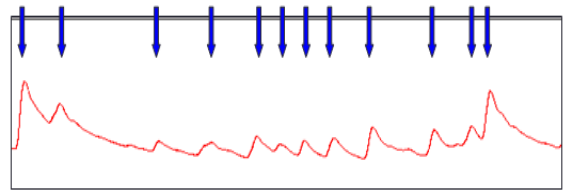

Figure 3. The example of EDA signals (200 sec)

Figure 4 is the law signal of ECG during 10 sec and the example of RRI tachogram extracted from ECG. RRI (msec) is the time interval between two R-peaks in the ECG.

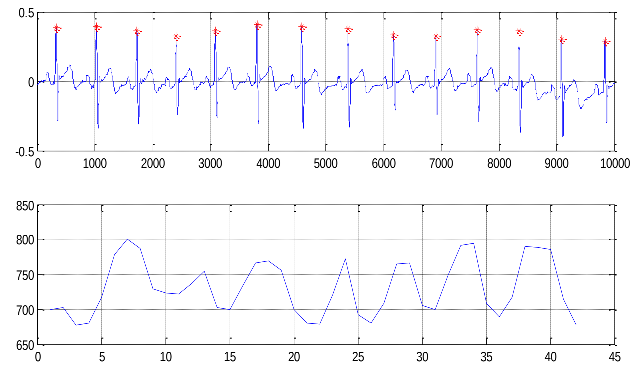

Figure 4. ECG signal (10 sec) and RRI tachogram (30 sec)

For extracting an emotional feature based on bio-signals, ECG analysis in the time (statistical and geometric approaches) and frequency domain (FFT and AR) was performed. RRI and heart rate (HR) offers the mean RRI (mean RRI) and standard deviation (std RRI), the mean heart rate (mean HR), RMSSD, NN50 and pNN50. RMSSD is the square root of the mean of the sum of the squares of differences between successive RRIs. NN50 is the number of RRI with 50msec or more and the proportion of NN50 divided by total number of RRI is pNN50. In addition to those, RRI triangular index (RRtri) and TINN are extracted from the histogram of RRI density as a geometric parameter. RRtri is to divide the entire number of RRI by the magnitude of the histogram of RRI density and TINN is the width of RRI histogram (M-N) as shown in Figure 5.

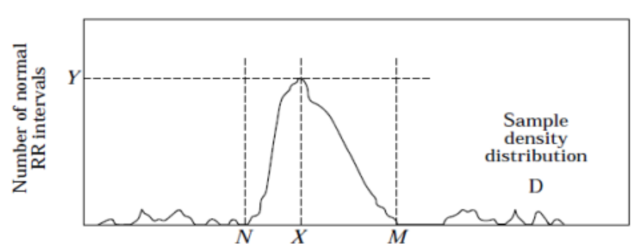

Figure 5. Histogram of RRI

The relations between RRI (n) and RRI (n+1) are called Lorentz or Poincare plot [12] as shown in Figure 6. Here, n and n+1 are n-th and n+1-th values of RRI, respectively. In the figure, L is the direction that is efficient for representing data, and T is the orthogonal direction of L. The standard deviations, SD1 and SD2, are gotten for T and L directions, respectively. The cardiac sympathetic index (CSI) is calculated by CSI = 4SD2/4SD1 and the cardiac vagal index (CVI) is obtained from CVI = log10 (4SD1 * 4SD2) as an emotional feature. SD1, SD2, CSI and CVI reflect short term Heart Rate Variability (HRV), long term HRV, sympathetic nerve activity and parasympathetic activity, respectively.

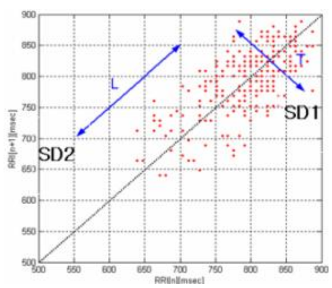

Figure 6. Lorentz plot of RRI

Also, we used the fast FFT and the AR power spectrum. Figure 7 shows Power spectrum density of FFT and AR on frequency domain. The band of low frequency (LF) is 0.04~0.15 Hz and the high frequency (HF) is 0.15~0.4Hz. The total spectral power between 0.04 and 0.15 Hz is apLF and the normalized power of apLF is nLF. apHF and nHF are the total spectral power between 0.15 and 0.4 Hz and the normalized power, respectively. L/H ratio means the ratio of low to high frequency power. These are resulted by averaging FFT and AR. LF and HF are used as indexes of sympathetic and vagus activity, respectively. The L/H ratio reflects the global sympatho-vagal balance.

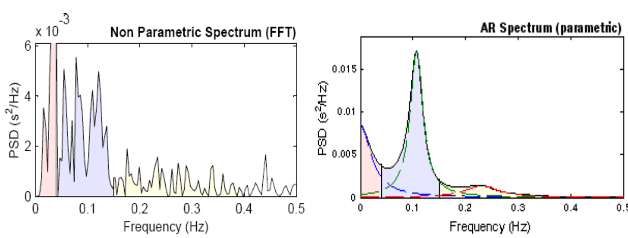

Figure 7. Power spectrum density of FFT (left) and AR (right)

The mean skin temperature (mean SKT) and maximum skin temperature (max SKT) and the mean amplitude of blood volume changes (mean PPG) are gotten from SKT and PPG, respectively.

D. Machine Learning Algorithm for Emotion Rrecognition

Fisher's LDA is one of the linear models to find a linear combination of features which characterizes or separates two or more classes of objects and is used in statistics, pattern recognition and machine learning. LDA finds the direction to project data on so that between-class variance in maximized and within-class variance in minimized, and then offers a linear transformation of predictor variables which provides a more accurate discrimination [7]. In LDA, the measurement space is transformed so that the separability between the emotional states is maximized. The separability between the emotional states can be expressed by several criteria.

$$W^{*} = \arg\max \left\{ \frac{W^{T} S_{B} W}{W^{T} S_{W} W} \right\} \tag{1}$$

CART is one of decision tree and nonparametric technique that can select from among a large number of variables those and their interactions that are most important in determining the outcome variable to be explained [10]. The fundamental principle underlying tree creation is that of simplicity. We prefer decisions that lead to a simple, compact tree with few nodes. In formalization this notion, the most popular measure is the entropy impurity (or occasionally information impurity):

$$i(N) = -\sum_{j} P(\omega_j) \log_2 P(\omega_j) \tag{2}$$

where, $P(\omega_j)$ is the fraction of patterns at node N that are in class ω_j. By the well-known properties of entropy, if all the patterns are of the same category, the impurity is 0; otherwise it is positive, with the greatest value occurring when the different classes are equally likely.

The Naïve Bayes algorithm is a classification algorithm based on Bayes rule and particularly suited when the dimensionality of the inputs is high [7]. When the dependency relationships among the features used by a classifier are unknown, we generally proceed by taking the simplest assumption, namely, that the feature are conditionally independent given the category, that is,

$$p(\omega_k \mid \mathrm{X}) \propto \prod_{i=1}^{d} p(\chi_i \mid \omega_k) \tag{3}$$

This so-called naïve Bayes rule often works quite well in practice, and it can be expressed by a very simple belief net.

SVM is the well-known emotion algorithms and non-linear model that support vector classifier can be extended to nonlinear boundaries by the kernel trick. SVM is designed for two class classification by finding the optimal hyperplane where the expected classification error of test samples is minimized and has utilized as a pattern classifier to overcome the difficulty in pattern classification due to the large amount of within-class variation of features and the overlap between classes, although the features are carefully extracted [8]. The distance from any hyperplane to a pattern y is |g(y)|/||a||, and assuming that a positive margin b exists

$$zk\ g(yk) / \|a\| \geq b,\ k = 1, \ldots, n; \tag{4}$$

 ISBN: 978-1-61208-340-7

The goal is to find the weight vector a that maximizes b. Here, zk is the class of k-th pattern, b is margin and g(y) is a linear discriminant in an augmented y space,

$$g(y) = aTy \tag{5}$$

III. EMOTION CLASSIFICATION

To examine the difference physiological responses of among four emotions, one-way ANOVA was performed using SPSS ver.18.0. We could identified that there are significant the differences in NSCR, meanSCR, meanPPG, TINN and FFT L/H ratio ($p<.05$) (Table 3). Results of post-hoc test (LSD) showed that the change of NSCR during stress is higher than other emotions, happiness, anger and surprise, and meanSCR induced by surprise increase more than other emotions significantly.

TABLE III. RESULT OF ONE-WAY ANOVA BY 28 FEATURES

Features		SS	df	MS	P
NSCR	between	30.451	3	10.150	.01
	within	986.756	357	2.604	
	sum	1017.206	360		
meanSCR	between	10.503	3	3.501	.00
	within	204.954	357	.541	
	sum	215.457	360		
meanPPG	between	4.130	3	1.377	.05
	within	245.480	357	.648	
	sum	249.610	360		
TINN (ms)	between	34.624	3	11.541	.05
	within	849.876	357	5.379	
	sum	884.500	360		
FFT L/H ratio	between	112.024	3	37.341	.04
	within	6125.350	357	16.162	
	sum	6237.374	360		

For recognition accuracy of emotions, the four machine learning algorithms were evaluated on only Training (TR), 10-fold Cross-Validation (CV) and Repeated Random Sub-sampling Validation (RRSV). For TR, the entire dataset is used to build a recognizer and evaluate the built recognizer. We have also applicated CV and RRSV cosidering that TR has the overfitting problem. In 10-fold cross-validation, the entire dataset is partitioned into 10 equal size subsets. Of the 10 subsets, a single subset is retained as the testing data for testing the recognizer, and the remaining 9 subdatasets are used as training data to build a recognizer. In RRSV, the 70% of the whole emotional patterns are selected randomly for training, the remaining patterns are used for testing purposes and this is repeated 10 times.

TABLE IV. RESULT OF EMOTION RECOGNITION ON FEATURE SPACE WITH 28 FEATURES

Machine Learning Algorithms	TR	CV	RRSV	
			Training	Testing
LDA	48.2	33.4	52.3±2.4	29.3±4.3
CART	86.7	35.8	83.3±1.1	34.3±5.6
Naive Bayes	77.7	43.7	78.9±1.4	38.4±5.9
SVM	100	31.9	100.0±0.0	28.0±4.9

We have performed feature normalization and the related parameters of algorithms using default values, which have offered with toolbox. Table 4 shows the recognition results by using the TR, CV and RRSV for 28 features. The accraccy of emotion recognition have higher values for training than testing. The CV exhibits the results for testing. To apply to real system, we have to discuss in the view point of testing. For 28 features, the results of emotion recogntion for CV has range of 31.9 to 43.7% when all emtions are recognized for test dataset. The accuracy of recognition for RRSV shows in range 28.0 to 38.4% for testing.

IV. CONCLUSIONS

The aim of this study was to classify four emotions, happiness, anger, surprise and stress, induced by audio-visual stimuli. For this, we have gotten the bio-signals based on ANS responses of the evoked emotions. Also, twenty-eight features have been analyzed and extracted from these signals. For four emotions classification, we have used four machine learning algorithms, namely, LDA, CART, Naïve Bayes, and SVM. The results were reported by only TR, CV and RRSV. However, we have only obtained very low recognition accuracy from 28.0 to 38.4% for testing and this means that there is a problem with improvement of recognition accuracy for the four emotions, becuase recognition results showed the low accuracy for testing. To apply to real system, we have to discuss in the view point of testing and this means that it needs to apply various methodologies for the accuracy improvement of emotion recognition in the future analysis. We will investigate various methodologies for dealing the accuracy improvement of emotion recognition in the future research (e.g., the deduction of the features such as NSCR or mean SCR are significant differences among emotions by statistical methods, data normalization or the use of enhanced algorithm). Although bio-signal offers a great potential for the emotion recognition in computer systems, in order to effectively exploit the advantages of bio-signals, there are limitations, such as standardization on the emotional model, the measures and feature extraction of bio-signals, signal patterns, and model for pattern recognition and classification [13]. Nevertheless, our result can be useful in developing an emotion recognition system based on bio-signals in HCI.

ACKNOWLEDGMENT

This research was supported by the Converging Research Center Program through the Ministry of Science, ICT and Future Planning, Korea (2013K00 0329 and 2013K000332).

REFERENCES

[1] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, 2001, pp. 1175-1191.

[2] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein, “Emotion recognition from physiological signals using wireless sensors for presence technologies,” Cognition, Technology and Work, vol. 6, 2004, pp. 4-14.

[3] W. Boucsein, Electrodermal activity, New York: Plenum Press, 1992.

[4] C. Maaoui and A. Pruski, Emotion recognition through physiological signals for human-machine communication. in Cutting Edge Robotics 2010, Vedran Kordic (Ed.), 2010, pp. 317.

[5] P. D. Drummond and S. H. Quah, “The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasian,” Psychophysiology, vol. 38, 2001, pp. 190-196.

[6] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” Biological psychology, vol. 84, 2010, pp. 394-421.

[7] R. O. Duda, P. E. Hart, and E. G. Stork, Pattern Classification, 2nd ed-4th print, John Wiley and Sons, Inc., New York, 2000.

[8] P. D. Wasserman, Advanced Methods in Neural Computing, New York, Van Nostrand Reinhold, 1993.

[9] https://www.biopac.com/

[10] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Wadsworth, 1984.

[11] J. J. Gross and R. W. Levenson, “Emotion elicitation using films,” Cognition and Emotion, vol. 9, 1995, pp. 87-108.

[12] O. M. Doyle, I. Korotchikova, G. Lightbody,W. Marnane, D. Kerins, and G. B. Boylan, “Heart rate variability during sleep in healthy term newborns in the early postnatal period,” Physiological Measurement, vol.30, 2009, pp.847-860.

[13] J. Arroyo-Palacios and D. M. Romano, “Towards a Standardization in the Use of Physiological Signals for Affective Recognition Systems,” Proceedings of Measuring Behavior 2008, Maastricht, The Netherlands, August 2008, pp. 121-124.

 ISBN: 978-1-61208-340-7