IJRESM
International Journal of Research in Engineering, Science and Management
Volume-1, Issue-9, September-2018
www.ijresm.com | ISSN (Online): 2581-5782

Implementing Emotion Recognition Algorithm Using Physiological Signals and Enhancing Classification Accuracy for Possible Application in Depression Diagnosis

Bala Priya C
Student, Department of ECE, Coimbatore Institute of Technology, Coimbatore, India

***Abstract*—Around 1,35,000 Indians commit suicide every year due to depression out of which 12.5% are from Tamil Nadu. Emotion recognition plays a pivotal role in depression detection. The proposed system aims at recognizing emotions automatically by pre-processing the physiological signals, applying feature extraction followed by classifier training and analyzing classification accuracy. From literature, it has been inferred that Support Vector Machine (SVM) is widely used for classification because of its high recognition rates and superior performance to Bayesian and Regression-based classifiers. A survey has been made on the available databases: DEAP, MAHNOB-HCI and Eight Emotion Sentics Data. The physiological signals; Electromyogram (EMG), Blood Volume Pressure (BVP) and Galvanic Skin Response (GSR) from Eight Emotion Sentics Database were considered and statistical and spectral features were extracted. A classification accuracy of 75% was obtained for five target emotions, namely, Joy, Grief, Anger, Hate and Reverence. An improved recognition rate of 91% was obtained by using k-fold leave out one cross-validation to reduce the observation dependence of prediction when there is limited training data.**

***Index Terms*—Emotion Recognition, Support Vector Machine, Eight Emotion Sentics Data, k-fold leave out one cross-validation**

I. Introduction

Human Computer Interaction (HCI) is of paramount importance in the world today. Handheld devices, internet kiosks and several mobile applications facilitate interaction using Graphical User Interface (GUI). With Voice UserInterface (VUI) facilitating speech recognition, the interaction between humans and machines has been progressing rapidly. Several approaches to facilitate multimodal interaction across the human-computer interface have led to the possibility of embedding emotional intelligence to the systems.

The process of Emotional Intelligence embedding leverages several techniques from signal processing, machine learning and computer vision. The accuracy of an emotion recognition system depends on its emotion perception capability. Emotion perception is modelled as a function of cognitive appraisal and subjective experience. Emotions can be perceived through a wide variety of modalities such as visual, auditory and physiological sensory processes. The relationship between the physiological signals and the different emotional states has been established in psychophysiology which states that the activation of the Autonomic Nervous System (ANS) is highly dependent on the emotional states elicited. Thus the field of Affective Computing or Artificial Emotional Intelligence involves the study and development of systems that can recognize and interpret human emotions.

An alarming annual suicide rate of 1,35,000 in the country, with the state Of Tamil Nadu contributing to 12.5% of these suicides, there is an undeniable need for depression detection. Early stages of depression can be cured by proper counselling. Emotion Recognition or Detection of human cognitive states is very important to address depression.

Thus a system which automatically recognizes emotional states of an individual using physiological signals will certainly help in mapping discrete human emotions to a certain pattern of physiological signals that can be analysed for the cognitive state that it corresponds to. However, the task of accurate recording, effectiveness of the stimuli in eliciting target emotions and underlying emotions of the subjects that influence the perception of the target emotional state are some of the challenges that should be overcome in implementing an emotion recognition system.

II. Related Work

Even though facial features and prosodic features from speech signals have been proven to aid in recognition of human affect, these modalities often tend to have greater dependence on individuals and are therefore very likely to be exaggerated or under emoted. This can make the system less generic and less reliable. However, physiological signals are more reliable. A survey on related works has been done. A summary of the existing methodology and the inferences from the same are elucidated in the following sections.

Maria S Perez-Rosero has considered the Eight Emotion Sentics Data for the implementation of an automatic emotion recognition system in [2].The classifier design was predominantly aimed at achieving higher recognition rates. A Support Vector Machine (SVM) was trained using the features extracted from the recorded signals corresponding to various emotions. A set of statistical and spectral features were extracted such as mean, variance, kurtosis, spectral power. These features were fed to the SVM as the training dataset. The recognition accuracy was found to be 70% for an SVM with linear kernel.

Choubelia Maaoui and Alain Pruski have implemented emotion recognition using physiological signals in [3]. Physiological signals from several modalities were used. The physiological signals used were Blood Volume Pressure (BVP), Electromyogram (EMG), Skin temperature (SKT), Respiration (Resp) and Electrodermal Activity (EDA).

The features extracted include

1) Mean of raw signals
2) Standard deviation of raw signals
3) Mean of absolute value of first and second differences of the raw and normalized signals.

Support Vector Machine (SVM) and Fisher Linear Discriminant were used as supervised learning models to classify the six discrete emotional states: Amusement, Contentment, Disgust, Fear, Neutrality and Sadness. SVM and Fisher Discriminant yield a subject-wise accuracy of around 92% and 90% respectively. However, the overall prediction accuracy drops to around 45% with SVM and even lower with Fisher Discriminant.

III. PROPOSED MODEL

A. Implementation

The proposed system is based on the use of physiological signals and the subsequent extraction of features and training the classifier to detect the various states of emotion. An attempt has been made to implement an emotion recognition system that uses signals from physiological modalities. These include Blood Volume Pressure (BVP), Electromyogram (EMG) and Galvanic Skin Response (GSR).

The implementation comprises of three phases: Feature extraction, Classifier training and testing the classifier for recognition accuracy.

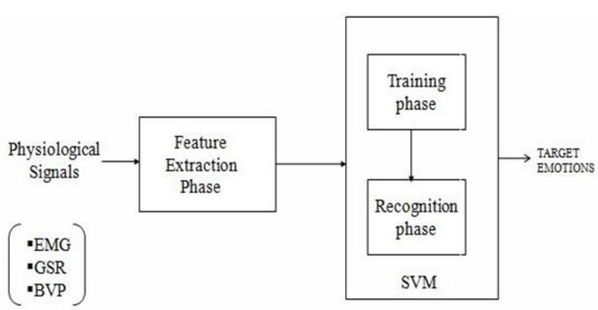

Fig. 1. Block diagram of proposed system

The Fig. 1 shows the various phases in the implementation of the emotion recognition system. The physiological signals considered are Electromyogram (EMG), Galvanic Skin Response (GSR) and Blood Volume Pressure (BVP).EMG is recorded from upper trapezius muscle of the jaw, GSR from electrodermal activity of the skin and BVP as percentage reflectance from the skin. The feature extraction phase comprises of computing various statistical features from the physiological signals. Training and testing the SVM classifier for recognition accuracy follows the feature extraction phase.

B. Workflow

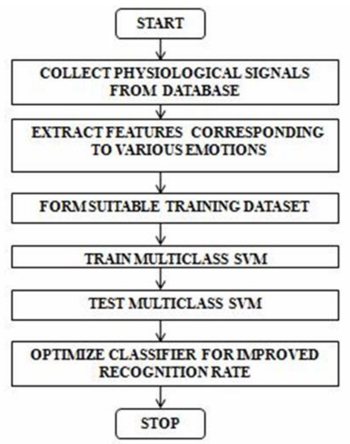

Fig. 2. Physiological-signal based emotion recognition

The Fig. 2 summarizes the various steps involved in the implementation of the emotion recognition system. The target emotions are fixed and the features are extracted from the physiological signals for the various emotions. The extracted features are grouped into feature vectors that form the training matrix. The test vectors are also found in the same way. The SVM classifier is trained and tested for classification accuracy of the target emotions.

C. Database Description

The Eight Emotions Sentics Data is a database of physiological signals created by MIT Media Lab, Affective Computing Group [5]. The emotional states are induced deliberately using suitable arousal techniques and the subject is asked to emote the various emotional states. During the expression of emotions, the physiological signals are recorded.

Data has been collected over a period of 32 days in an office setting. Klynes' Theory of Emotional model covering the following emotions; anger, joy, grief, hate, neutral, love, romantic love and reverence has been chosen. The method of eliciting the emotions and time of recording are maintained the same throughout the experimental duration of 32 days.

The database comprises of two overlapping sets Set A and Set B.Set A comprises of the readings recorded on all the 32 days, from each of the sensors. However, in set A there are failures in one or more of the sensors during the experiment. Detachment of electrodes from the body of the subject, failure of Analog to Digital Converters (ADC) etc. occurred. Set B comprises of twenty good datasets in which none of the electrodes failed during the recording session. Set B has been considered in the work proposed. Anger, Joy, Grief, Hate and Reverence have been chosen as the target emotional classes.

D. Feature Extraction

From literature survey [2],[3],[4], a set of statistical and spectral features which were found to be effective in recognition of cognitive states has been determined. The features extracted are described below.

a)Mean:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (1)$$

where, x_i are the sample values, i ε (1,2,3,…,N)
N is the total number of samples

b)Variance :

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} \quad (2)$$

where, μ is the mean

c)Minimum value :

$$\text{Min\_value} = min\{x_i\} \quad (3)$$

where, x_i are the sample values, i ε (1,2,3,…,N)

d) Maximum value :

$$\text{Max\_value} = max\{x_i\} \quad (4)$$

where, x_i are the sample values, i ε (1,2,3,…,N)

e) Signal Power :

$$\text{Power} = \frac{\sum_{i=1}^{N}|x_i|^2}{N} \quad (5)$$

where, x_i are the sample values, i ε (1,2,3,…,N)
N is the total number of samples

f) Power Spectral Density

$$\text{Spec\_power} = |S_{xx}\ (f)|^1 \quad (6)$$

where, $|S_{xx}\ (f)|$ is the Power Spectral Density (PSD) of $R_{xx\ (\tau)}$

g) Mean of first difference :

$$\check{\mu} = \frac{\sum_{i=1}^{N}(x_{i+1}-x_i)^1}{N} \quad (7)$$

E. Classifier Training

Support Vector Machine (SVM) with linear kernel has been chosen as the classifier.

a) Linear Kernel SVM

Given a set of labelled training samples, $\{(x1,1), (x2,y2), (x3,y3), \ldots(xl,yl)\}$, where yi ε {1, -1} corresponding to the two classes Class 1 and Class 2.

The SVM linear kernel learns a decision function given by

$$h(x) = \text{sgn}\ (\langle w.x\rangle + b) \quad (8)$$

In equation (8),
x denotes the data point to be classified
w denotes the support vectors
b denotes the bias which quantifies how far the separating hyperplane is from the nearest data point in either class .

The hyperplane of maximal margin is found. Maximal margin is defined as the sum of the distances of the hyperplane from the nearest data point in each class. The optimization of the margin is done by considering a dual optimization problem to find the Lagrange multipliers αi.

Thus, SVM learns non-linear functions of the form

$$f\ (x) = \text{sgn}(\textstyle\sum_{i=1}^{l} \alpha_i\, K(x_i\,, x)\ ) \quad (9)$$

In equation (9),
αi are Lagrange multipliers
(,) is the linear kernel that computes the inner product of the data point to be classified and the *i*th support vector

IV. Experimental Results

A. Windowing out the Target Emotional Classes

Using the index provided in the database, common windows across the physiological signals from various modalities are found for the target emotional classes.

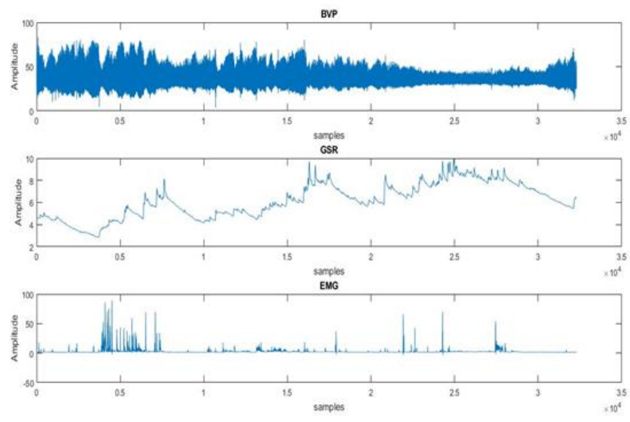

Fig. 3. BVP, GSR & EMG signals corresponding to dataset-1 (Comprising all emotions)

The Fig. 3 shows the waveforms corresponding to BVP, GSR and EMG signals. The waveforms correspond to one entire set of recording, comprising all eight emotions in succession.

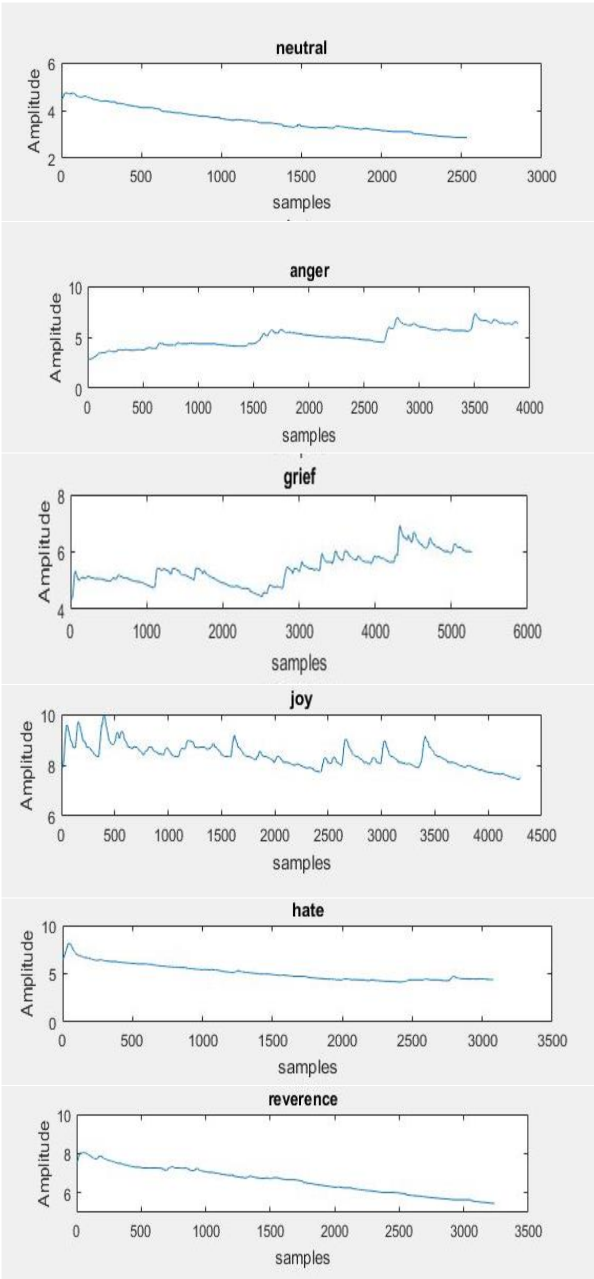

Fig. 4. GSR signals corresponding to various emotions

The Fig. 4 shows the GSR signals that are associated with different emotional states. It can be seen that the GSR signals have comparable peak values during ‘joy’and ‘grief’ and the signals are not sufficiently distinct during ‘reverence’ and ‘hate’ too.

The GSR signals are therefore less likely to aid in emotion recognition for the considered emotional classes. However, GSR related features were included in the training dataset and were subsequently dropped due to deterioration in the classification accuracy.

In Fig. 5, the EMG signals during the target emotions are displayed. It can be seen that the EMG signals are sufficiently distinct for the various emotional classes. It can be seen that the peak value and number of peaks in EMG are substantially higher during ‘anger’ than during any other emotion. The EMG signals during ‘neutral’ and ‘reverence’ are almost similar. The EMG signals are not remarkably distinct for ‘joy’ and ‘grief’. However, they provide better distinction among the emotional classes as compared to GSR signals.

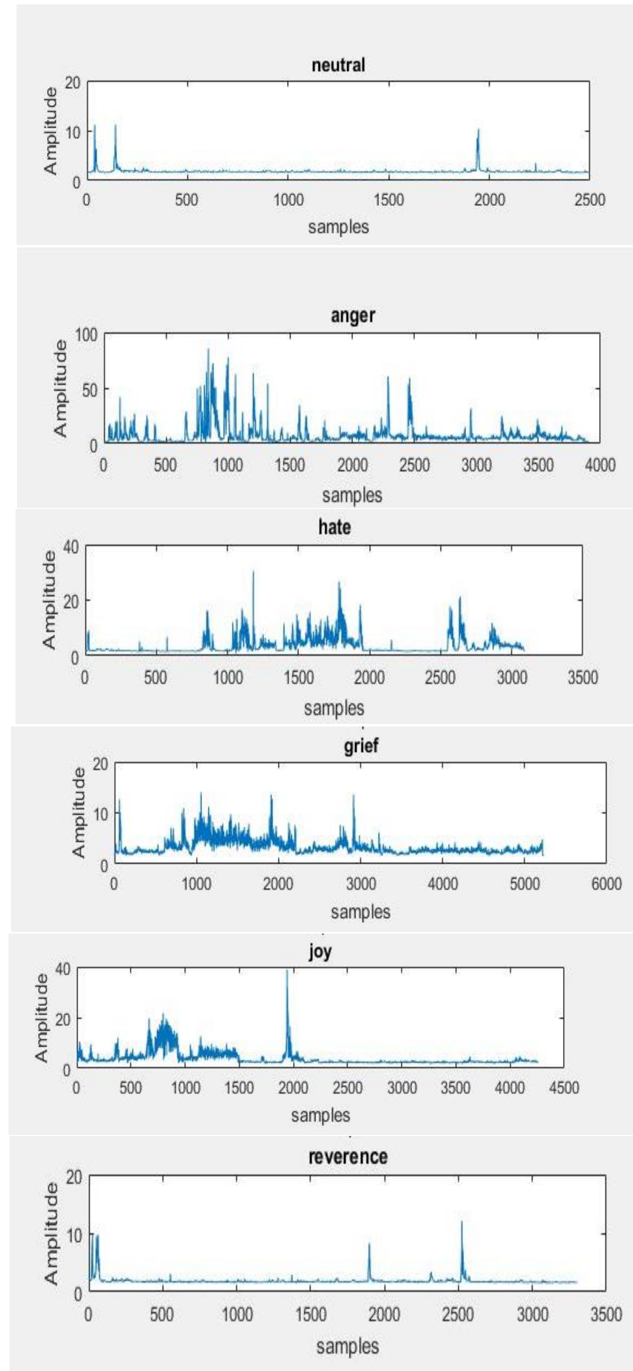

Fig. 5. EMG signals corresponding to various emotions

The Fig. 6 shows BVP signals for the different emotional classes. It can be seen that they are not consistent across observations from the other recordings in the database.

B. Feature Extraction and Classifier Training

The features are extracted from the signal set corresponding to different data sets. The features mentioned in section II.C corresponding to a particular emotion for a particular dataset are grouped to form a feature vector.

The feature vectors corresponding to the datasets chosen to be a part of training set were grouped into a matrix to form the training set. SVM is trained using the training set where each row is mapped onto a corresponding class label, defining the class to which each feature vector belongs. The class labels are grouped into a matrix called train_labels matrix.

C. Recognition Accuracy

The features are also extracted from the target test class. The test vector is predicted and the test label is displayed. The SVM was trained and tested and the results are summarized below

Case (i): Testing Binary SVM (2 Target Emotions 'Joy' & 'Grief') (using BVP, GSR and EMG features)

TABLE I
CONFUSION MATRIX FOR 'JOY' AND 'GRIEF'

EMOTIONS	JOY	GRIEF
JOY	50	50
GRIEF	50	50

From Table-1, it can be found that the recognition rate is only 50%.BVP features were found to be inconsistent across emotions and were therefore not included in further tests.

Case (ii): Testing Binary SVM (2 Target Emotions –'Joy' & 'Grief') (using GSR and EMG features)

TABLE II
CONFUSION MATRIX FOR 'JOY' AND 'GRIEF' (DROPPING BVP FEATURES)

EMOTIONS	JOY	GRIEF
JOY	100	0
GRIEF	0	100

The Table-2, shows that the recognition accuracy is enhanced when BVP related features were discarded from the training dataset.

Case (iii): Testing Binary SVM (3 Target Emotions –'Joy', 'Grief' & 'Anger') (using GSR and EMG features)

TABLE III
CONFUSION MATRIX FOR THREE EMOTIONS: 'JOY', 'GRIEF' AND 'ANGER'

	JOY	GRIEF	ANGER
JOY	100	0	0
GRIEF	0	100	0
ANGER	0	50	50

From Table-3, it can be seen that there are misclassifications between 'anger' and 'grief'. The overall recognition accuracy in this case is 66.67%.

Case (iv): Using EMG and GSR related features (4 Target Emotions)

TABLE IV
CONFUSION MATRIX FOR 'JOY', 'GRIEF', 'ANGER' AND 'HATE'

EMOTIONS	JOY	GRIEF	ANGER	HATE
JOY	100	0	0	0
GRIEF	0	100	0	0
ANGER	50	50	0	0
HATE	0	100	0	0

From Table-4, it can be seen that Hate and Anger are misclassified because of GSR related features not being distinct enough for the target classes.

Case (v): Using EMG related features (4 Target features)

TABLE V
CONFUSION MATRIX FOR FOUR TARGET EMOTIONS

	JOY	GRIEF	ANGER	HATE
JOY	75	25	0	0
GRIEF	25	75	0	0
ANGER	0	0	75	25
HATE	25	0	0	75

From Table-5, it can be seen that an overall recognition accuracy of 75% is obtained when only EMG related features are considered

Case (vi): Using EMG related features (5 Target emotions)

TABLE VI
CONFUSION MATRIX FOR 'JOY''GRIEF', 'ANGER', 'HATE' AND 'RESPECT'

EMOTIONS	JOY	GRIEF	ANGER	HATE	RESPECT
JOY	75	25	0	0	0
GRIEF	25	75	0	0	0
ANGER	0	0	75	25	0
HATE	0	0	25	75	0
RESPECT	25	0	0	0	75

From Table-6, it can be seen that the overall recognition accuracy is still maintained at 75% even when extended to all five target emotional classes.

D. Increasing Recognition Accuracy Using k-Fold leave out one Cross Validation

- The feature vectors corresponding to the 5 target emotional classes are computed.
- The training matrix consists of 100 feature vectors

Steps in k-fold leave out one-cross validation

i. Use (k-1) feature vectors in the training data set
ii. Test the k^{th} data set
iii. Repeat I and ii for the entire feature vector set

When there is limited training data, the problem of overfitting occurs which can be overcome by reducing observation dependence using k-fold leave out one cross-validation.

In k-fold leave out one cross-validation, when a test vector is being tested, it would have been in the training set (k-1) times and exactly once in the test set. This can average out the randomness and mitigate the inconsistencies due to bias or erroneous recording in the observations that go into the training dataset.

TABLE VII
CONFUSION MATRIX FOR INCREASED RECOGNITION RATE

	JOY	GRIEF	ANGER	HATE	RESPECT
JOY	95	05	0	0	0
GRIEF	15	85	0	0	0
ANGER	0	0	90	10	0
HATE	0	0	10	90	0
RESPECT	05	0	0	0	95

From Table-7, it can be seen that the misclassification rate is greatly reduced and a recognition accuracy of 91% is obtained.

V. CONCLUSION AND FUTURE WORK

The emotion recognition algorithms have been implemented using physiological signals. Several techniques including reduction in dimensionality of the feature set, retaining only the most distinctive features across emotions have been adopted to achieve better recognition accuracy of 91%.

Although the physiological signal sources are less susceptible to individual-dependent variations, the recognition accuracy can certainly be increased if features from facial and speech signals are also added to the emotion recognition system.

A multimodal approach that involves fusion of facial, prosodic and physiological features can augment in building a system with greater emotional intelligence. Depression is largely due to the cumulative effect of emotions that are rated negative in valence scale such as grief, anger, hate and distress.

Pattern recognition techniques can be applied to recognize specific recurrent patterns of such emotions whose prolonged occurrence may eventually lead to depression. Such an analysis may aid in estimating a depression quotient. Depression diagnosis followed by proper counselling sessions may help in reducing the suicide rates. A database dedicated entirely towards emotion recognition for depression detection can certainly be of great use for future work in this interdisciplinary field.

REFERENCES

[1] https://en.wikipedia.org/wiki/Suicide_in_India
[2] Maria S Perez-Rosero, "Decoding Emotional Experience through Physiological Signal Processing", *IEEE Conference on Acoustics, Speech and Signal Processing(ICASSP)*, vol. 7, no. 01, pp. 189-200, January-March 2017
[3] Choubelia Maaoui, Alain Pruski, "*Emotion recognition through physiological signals for Human-Machine Communication*", International Journal on Cutting Edge Robotics,vol.01,pp. 318-333, 2010
[4] Verma G.K. and Tiwary U.S., "Multimodal Fusion Framework : A multiresolution approach for emotion classification and recognition from physiological signals", *NeuroImage*, vol.01, pp. 234-245, 2013
[5] https://www.media.mit.edu/groups/affective-computing/
[6] S. Koelstra, C. Muhl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi,T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis Using Physiological Signals", *IEEE Transactions On Affective Computing*, Vol. 3, no. 01,pp. 112-286, January-March 2012
[7] Pascal Ackermann, Christian Kohlschein, AgilaBitsch, Klaus Wehrle and Sabina Jeschke, "EEG-based Automatic Emotion Recognition: Feature Extraction, Selection and Classification Methods",*International Conference on Communication and Distributive Systems*, vol. 8, pp. 49-59, 2012
[8] Shiyu Chen, Zhen Gao and Shangfei Wang, "Emotion Recognition from Peripheral Physiological Signals enhanced by EEG", *IEEE Transactions On Affective Computing*, vol. 8, no. 02, pp. 2827-2831, March 2016
[9] Busso. C, Deng. Z, Yildirim. S, Bulut. M, Lee. C, M, Kazemzadeh. A, Lee. S, Neumann. U, and Narayanan. S, "Analysis of emotion recognition using facial expressions, speech and multimodal information", *Proc. 6th Int. Conf. Multimodal Interfaces*, pp. 205–211, 2014.
[10] Johannes Wagner, Jonghwa Kim, Elisabeth Andre, "From Physiological Signals to Emotions: Implementing and Comparing Selected methods for Feature Extraction and Classification", IEEE Conference on Machine Intelligence, vol. 6, no. 10, pp. 32-80,2010.
[11] Nicholas Cummins, Stefan Scherer, Jarek Krajewski, Sebastian Schnieder, Julien Epps, Thomas F. Quatieri, " A Review of Depression and Suicide Risk Assessment using Speech Analysis", Speech Communication, Elseiver, vol.6, pp. 10-49,2015
[12] Nicolle. J, Rapp. V, Bailly. K, Prevost. L, and Chetouani. M, "Robust continuous prediction of human emotions using multiscale dynamic cues," in Proc. ACM Int. Conf. Multimodal Interaction, pp. 501–508, 2012.