A Novel Method to Detect Inner Emotion States of Human using Artificial Neural Networks

Thejaswini S, K M Ravi Kumar

Abstract: Human computer interaction is a fast growing area of research where in the physiological signals are used to identify human emotion states. Identifying emotion states can be done using various approaches. One such approach which gained interest of research is through physiological signals using EEG. In the present work, a novel approach is proposed to elicit emotion states using 3-D Video-audio stimuli. Around 66 subjects were involved during data acquisition using 32 channel Enobio device. FIR filter is used to preprocess the acquired raw EEG signals. The desired frequency bands like alpha, delta, beta and theta are extracted using 8-level DWT. The statistical features, Hurst exponential, entropy, power, energy, differential entropy of each bands are computed. Artificial Neural network is implemented using Sequential Keras model and applied on the extracted features to classify in to four classes (HVLA, HVHA, LVHA and LVL A) and eight discrete emotion states like clam, relax, happy, joy, sad, fear, tensed and bored. The performance of ANN classifier found to perform better for 4-classes than 8-classes with a classification rate of 90.835% and 74.0446% respectively. The proposed model achieved better performance rate in detecting discrete emotion states. This model can be used to build applications on health like stress / depression detection and on entertainment to build emotional D.].

Keywords: DWT, EEG, Emotion states, ANN, Keras.

I. INTRODUCTION

The requirement and importance of emotion detection using automated methods has increased its role in human computer applications. In human life, emotion is playing a predominant role, detection of such emotion states helps to develop the intellectual BCI systems [1-2]. The emotion states can be detected using facial expressions, speech signals and physiological signals like blood pressure, activity of central nervous system, heart rate etc. There are some problems in detecting emotion states using facial expressions and speech signals. During speech and face recognition, the subject needs to speak or look in the camera direction. Moreover, the emotion states can be controlled or hidden or faked during these methods. Physiological signals can be used to avoid these problems [1,3]. The study of emotion detection using electroencephalogram signal is rapidly growing field of research nowadays. Since it is difficult to influence brain activity, measuring emotions using EEG signals is advantageous. A variety of EEG wireless devices which are portable and easy to use are available in market are helping to develop emotion recognition systems or algorithms which can be used in many applications like health care, entertainment, marketing, decision making etc. [4].

For theoretical emotion representation, various models are proposed by researchers in discrete form and in dimensional form. In the discrete model, a finite number of discrete states are defined corresponding to one of the core emotions, or a combination of them are used to represent different states of emotion. The dimensional model as shown Figure 1 spreads spatially the interpreted levels of emotion states along valence-arousal dimensions. These emotion models have been used for systematic and multilateral analyses of emotion states [5].

The related work implemented on bench marked database: DEAP which is publicly available is discussed below. Sharma et al. [6] proposed a classification technique using non-linear Higher order statics and LSTM algorithm. Their proposed system for Valence-Arousal score attained an average accuracy of 82.01% for 4-classes and 84.68% for 2-class. Guo et al. [7] developed an emotion classification system using a hybrid of SVM and FCM classifier for valence, arousal, dominance and liking states with an average rate of 78.03%. Xing et al. [8] suggested a framework using SAE and RNN-LSTM with classification rate as 81.10 for valence and 74.38% for arousal. Pandey et al. [9] detected happy and sad emotion states happy using MLP with an average accuracy of 58.5% for frontal and theta band. Hongpei et al.
[10] used KNN classifier and achieved 95% in gamma band. Zamanian et al. [11] extracted Gabor and IMF along with time-domain features and using multiclass SVM as classifier, they obtained an accuracy of 93% for 3 and 7 channels. Chao et al. [12] explored deep learning framework achieving a rate of 75.92% for arousal and 76.83% for valence states. Liu et al. [13] classified the data using time and frequency domain features and obtained 70.3% and 72.6%, using SVM. Zeynab et al. [14] classified emotions using EEG signals in arousal-valence dimensions using 10 channels and frequency bands. The frequency bands were decomposed using DWT and features like entropy and energy were extracted. They have used SVM and KNN classifiers to detect the emotions. The accuracy obtained was 86.75% and 84.05% for arousal level and valence level.

Xian et al. [15] proposed SVM emotion detection system effectively using EEG by selecting only 8 channels. They calculated features in time domain, frequency domain and non-linear dynamic features and applied for SVM classifier. Among channel based classifier, channel FP1 obtained an average highest accuracy of 80.01% in arousal and 78.43% in valence states. Adrian et al. [16] implemented an algorithm for emotion recognition via EEG signals using wavelet transform and time-frequency. EEG data was acquired from 22 subjects and the channels used were FP1 and FP2 with one reference channel CZ. To stimulate discrete emotions happy and sad, they have used IAPS images. Mean, standard deviation, wavelet coefficients at db4 and sym6 features were extracted. They used ANN classifier for only two features mean and standard deviation and obtained an accuracy of 72.7% and 81.8% for happy emotion. For sad emotion frequency domain features performed better with an 72.7% accuracy.

Raja et al. [17] developed an algorithm using KNN and SVM to detect happy, scared, sad and clam emotion states. The emotions were elicited using picture (IAPS) database. Raw signals were acquired through 16 channel Emotive device. Extracted Statistical and frequency domain features were given to classifiers whose accuracy was 55% and 58% respectively. Adnan et al. [18] analyzed the music effect on human emotion states using EEG signals recorded from neurosky single channel headset for 32 audio tracks. For MLP, SVM and KNN classifier the input was extracted features of time, wavelet and frequency and obtained an accuracy of 78.11%, 75.62% and 72.8%.

From the review it is found that major work is carried out using available data base (DEAP). Some of the researchers acquired real time signals using less number of channels (2-16 channels). In both the cases signal is acquired by showing video or IAPS images. From the literature it is seen that the emotion recognition systems were limited to few discrete emotions states and some of them related these discrete states with valence-arousal model for their work. Our work focuses on developing a novel approach to elicit eight different emotion states and they are mapped across valance and arousal model.

II. METHODOLOGY

In the proposed work, a different approach using VR audio-video clips are used to elicit emotion states. The statistical and wavelet features are extracted and classified emotions into four states (HVLA, HVHALV, LVHA and LVLA) as well as 8 discrete emotion states (relax, calm, happy, excited, sad, tensed, fear, and bored) using Artificial Neural Network. The proposed methodology is shown in Figure 2.

![Figure 2: Proposed Methodology](image)

A. Experiment Set Up

From 66 healthy subjects (36 males and 30 female) whose age is between 20 to 45 years with mean age of 35, EEG data is acquired using 32-channel Enobio device with a sampling rate 500 samples/sec. The volunteers are UG/PG students as well as research scholars of SGGSIT, Nanded, India. For each subject, before starting the experiment, they were made familiar with the protocol which is used and are asked to fill concern form. Once the EEG Data acquisition device is worn, the subject is instructed to relax for about 60 sec.
The protocol for data acquisition is based on a little modification on the procedure which was used by available data base [19-20]. In DEAP data base [19], they had used 40 one-minute music video to elicit emotion states and have acquired data for 32 subjects. In SEED data base [20], 15 Chinese videos were used to elicit emotion for negative, neutral and positive states for 15 subjects. In the present work a novel technique is used to elicit emotion states by making the subjects to experience VR 3D-360 videos using VR virtual glasses (IRUSU PLAY VR headset). The advantage of using VR videos is to express emotions better [21]. Eight VR videos are used to stimulate emotion states. The video clips to elicit the eight emotion states relax calm, Joy/ excited, happy, tensed, fear, sad and bored are selected. The experiment set up for real-time signal acquisition using Virtual Reality Videos is illustrated in Figure 3 and the time duration of each clip is 50 to 300 seconds depending on the emotion state. The 32 electrodes used for the data acquisition are CP2, CP1, FPZ, CP5, T7, CP6, C3, P3, CZ, P4, O1, P7, FC5, P8, PO3, O2, PO4, T8, FP1, FC1, FP2, FC2, F7, FC6, F3, F4, FZ, AF3, F8, AF4, CZ. The reference electrodes used are CMS & DRL [22]. The self-assessment is done using a questionnaire session for 30sec after each emotion clip, where each subject assesses, his / her emotion state for both the experiment set up.

![Figure 3: Protocol for data acquisition](image)

The experiment set up and instructions are listed below. The protocol has 8 trails to elicit 8 emotion states like relax, calm, joy, excited, tensed, fear and sad and bored. After the Enobio device EEG cap is placed, the subject is asked to close eye for 1 min. This state is called Relaxation stage [RS]. Then the subject is made to experience emotion states by watching 8 audios –video clips using VR device. He /she is advised not to do body movement. Eight trails are presented and each trail includes the following:

- Base line: displays trail number.
- Audio-video clip: the subject is instructed to watch VR-videos of 3-4 min duration.
- Self-assessment: after each video clip the subject is asked to fill the self-assessment form and is informed to close his/her eyes there by relaxing themselves. [22].

The process of experiment carried out using VR device is illustrated in Figure 4.

**B. Preprocessing**

The raw EEG signal has some artifacts like eye blinks, muscle movement etc. The power line interface is eliminated using 50 Hz notch filter during data acquisition only. To improve the quality of signal to noise ratio, moving average referencing technique is applied. The signal is further preprocessed using a 20th order band pass FIR filter allowing frequencies from 2 to 50 Hzs. To use zero phases, “filtfilt” command of MATLAB is used [23-24].

![Figure 4: Data acquisition using VR device](image)

**C. Feature Extraction**

For event separated preprocessed data [24], features such as time, wavelet features in time-frequency domain listed below are extracted. From the related work explored, it was observed that resolution of wavelet transforms was good compared to FFT or STFT. In the present work we have applied 8-level decomposition using DWT to extract desired frequency bands gamma (31-50Hz), beta (14-30Hz), delta (1-3Hz), theta (4-7Hz) and alpha (8-13 Hz) [24]. The wavelet features for all five bands of 32 channels are extracted as shown in Table - I. The features like differential entropy, PSD, power, average energy of all 5 bands (e_a, e_t, e_b, e_g and e_d ) are calculated.

The channel wise features listed below in Table –II are computed using MATLAB and stores separately so that each channel features are applied for channel wise classifier. The scatter plot of a few features of happy, angry sad and calm for Power of alpha band (blue), mean (red), energy of beta band (yellow) and Hurst Exponential (green) is presented in Figure 5. The features are calculated based on the formulae listed in Table –III.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time domain (Statistical Features)</td>
<td>Mean, Root mean square, Std-deviation, First &amp; second difference, Normalized first &amp; second difference, Skewness, Kurtosis, Variance, Mobility and Complexity (Hjorth Parameters)</td>
</tr>
<tr>
<td>Wavelet or Time-Frequency Domain features of 5 band</td>
<td>Band energy, Differential entropy, Power spectral density, Average band power</td>
</tr>
<tr>
<td>Other features</td>
<td>DASM, RASM, Hurst exponential and Permutation entropy</td>
</tr>
</tbody>
</table>

**Table -I: 8-level decomposition using DWT**

<table>
<thead>
<tr>
<th>SI. No.</th>
<th>Decomposition Coefficients</th>
<th>Frequency band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CD5</td>
<td>Gamma</td>
</tr>
<tr>
<td>2</td>
<td>CD6</td>
<td>Beta</td>
</tr>
<tr>
<td>3</td>
<td>CD7</td>
<td>Alpha</td>
</tr>
<tr>
<td>4</td>
<td>CD8</td>
<td>Theta</td>
</tr>
<tr>
<td>5</td>
<td>AD8</td>
<td>Delta</td>
</tr>
</tbody>
</table>
A Novel Method to Detect Inner Emotion States of Human using Artificial Neural Networks

Figure 5: Scatter plot of features for a 4 emotion states of four frequency bands.

Table –III: Feature calculation [23]

<table>
<thead>
<tr>
<th>Features</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\mu = \frac{1}{T} \sum_{t=1}^{T} s(t)$</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$\sigma = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (s(t) - \mu)^2}$</td>
</tr>
<tr>
<td>First difference</td>
<td>$\delta = \frac{1}{T-1} \sum_{t=1}^{T-1}</td>
</tr>
<tr>
<td>Normalized first difference</td>
<td>$\delta' = \frac{\delta}{\sigma}$</td>
</tr>
<tr>
<td>Second difference</td>
<td>$\gamma = \frac{1}{T-2} \sum_{t=2}^{T-1}</td>
</tr>
<tr>
<td>Normalized second difference</td>
<td>$\gamma' = \frac{\gamma}{\sigma}$</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>$k = \frac{(t - \mu)^4}{\sigma^4}$</td>
</tr>
<tr>
<td>Mobility</td>
<td>$\sqrt{\frac{\text{variance}(s(t))}{\text{variance}(s(t))}}$</td>
</tr>
<tr>
<td>Complexity</td>
<td>$\sqrt{\frac{\text{mean}(s(t))}{\text{mean}(s(t))}}$</td>
</tr>
<tr>
<td>Differential Entropy</td>
<td>$\mathcal{G}(x) = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \log_{2} \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \right) dx$</td>
</tr>
</tbody>
</table>

D. Classification

The extracted features are classified using basic deep learning algorithm implemented using KERAS [Sequential ANN]. The classifier is implemented to classify the extracted features into four classes (HVLA, HVHALV, LVHA and LVLA) as well as 8-discrete classes. The classifier for each channel is implemented using Keras library in python. In this paper 66 subjects each of 8 trials are used for the classification. So totally 66*8 = 528 trials are obtained from which 70% trials are used in training and 30% for testing. So 47*8 = 376 trails are used to train the classifiers.

Remaining 19*8 = 152 trails are used to test the trained network. The classifiers are implemented for Time-domain features, Time – frequency features and even on DASM & RASM features separately. A combined feature is also fed to the classifier. The concept of dropouts at a rate of 0.1% and cross validation of 15-fold parameter tuning is used.

The Algorithm involved in building ANN is as indicted below.

Step 1: Importing the feature vector
Step 2: Splitting the 70% feature vector as training set and Remaining as testing set using train_test_split command
Step 3: Feature scaling using Standard scaler
Step 4: Building an ANN using sequential model with input layer, hidden layer and output layer
Step 5: Making prediction and evaluating the model and saving the same for further testing of untrained signals
Step 6: predicting the test results and extracting confusion Matrix
Step 7: Repeat step 1 -6 for all 32-channels
Step 8: Calculate the average prediction rate.

In this work, in order to classify sequence data, sequential classifier is used to build ANN. By importing Keras library using Tensor flow at backend 2 models are used to build ANN. One model is sequential mode which is used for initializing ANN and is done by sequence of layers using the function Sequential ( ). The other is dense model through which different layers of ANN are built. The proposed ANN architecture uses Keras sequential model and it consists of 3 layers input layer, output layer and one hidden layer. The input nodes depend on size of the feature matrix. The activation function used for input and hidden layer is “ReLU”. The size of the output nodes depends upon number of classes i.e 4 or 8. The activation function used for output layer is “softmax”. We have used 16 nodes in hidden layers.

III. RESULT AND DISCUSSION

In the present work, signals acquired from 66 subjects are used for analysis. Totally, from 66 * 8 = 528 trails, 70% is used for training the model and remaining 30% to test the trained model. The 4 emotion states are classified along 2-dimensional models as HVLA, HVHALV, LVHA and LVLA and eight discrete emotion states such as relax, calm, joy, excited, tensed, fear, sad and bored.

The extracted 14 time-domain features, 20 time-frequency domain features and total of 34 complete features are applied to ANN classifier. Ten-fold cross validation and drop out of 0.1% is carried out. A separate classifier is developed for each channel and at the end channel fusion is carried out. The accuracy of the classifier is computed based on the average accuracy rate of all the channels. The percentage of prediction rate for 4 class by the models are 90.835%, 87.1% and 70.14% for all features, wavelet features and time domain features respectively.
Similarly, for 8-class, the classifier accuracy is 74.044%, 68.61% and 48.67% respectively and is depicted in Table IV and Figure 6.

Further channel reduction is applied for observing the behavior of channels on emotion states. Channels are reduced based on pre-frontal, parietal, temporal and occipital electrodes. For channel reduction from 32 to 20 channels FP1, C3, FP2, FOZ, F3, CZ, F4, FZ, F7, T7, P4, T8, C4, P3, O2, P7, P8, PZ and O1 channels are selected. Further 8 channels such as FP1, FP2, FPZ, P3, P4, PZ, T7 and T8 are analyzed. This was followed in reducing further to 5 channels such as FP1, FP2, P3, P4, and PZ. The classification rate for channel reduction is tabulated in Table V and plotted in Figure 7.

It is observed from the channels performance that, the emotions are reflected mostly in frontal electrodes of alpha and beta band. The performance of five channels is illustrated in Figure 8.

As channel reduction is done the classification accuracy also increases for 4 classes but however for 8 classes it is observed that use of 20 channels gives a better performance.

Table IV: Performance of ANN classifier

<table>
<thead>
<tr>
<th>Features</th>
<th>4-Classes</th>
<th>8- Emotion states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Time and wavelet features</td>
<td>90.835%</td>
<td>74.0446%</td>
</tr>
<tr>
<td>Wavelet Features</td>
<td>87.1%</td>
<td>68.61%</td>
</tr>
<tr>
<td>Statistical Features</td>
<td>70.14%</td>
<td>48.67%</td>
</tr>
</tbody>
</table>

Figure 6: Prediction rate of ANN based on features

Table V: Performance of ANN classifier after channel reduction

<table>
<thead>
<tr>
<th>No. of channels selected</th>
<th>4-Classes</th>
<th>8- Emotion states</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 channels</td>
<td>90.835</td>
<td>74.045</td>
</tr>
<tr>
<td>20 channels</td>
<td>89.813</td>
<td>89.815</td>
</tr>
<tr>
<td>8 channels</td>
<td>93.57</td>
<td>77.62</td>
</tr>
<tr>
<td>5 channels</td>
<td>93.048</td>
<td>76.48</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

In this paper protocol for data acquiring is done using a novel technique to elicit emotion states using VR device. For 66 volunteer subjects, EEG signals were recorded through 32 channels of 500 Hz sampling rate. A Deep Learning Neural network based on sequential Keras model is developed to detect 4-classes of two dimensional states and 8-discrete emotion states from the acquired data.

The statistical and Time-frequency domain using DWT features, entropy features are used for ANN classifier. The accuracy for proposed system is 90.835% and 74.0446% for 4 class and 8 class respectively. Channel reduction method is incorporated and observed that for discrete emotion states more number of channels are required for better prediction results.

Future more deep learning algorithms for time series data can be explored. This trained model is performing better for the acquired data. The model can be tested on online available data there by developing a robust model and can be tested on various others classifiers for better classification rate. The model can be used to build for different applications.

ACKNOWLEDGMENT

The Authors would like to thank, Dr. Manesh B Kokare, Dean Research and Development Associate Professor and PI of Center of Excellence in Signal and Image Processing and Dr. D D Doye, Professor, Department of Electronics and Telecommunication SGS Institute of Engineering and Technology, Vishnupuri, Nanded, Maharastra, India, for permitting to acquire EEG data as well sharing the same.
REFERENCES


21. Alice Chiricolo, Pietro Cipresso 2, David B. Yaden 3, Federica Biasionni1, Giuseppe Riva1,2 & Andrea Gaggioli1, “ Effectiveness of Immersive Videos in Inducing Awe: An Experimental Study”, Scientific Reports | 7: 1218 | DOI:10.1038/s41598-017-01242-0 Scientific Reports volume 7, Article number: 1218 (2017)


AUTHORS PROFILE

Mrs. Thejaswini S, is working as Assistant Professor, Department of Electronics and Telecommunication Engineering, B M S Institute of Technology and Management, Bengaluru, Karnataka, India. She has completed her BE in E&CE and M. Tech in digital Communication and Networking. she is pursuing her interest is BCI Machine Learning & signal Processing. She has a teaching experience of 18 years thejaswini79@gmail.com

Dr. K M Ravikumar, is working as Principal and Professor Department of Electronics & Communication, Engineering, S J C Institute of Technology, Chickballapur, Karnataka, India. He has obtained his Ph.D. in the area of Digital signal processing. He a has around 22 years of teaching experience and more than 10 years of research experience in the fields of speech processing, Brain Computer Interface, Image processing domains. He has more than 20 publications in peer reviewed Journals and more than 15 papers in International Conferences. Some of his achievements Are Special Officer, VTU-Regional, Expert Committee Member for National Board of Accreditation (NBA), New Delhi, Reviewer: Elsevier Editorial Systems, Member LIC (Local Inquiry Committee), VTU, Belgaum. He has received various awards and few are Bharat Vidya Shromani Award, Best Engineering College Principal of the year- 2016, Rashtriya Vidya Gaurav Gold Medal Award etc. kmravikumar75@gmail.com