Features Extraction and Depression Level Prediction by Using EEG Signals

[1]Renu Gautam , [2]Mrs. Shimi S.L

[1]ME Scholar, Electrical Engineering DepartmentNITTTR, Chandigarh
[2]Assistant Professor Electrical Engineering Department NITTTR, Chandigarh

---***---

Abstract - Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest. It affects how you feel, think and behave and can lead to a variety of emotional and physical problems. Depressed people may feel sad, anxious, empty, worthless, guilty, irritable, or restless. They may lose interest in activities that once were pleasurable, experience loss of appetite or overreacting, or problems concentrating, remembering details or making decisions. According to the WHO in India 36% (data of 2016) of population suffering from depression. Day by day it is becoming global crisis. It is estimated that by the year 2020 if current trends for demographic and epidemiological transition continue, the burden of depression will increase to 5.7% of the total burden of disease and it would be the second leading cause of disability-adjusted life years.

Severe cases of depression interfere with the common live of patients. For those patients a strict monitoring is necessary in order to control the progress of the disease and to prevent undesirable side effects. Diagnosing depression in the early curable stage is very important. Electroencephalogram (EEG) signals are obtained for publicly available database are processed in MATLAB. This can be useful in classifying subjects with the disorders using classifier tool in it.

Primarily the EEG signals were read using EDF browser software and the signals were fed into MATLAB to get log power spectral density from EEG bands. The features are extracted from different frequency bands.

Keywords: EEG, Depression disorder, Depression recognition, MATLAB

1. INTRODUCTION

With the increasing pace of life and stress of work, more people are growing mental depression disease that is harmful physically and mentally. However, traditional medicines do not have good effect on depression disease, meanwhile have relatively big side effects. Therefore, Depression is one of the common types of mental disorder worldwide. There are almost 350 million people suffering from depression globally. It is categorized into three types: mild depression, moderate depression & major depression.

It is classified that If the person has persistent at least few symptoms for two weeks or more. However there are several types of depressive disorder. If you have been diagnosed with clinical depression, you may be having trouble getting to sleep. There is a reason for that. There is a definite link between lack of sleep and depression. In fact, one of the common sign of depression is insomnia or an inability to sleep [1].

Depression symptoms include: an increase or decrease in the need for sleep. The relationship between sleep and depressive disorder is complex, depression may cause sleep problems and sleep problems may cause or contribute to depressive disorder. For some people, symptoms of depression occur before the onset of sleep problems.

Mood disorders are inherently related to emotion. In particular, the behavior of people suffering from mood disorders such as unipolar depression shows a strong temporal correlation with the affective dimensions valence and arousal. In addition, psychologist and psychiatrists take the observation of expressive facial and vocal cues into account while evaluating a patient condition. Depression could result in expressive behavior such as dampened facial expressions, avoiding eye contact, and using short sentences with flat intonation [2].

A human brain is the organ that gives the person the capacity for art, language, rational thoughts and moral judgments. It is also responsible for each individual's personality, movements, memories and his perception about the world. It is one of the body's biggest organs, consisting of some 100 billion nerve cells that not only put together and highly coordinated physical actions but regulate our unconscious body processes, such as digestion and breathing.

EEG signal is an important information source of underlying brain processes. The communication based on EEG between human brain and computer is a new modality of human-computer interaction. With the improvement of science and technology, people know the deepening of brain function and the appearance of high performance peripheral component (PC) [3]. The era of EEG began with the pioneering work of Hans Berger (1929). Hans Berger was a German psychiatrist. Who spent some years thinking about electrical activity generated by human brain and how could be recorded. The researchers have found that these signals indicate the brain function and status of whole body. Therefore EEG signal provides valuable information of the brain function and neurobiological disorders as it provides a visual display of the recorded waveform and allows computer aided signal processing techniques to characterize them [5].

Electroencephalography (EEG) is one effective way of recording and analyzing the brain activity due to its ease of use, affordable cost and fine resolution. The term EEG refers that the brain activity emits the signal from head. Electroencephalogram (EEG) signal is the recording of spontaneous electrical activity of the brain over a small period of time. It is produced by bombardment of neurons within the brain. It is measured for a short duration of 20-40 minutes with the help of placing multiple electrodes over the scalp [3].
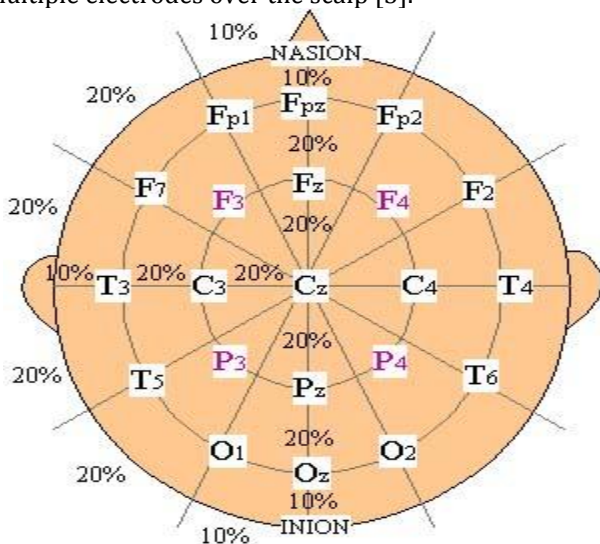

Figure 1.1 Electrode Positions for the Placement of 21 Electrodes [4]

To record either evoked or ambient neural activity, an assembly of surface electrodes placed at standardized locations on the scalp known as 10/20 system. The positions are measured on the scalp relative to the known skull landmarks, from the fronts at nasion to the inion at the back of the head and side to side from the two ear canels. The electrodes are from 5 to 10 mm

diameter and for the macro recording of EEG electrode size in this range is not factor in determining signal characteristics [4].

The 10/20 system or International 10/20 system is an internationally recognised method to describe the location of scalp electrodes. The system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The numbers '10' and '20' refer to the fact that the distance between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull. No central lobe exists; the 'C' letter is used for identification purpose only. The 'Z' zero refers to an electrode placed on the mid line. Even numbers (2, 4,6,8) refer to electrode positions on the right hemisphere. Odd numbers (1,3,5,7) refer to electrode positions on the left hemisphere.

The aim of this work is to find the various depression levels in subjects. During the experiment, EEG data was recorded in order to process, format and evaluate against different machine learning techniques for the classification of EEG data. The experiment helped to confirm the selection of right features and technique, i.e., that providing the most accurate results. The experiment involved planning, preparation, execution and reporting the results as explained in Figure 1.2
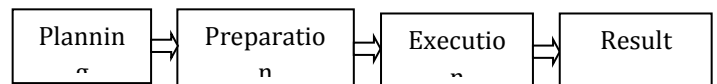

Figure 1.2 Procedures for the EEG signal processing

2.AN OVERVIEW OF EEG SIGNALS

Large numbers of literature are available on Depression and depressive disorder based on EEG. Only EEG is a basic tool which is used to extract the features from the brain waves. But only few are present which combines them. Outcome of some of the literature are discussed briefly in the following paragraphs.

Electroencephalography EEG is the study of electric activity of the brain. The first attempt at measuring this activity was made in 1875 by a British Physician named Richard Caton. Afterward, advancements in neurophysiology were made throughout all of Europe, but slowed to a crawl during both World Wars. After the Second World War, the United States took the lead in EEG research, and the American EEG Society was founded in 1947. In the decades that followed, EEG research in both Europe and America flourished

and every major university hospital had an EEG machine by 1950s [13].

EEG signal acquisition system has to fulfill some basic conditions. To ensure patient safety the device has to be galvanically disconnected from his body. Bio-Semi Active-Two System is a widely used system to achieve advanced EEG data and uses an active-electrode technology with open architecture [9]. The system is an improved version of Active-One and purely designed for research purposes. It features multi-channel and supports high resolution measurements for bio-potentials. With the help of active electrode technology, requirements like impedance measurement, skin preparation and faraday cage are not required to be used [6].

The Active-Two system sets new standards for multi-channel, high resolution bio-potential measurement systems for research applications. The system is a further development of our successful Active-One system, the first commercially available system with active electrodes. Advances in technology have allowed us to significantly increase the number of channels, digital resolution, input range, and sample rate, without any increase in size, power-consumption or costs. Second generation active electrodes are smaller in size with less cable weight, while offering even better specs in terms of low-frequency noise and input impedance.

The engineers Robert Honsbeek, Ton Kuiper and physicist Coen Metting van Rijn have introduced BioSemi system for researches in 1998 at Medical Physics department of the University of Amsterdam.
BioSemi ActiveTwo System used in the experiment involved the following components:
• BioSemi Active-Electrodes
• BioSemi Headcaps
• ActiveTwo AD-box
• USB2 Receiver
• ActiView

Figure 2.1 BioSemi ActiveTwo System

EEG signals are primarily analyzed by their frequency content. That is, the interpretation of the EEG signal is based on the power of the frequencies that it contains. The primary range of interest for EEG signal is from 1 to 100 Hz. Five main frequency ranges are normally included in all EEG studies: Delta (0.5 – 4.0 Hz), Alpha (4.0-8.0 Hz), Beta (8.0-12.0 Hz), Theta (12.0-30.0 Hz) and Gamma (30.0-100.0 Hz).

Most of studies on depression have been based on linear methods (i.e., amplitude, power, frequency or wave symmetry of brain waves [18] [19], or brain activities in specific frequency bands [20]) and there are few on quantitative nonlinear analysis of brain activity. Here, a quantitative wavelet based analysis of EEG which is a suitable nonlinear method to evaluate EEG signals for studying the brain state under different mental conditions.

Nonlinear changes of brain dynamics, in accordance with clinical improvement, at least to a correspondence between a mental state (depressive mood) and a cerebral neuronal state (brain dynamics characteristics) [21]. The brain activity changes during cognitive tasks, e.g. mental arithmetic showed an asymmetrical increase of EEG spectral entropy. And since depression disease could be associated with abnormality in the neuronal activity [22], we speculated that there were entropy difference in EEG activity between depression patient and healthy people, and the group effect may be different under different cognitive conditions.

In [23] have proposed the mobile care mediation system that enables depressed individuals and caregivers to share mood and availability information, thus improving communication in care giving and care-receiving and reducing the care giving burden. They prepared phone-based ESM system was prototyped on a Dopod p800w smart phone, To simplify user entries on a phone and reduce the amount of time needed to answer these questions, input for the ESM questions is multi-modal: radio buttons for single-choice questions, user interaction with mobile devices and applications, and supports automatic logging of device states and sensor data.

In [24] have expressed either verbally through emotional vocabulary, or by expressing non-verbal cues such as intonation of voice, facial expressions and gestures. Most of the contemporary human-computer interaction (HCI) systems are deficient in interpreting this information and suffer from the lack of emotional intelligence. In other words, they are unable to identify human emotional states and use this information in deciding upon proper actions to execute. The goal of affective computing is to fill this gap by detecting

emotional cues occurring during human-computer interaction and synthesizing emotional responses.

3. METHODLOGY

Step 1   First EEG signals will be obtained from subject.

Step 2 EEG signal from the EEG machine is available in the EDF format (European Data Format). It is first converted into .Wav format which is suitable for processing.

Step 3 The software EDF2WAV is used for this purpose. EDF browser is an open source free viewer and toolbox for EEG data files it is used for the storage of multichannel biological and physical signals. It has various options such as converting one file format to another, annotation/event manipulation, header editing , heart rate detection, data reduction and cropping down sample signals.

Step 4 These WAV files can be directly referred into MATLAB programs, as input data.

Step 5 Compare the testing data with the training data.

EDF browser is an open source, free viewer and toolbox for EEG data files [7]. EDF is an abbreviation of European Data Format which is a file format for storage of multichannel biological and physical signals. It was developed by a few medical engineers which later become de-facto standard for EEG and PSG recording in commercial equipment and multicenter research projects

The pre-processed data was further processed to get the real values for the signals using EEGLAB Toolbox. MATLAB having real value for signals make it easier to analyze and classify them. The following features were extracted from the real values of each signal based on findings from [8] in order to process the data further:

• Minimum Value
• Maximum Value
• Mean Value
• Standard Deviation

EEGLAB is a toolbox and graphical user interface that runs under MATLAB (The Mathworks, Inc.) [11]. It can process collection of EEG data for any number of channels. It support various functions of EEG data processing such as importing channel and event information, visualization of data (plus multi-trial ERP-image plots, scalp map and dipole model plotting, scrolling), data pre-processing (including filtering, epoch selection, averaging and artifact rejection), independent component analysis (ICA) and time/frequency decompositions including channel and component cross-

coherence supported by bootstrap statistical methods based on data re-sampling [8].

Three layers are designed to divide the functionalities of EEGLAB. First layer lets the user interact with data using graphical interface without using any syntax of MATLAB. User can also mold the operation of MATLAB according to memory available. Data processing, command history and interactive pop functions can be performed under middle layer.
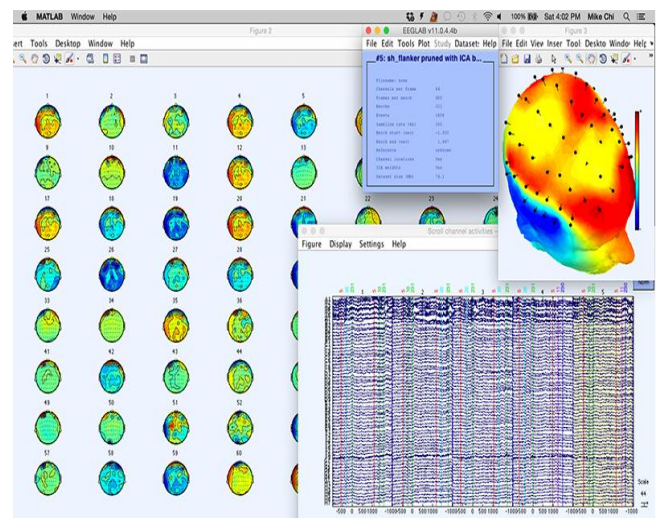

Figure 3.1 EEGLAB Toolbox [12]

EEGLAB data structures and individual signal processing function can be used to write custom scripts by experienced users of MATLAB.

Each dataset having EEG data in MATLAB in dataset into two models: model A and model B. These two dataset will be classified by using two algorithms.

K-Nearest Neighbor (KNN)

KNN is one of the most basic and simple classification techniques [32]. KNN is usually considered when there is no or very little knowledge available for the distribution of data. It is a potential non-parametric classification technique which completely bypasses the problem of probability densities [33]. It is an algorithm that classifies and divides multiple types of data. When given a new test point, the KNN classifier

- Calculate the distance from the test data to all training data points.
- Select k number of training data points that are the closest to the test data point.
- Identifies test data point as same as the most common class among the k nearest training data points.
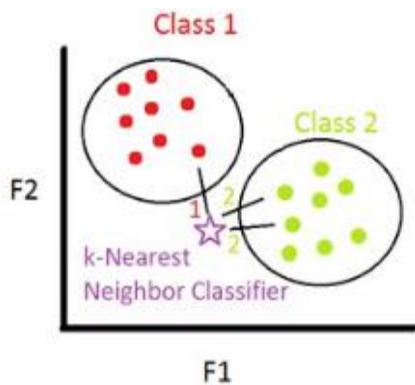
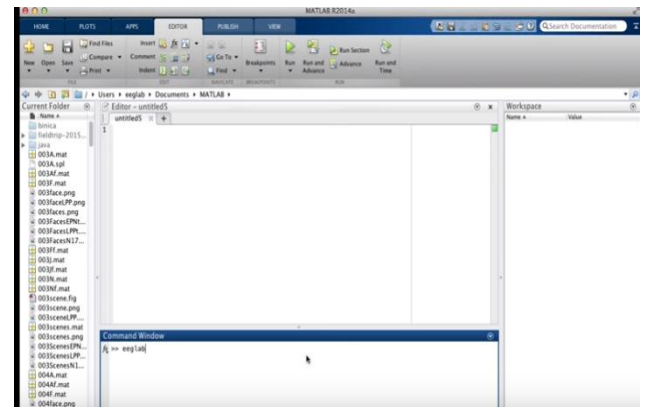Figure 3.2 Shows KNN Classifier for Test and Training Data

Regression Tree (RT)

Regression Tree (RT) takes an object or situation as input which is categorized a number of properties and provides a decision as an output [34]. One input feature is represented by every node and possible test result values are represented by the branches of node. The positive test values can either be positive or negative. If that node is reached, leaf nodes, also called as terminal nodes represent the decision value.

4. RESULT & DISCUSSION

Feature selection is one of the key challenges in affective computing due to phenomena of person stereotype. This is because different individuals express same emotion with different characteristic response patterns for a same situation. Each subject involved in the experiment was having diverse physiological indices that showed high correlation with each affective state. From the obtained EEG data, it was observed that physiological features were highly correlated with the state of arousal among two subjects.

The raw EEG data is then processed to extract the selected features and real value for them. Different signal processing techniques are available for this purpose such as wavelet transform, thresholding, and peak detection. There are some tools available for the same purpose as well such as EEGLAB Toolbox for MATLAB.
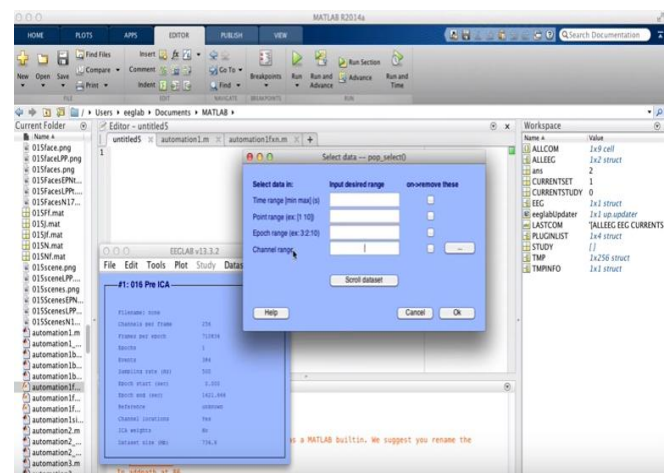

Figure 4.1 EEG raw data loaded MATLAB

Fig 4.2 Show the Selection of Channel of EEG

Fig 4.2 shows that how the channel ranges is selected from EEG waveform. By using ICA channel range are selected single channel and multiple channel can be selected.
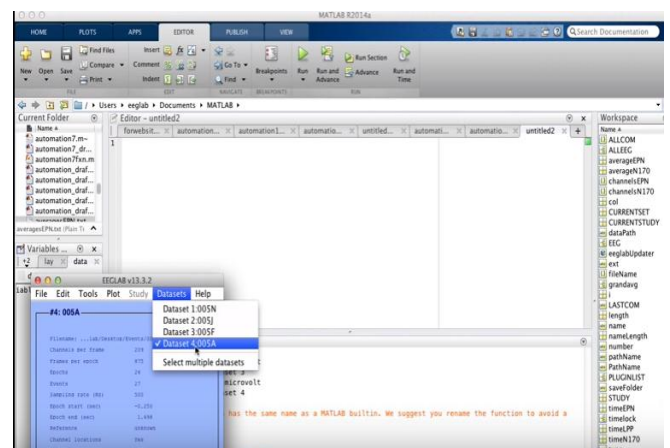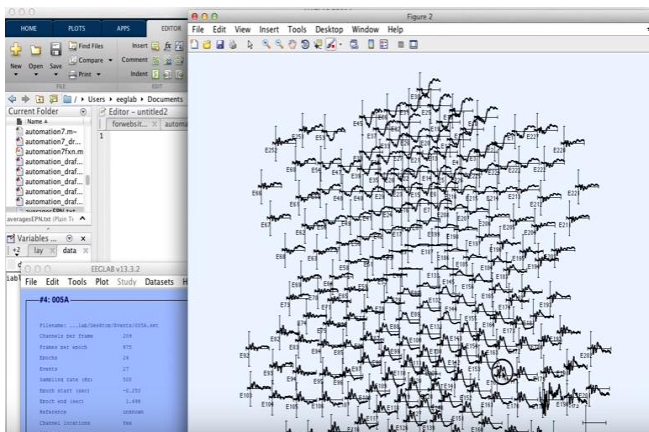

Figure 4.3 Selection of Dataset

Figure 4.4 Selection of particular EEG signal

By using filters in MATLAB low pass filter, high pass filter and notch filter particular band of frequency can be selected. Identifying the EEG data feature in order to achieve better results for classifying EEG data using different the number of emotions taken in consideration. A clear diversity in results was observed when the number of emotions is different even if the same technique is applied for classification of EEG data. Not only this, different dataset for each subjects shows a diverse accuracy results which means the diversity is just not limited to the number of emotions.

All datasets compiled for all subjects were passed to classification software MATLAB one after another for their offline analysis. The selected techniques KNN and RT were used for classification in MATLAB. Each dataset was trained using each selected technique to classify negative or positive arousal/valence values as correctly classified and neutral values as incorrectly classified. The techniques used had all the default parameter values as implemented in MATLAB and 10-fold cross validation (standard value for cross validation). The sequence and description by which datasets were passed to MATLAB are available in Table 4.1:

Dataset	Description
Dataset A	KNN and RT with 10-fold Cross Validation
Dataset B	KNN and RT with 10-fold Cross Validation

Table 4.1 Dataset and their description while passing to MATLAB

Two datasets A and B were passed to MATLAB. Both dataset were formatted using difference models in order to analyze the EEG data in various ways and to find the possible difference among them. The classification accuracies by selected techniques for dataset A are shown:

Techniques	Accuracy
K-Nearest Neighbour	52.88%
Regression Tree	50.00%

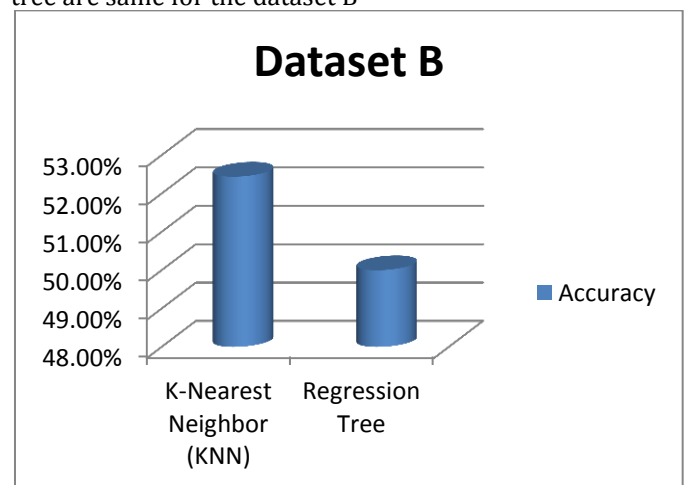Table 4.2 Classification Accuracies by selected techniques for dataset A
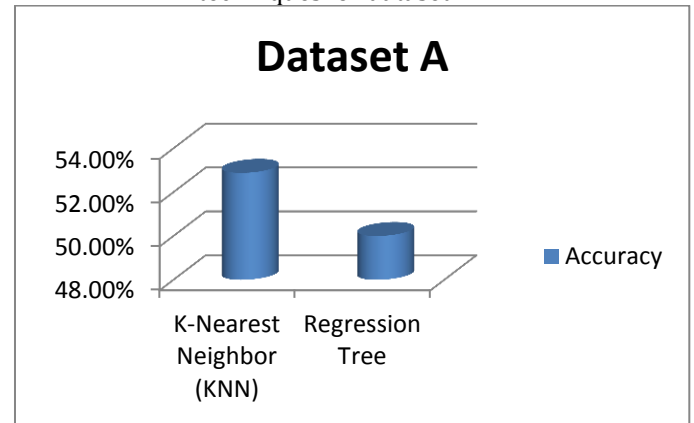
Graph 4.1 Classification Accuracies by selected techniques for Dataset A

Graph 4.1 shows the KNN and regression tree accuracy for the data set A. Graph shows that KNN having better accuracy than the Regression tree for the set A. The classification accuracies by selected techniques for dataset B are shown:

| Techniques	Accuracy
K-Nearest Neighbour	52.00%
Regression Tree	50.00%

Table 4.3 Classification Accuracies by selected techniques for dataset B

In table 4.3 the accuracy for both in KNN and Regression tree are same for the dataset B

Graph 4.2 Classification Accuracies by selected techniques for Dataset B

Graph 4.2 shows the accuracy of dataset B in this KNN has higher accuracy than regression tree
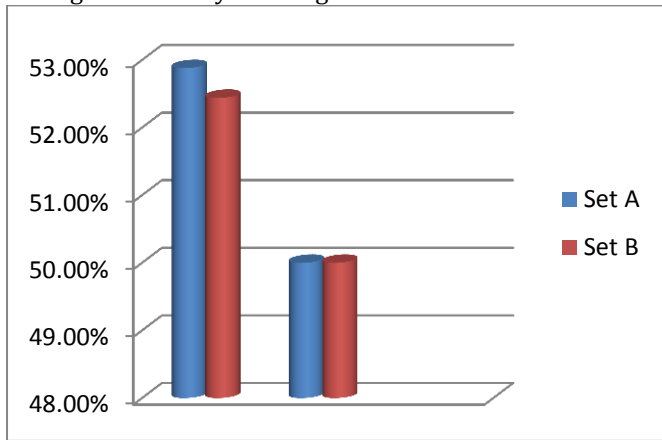

Graph 4.3 Comparison of result obtained from Set A & Set B

Graph 4.3 shows the comparision of data set A and B. In this the KNN shows the higher accuracy than the regression tree. Dataset A accuracy is better in both . Due to the difference in accuracies obtained for using different models for datasets formulation, a comparison of both sets have been carried out in order to identify the reasons and to select a better set to continue the research operations.

Classification of EEG by using KNN Classifier Algorithm						
Estimated Class		**Ground Truth Class**				
		05 Hz	10 Hz	15 Hz	20 Hz	25 Hz
	05 Hz	1	0	0	0	0
	10 Hz	0	1	0	0	0
	15 Hz	0	0	1	0	0
	20 Hz	0	0	0	1	0
	25 Hz	0	0	0	0	1

Table 4.4 EEG Data classification by KNN

Table 4.4 shows the separation the training set from the test set by segmenting the data into 27 numbers of sections. The classifier will test on one section and train the remaining sections. One advantage is this that it prevents over fitting. Same process is done as for set B.

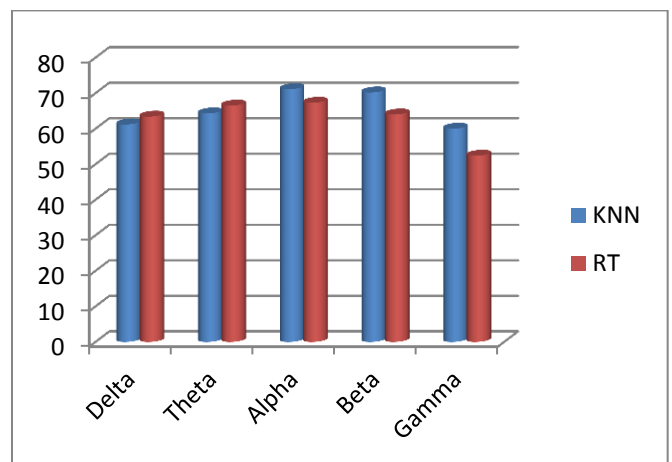| Classifier /Feature	Delta	Theta	Alpha	Beta	Gamma
KNN	65.3	62.3	76.0	66.3	52.5
RT	63.2	68.2	70.0	62.4	45.2

Table 4.5 Results of classification accuracy for each sub band features for set A
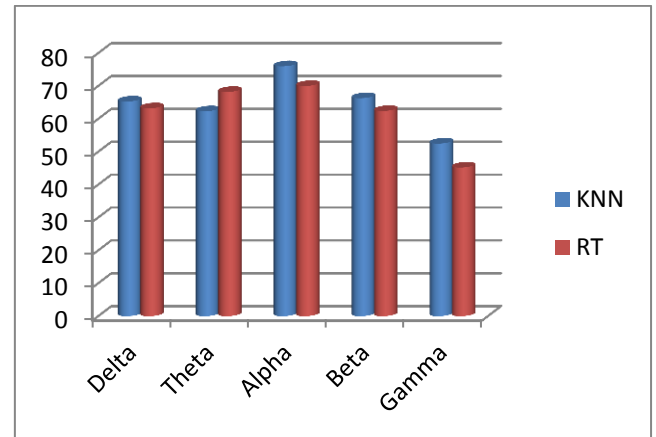
Graph 4.4 Results of classification accuracy for each sub band features for set A

Classifier /Feature	Delta	Theta	Alpha	Beta	Gamma
KNN	61.2	64.3	71.1	70.2	60.0
RT	63.4	66.5	67.3	64.0	52.4

Table 4.6 Results of classification accuracy for each sub band features for set B

Graph 4.5 Results of classification accuracy for each sub band features for set B

Offline Classification of EEG data by using SAM

Six basic methodologies for control self-assessment have been defined:

- Internal Control Questionnaire (ICQ) self-audit and Customized questionnaires
- Control guides
- Interview techniques
- Control model workshops
- Interactive workshops

The National Institute of Standards and Technology control self-assessment methodology is based on customized questionnaires. It is a focused methodology suitable for assessing system based controls. It provides a cost-effective technique to determine the status of information security controls, identify any weaknesses and, where necessary, define an improvement plan. The methodology uses a questionnaire that contains specific control objectives and techniques against a system or group of systems can be tested and measured.
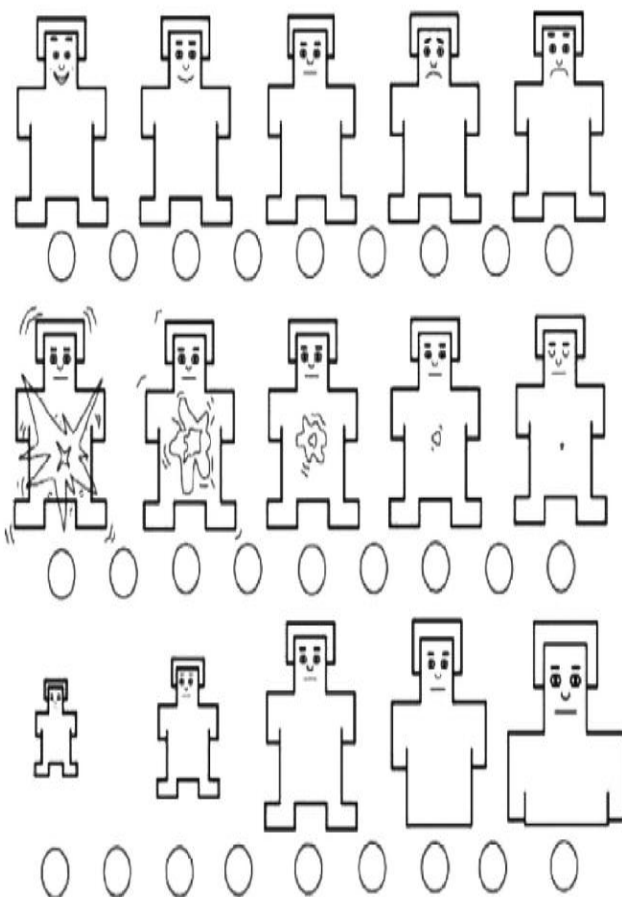

Figure 4.5 Shows SAM process of Experiment

During psychiatric illness and had normal or corrected-normal vision. All of the subjects were informed the scope and design of the study. In order to get quality data, the subjects were instructed to keep their eyes open and view each movie clip for its entire duration in the experiment. The process of our experiment is depicted in Fig.ure 4.5. Movie clips inducing different emotional states were presented in a random order. Each movie clip was presented for 3 – 5 min, preceded by 5 s of blank screen as the hint of start. At the end of each clip, the subjects were asked to assign valence, arousal and dominance ratings and rate the specific emotions they had experienced during movie viewing. The rating procedure lasted about 45 s. An inter-trial interval (15 s) of blank screen lapsed between movie presentations for emotion recovery. Valence, arousal and dominance ratings were obtained using the self-assessment manikin (SAM) shown in. The given self-reported emotional states were used to verify EEG-based emotion classification. In this paper, we only chose the sessions whose dominance ratings were equal to or larger than 3. The reason was that if the dominance rating was smaller than 3, we assumed that this session did not successfully arouse a certain emotion of the subject. Furthermore, we only cared about two kinds of emotional states, namely positive and negative emotions. If the valence rating was smaller than 5, then this session belonged to the class of negative emotion, else it belonged to the class of positive emotion. A 128-channel electrical signal imaging system (ESI-128, Neuro-Scan Labs), SCAN 4.2 software, and a modified 64-channel QuickCap with embedded Ag/AgCl electrodes were used to record EEG data. Each subject was fitted with a 100-channel electrode cap, whose electrodes were arranged according to the extended inter- national 10 – 20 system. The ground electrode was attached to the center of the forehead, and the reference electrode was located at the vertex of the head. The impedance was kept below 5 k Ω. The EEG data were recorded with 16-bit quantization level at the sampling rate of 1000 Hz. Electrooculogram (EOG) was also recorded, and later used to identify blink artifacts from the recorded EEG data. The EEG data we got from the experiments were analyzed through several procedures, including signal pre-processing, feature extraction, feature smoothing, feature dimensionality reduction, emotional state classification and trajectory of emotion changes, as shown in Fig. 4. First, the EEG signals were down-sampled to a sampling rate of 200 Hz to reduce the burden of computation. Second, the time waves of the EEG data were visually checked. The recordings seriously contaminated by Electromyogram (EMG) and Electrooculogram were removed manually. Third, each channel of the EEG data was divided into same-length epochs without overlapping using time-windows. Finally, all features discussed below were computed on each epoch of the EEG data.

CONCLUSION

The main purpose of this research study was to evaluate depression level prediction from urban and rural area of Lucknow city two different machine learning techniques

are used to classify EEG data obtained while a task is performed. To achieve this, two machine learning techniques selected based on literature study, their use in empirical studies and accuracy results reported by different authors. The selected techniques such as K-Nearest Neighbor (KNN), Regression Tree (RT) were evaluated.

For validation, the selected techniques were analyzed using EEG data collected from 50 subjects through an experiment in a controlled environment. The data obtained was processed to remove the artifacts and extract features. The data was formatted using a dataset in a form that is acceptable by the classification software to analyze the classification accuracies of the selected techniques.

From above study of dataset A and B it is also concluded that womens are found in more depressed as compared to men.

REFERENCES

[1.] Mallikarjun H.M, and Dr. H N Suresh, "Depression Level Prediction Using EEG Signals Processing", International Conference on Contemporary Computing and Informatics (IC31) pp. 928-933, 2014.

[2.]Ashish Panat and Anita Patil, "Analysis of emotion disorders based on EEG signals of Human Brain", International Journal of Computer Science Engineering and Application (IJCSEA), vol. 2, no. 4, pp. 20-24, 2012.

[3.] M.Teplan, "Fundamentals of EEG measurements," Measurement Science. Vol. No. 2, pp. 13-15, 2002.

[4.] Shih-Chung Chen, Shih-Chang Hsieh, and Chih-Kuo Liang, "An intelligent brain computer interface of visual evoked potential EEG", 8th International Conference on Intelligent Systems Design and Applications, pp. 343-347, 2008

[5.] Subha D.P., Paul K. Joseph, R. U. Acharya, L. C. Min, " EEG signal Analysis: A Survey," Journal of Medical Systems, Springer, vol. 34, no. 2, pp. 195-212, March 2010.

[6.] "Biosemi EEG ECG EMG BSPM NEURO amplifier electrodes", [Online]. Available: http://www.biosemi.com/products.

[7.]"EDF browser", [Online]. Available: http://www.teuniz.net/edfbrowser.
[8.]A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single trial EEG dynamics including independent component analysis", Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[9.] "Starlab - Living Science", [Online]. Available: http://starlab.es/products/biosemi.

[10.] "European Data Format (EDF)", [Online]. Available: http://www.edfplus.info.

[11.] "MathWorks - MATLAB and Simulink for Technical Computing", [Online]. Available: https://www.mathworks.com.

[12.] "EEGLAB - Open Source MATLAB Toolbox for Electrophysiological Research", [Online]. Available: http://sccn.ucsd.edu/eeglab.

[13.] C.W. Wu, H. Gu, H. Lu, E. A. Stein, J. H. Chen and Y.Yang, "Frequency specificity of functional connectivity in brain networks", Neuroimage , pp. 1047-1055, 2008.

[14.] M. Macas, M. Vavrecka, V. Gerla, and L. Lhotska, "Classification of the emotional states based on the EEG signal processing", in Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on, 2009, pp. 1–4, 2009.

[15.] Yang Li, Yingjie Li, Shanbao Tong, Yingying Tang, Yisheng Zhu, "More normal EEGs of depression patients during mental arithmetic than rest", Proceedings of NFSI & ICFBI, October 12-14 , pp. 165-168, 2007.

[16.] "Biosemi EEG ECG EMG BSPM NEURO amplifiers systems", [Online]. Available: http://www.biosemi.com/faq/file_format.htm.

[19.] M. Ungureanu, C. Bigan, R. Strungaru, and V. Lazarescu, "Independent component analysis applied in biomedical signal processing", Measurement Science Review, vol. 4, no. 2, 2004.

[20.] C., Shagass, R.A. Roemer, J.J. Straumanis, and R.C. Josiassen, "Psychiatric diagnostic discriminations with combinations of quantitative EEG variables", Br. J.Psychiatry, vol.144, no.6, pp. 581-592, 1984.