Affective State Recognition through Analysis of Electroencephalogram Signals by using Extreme Gradient Boosting

by

Md. Sakib khan
17101191
Nishal Salsabil
18301286
Rayeed Amir
17101428
Moumita Khandaker
17101065

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science And Engineering

Department of Computer Science and Engineering
Brac University
January 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sakib

Md. Sakib khan
17101191

Nishat

Nishat Salsabil
18301286

Rayeed

Rayeed Amir
17101428

Moumita

Moumita Khandaker
17101065

Approval

The thesis/project titled "Affective State Recognition through Analysis of Electroencephalogram Signals by using Extreme Gradient Boosting" submitted by

1. Md. Sakib khan (17101191)
2. Nishat Salsabil (18301286)
3. Rayeed Amir (17101428)
4. Moumita Khandaker (17101065)

Of Fall, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science And Engineering on 11th January, 2021.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam,PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam,PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Prof. Mahbub Majumdar
Chairperson
Dept. of Computer Science & Engineering
Brac University

Mahbub Majumdar,PhD
Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Emotion analysis has become a very important aspect in everyday life. It gives a detailed understanding of the behavior of human. In this research, we have focused on three dimensions of emotion. These are arousal (calm or excitement), valence (positive or negative feeling) and dominance (without control or empowered). We have collected dataset called DREAMER from a secondary source, consisting of Electroencephalogram signals from 23 participants, recorded on different 18 stimulus tests for each participant, and also pre-trial signals, along with self-evaluation ratings of all the dimensions for each stimuli at a scale between 0 and 5. In our proposed work, we have applied various feature extraction techniques which are FFT, DCT, poincare, power spectral density, Hjorth parameters which are activity, mobility, complexity, statistical features such as mean, median, maximum, variance, skewness. Additionally, chi-square and recursive feature elimination technique was used to select the discriminative features. Then, we have used, machine learning models such as support vector machine and extreme gradient boosting for classification. Finally, the 10 fold cross validation technique was performed to find the accuracy for each dimension separated in two classes (high or low). Here, extreme gradient boosting provided us better results with mean accuracy of 95.174% for arousal, 87.456% for valence and 84.541% for dominance, which is significantly higher than the state-of-the-arts.

Keywords: Affective State recognition, Emotion, EEG, Statistics, FFT, DCT, Poincare, Hjorth Parameters, SVM, XGBoost

Dedication

It is our sincere gratitude and warmest respect that all our well-wishers, particularly our parents who have always supported us and the teachers who have helped us in order to broaden our knowledge.

Acknowledgement

We would like to thank Almighty Allah at the very beginning for showing us the right direction, as well as for leading us to follow our engineering career.

Secondly, we would like to express our sincere gratitude to Prof. MD Golam Rabiul Alam, Bracu University, our thesis supervisor, for his motivation, support, and helpful guidance in enhancing our implementation ideas.

In addition, we take this sincere opportunity for expressing gratitude to all those who have helped us or at least intended to help us. Also we would like to thank our parents, family and friends for their kind motivation and continuous support towards our work.

Finally, we record our sense of gratitude to one and all of those who have been directly or indirectly involved in our work.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

BCI Brain computer interfacing

DCT Discrete cosine tranform

EEG Electroencephalogram

FFT Fast fourier tranform

SVM Support vector machine

XGBoost Extreme gradient boosing

Chapter 1

Introduction

1.1 Motivation

Emotion is fundamentally related to the way people interact with each other as well as with devices. This is why, the core of what makes us human is emotions. A human being can understand another human being's emotional state and act in the most efficient way to boost connectivity in a specific situation as it can be identified through verbal communication, facial expressions, and body language.On the other hand, it is opposite in terms of machines. But for quite some time now humans and machines have co existed, the presence of machines have grown so much that humans and machines are interdependent on each other. Without humans the machines will not be able to function and without machines our lifestyle diminish drastically. In fact some people depend on machines to simply stay alive. But it is also a matter of fact that, despite occupying such an important role machines have no understanding or features related to emotions. Therefore, it has become a mandatory step to do more research on affect recognition of humans, specially using brain signals, as everything generates from our brain and it decides how to react, when to and so on. This area of research is called Brain Computer Interfacing(BCI) and for the last many years, BCI term has become very popular to research on.

1.2 Problem Statement

In this technology-based world, people are heavily dependent on the technology and becoming involved in almost all the activities with the use of technology. Thus, it has become a necessity to improve the communication between devices and humans by automated computing. Up to now some BCI systems make the use of brain functions to exchange information and operate devices without requiring neuromuscular brain production pathways.[13]. Furthermore, human beings can comprehend other people's thoughts, but the machine cannot actually do this. The job we carry out is the same as faithfully as possible, but text, voice, face or expression does not actually generate actual feelings that come from brain signals. Additionally, EEG based approaches for emotional identification have had daunting problems surrounding the induction of emotional states and the extraction of traits for optimal classification. This paper has achieved the classification precision for single- and combined-channel situations of 62.3% (QDA) and 83.33% (SVM). [8] The visual and audio inputs are taken to implement neural network that lacks accuracy due to consistency deviation of human emotion over voice and expression, it will not bring high-end results, as emotion detection based on appearance only does not reflect complete human emotion reference. These techniques are not effective for reticent individuals and people with physical disabilities when they are unable to articulate their feelings or to interact fully.[6] There is another problem that has been mentioned in [27], that, mostly the research on EEG signals are conducted by using medical graded equipment's, which are very expensive. This becomes an issue, if we want to do brain signal analysis on daily basis. This will cause slowness in the research due to the expenses and not becoming available to all. Here we can observe from some of the problems that in the area of BCI, there are lot of scope of improvements by using better feature extraction, band extraction, scaling, classification models and by using low cost devices such as wearable devices [27], so that it can be used in day to day activities in near future.

1.3 Research Contributions

We have proposed this model to contribute in the area of brain computer interfacing, specifically, brain signal analysis. This research can help to assess the affective status of human in a given situation. In this sense, the emotions which are arousal, valence, and dominance are further categorized in two groups, high and low, in our proposed machine learning models. In our research we have extracted far more features than most previous papers and we have also implemented classification models to differentiate between the emotions.

1.4 Research Objective

Computers and machines have been able to outperform humans in almost every field when it comes to a specific task. But till today's date due to some limitations such as, machines can not interact with humans, as good as human does, it lack in understanding human emotions properly. So our goal is to train a model to read Electroencephalogram (EEG) signals and determine different dimensions of emotion of a human. This area of research is known as Brain Computer Interfacing (BCI). We have chosen this area of study due to the problems mentioned in the problem statement and still lack of applications of BCI in day-to-day activities. As we are mostly using our brain in most situations, the vital role of emotional state cannot be ignored anymore. In many situations, we may need to know the emotional state of a person, caused by any situation, especially introverted people as they are not comfortable enough to express their emotions, which results in taking wrong decisions in their life, sometimes life threatening. So it is very important that, the study of human emotions should be taken into consideration with higher priority. Therefore, In this field of study we want to contribute to improve analysis of collected brain signals, which will help the whole area of Brain computer interfacing. Eventually it will also benefit us, in our further future works in this area.

1.5 Thesis Outline

This report puts an emphasis on constructing a classification model that can identify emotions. Our aim is to formulate a system that will let the machine read and identify a person's emotion. The overall report focuses on the steps that were taken to achieve the goal.

Firstly, the introduction part(Chapter 1) States the motivation behind research which inspired us to address this particular problem statement. The goals of our research and summary of work is briefly discussed here.

In Related Work Section(Chapter 2), we have discussed about other research in the similar area of ours. The primary purpose of this chapter was to find out the shortcomings in the previous researches and improve upon them.

In Background Study(Chapter 3), we have discussed the pre-existing information that were already available so that the readers find it easier to understand the report. Here we explained the principles of the brain waves and the machine learning models that we used.

Then in Proposed Model(Chapter 4), we stated our proposed model,overall process and further explained the algorithms we used. This portion also included the description of our data set.

We have published the result in results and discussion chapter(Chapter 5), and also discussed its accuracy. The results are presented using multiple graphs for better visual representation.

Lastly we concluded with our research overview, the limitations of our research and the potential future works that can be done on it.

Chapter 2

Related Works

There are many suggested approaches for affect recognition from Brain signals.Various methods have been used to influence the acquisition of data, originating from facial expressions, peripheral physiological signals etc. The use of pattern recognition methods for impact recognition involves the collection of data on a subject's affective state of brain during the time in which a particular emotion is expressed.Emotion has a significant part in interpersonal interaction and correspondence between individuals. There are a number of emotional models used in affective computing since various hypotheses of how emotions occur and have been articulated.Recently the rating of EEG data emotions has become very attractive in terms of exponential development in dry electrode technology, machine learning algorithms, deep learning and numerous applications in the real world. In this chapter, we will discuss some of the previous works studied.

In the research[16], they have applied deep learning to building EEG data recognition. They used some classifiers which were K-Nearest neighbor , Support-vector machine, Graph regularized extreme learning machine , DBN and DBN-HMM and compared the performance of the deep models to those.Thus, the end-of-paper experimental results indicate that high-frequency (beta and gamma) characteristics were associated with emotional recognition more closely and DBN and DBN-HMM obtained the most reliable results. The authors Ayush Bhardwaj, Ankit Gupta, Pallav Jaiv and Asha Rani indicated [19] that they had the dataset from the International Affective Picture System . They feature from EEG signals by using Power Spectral Density and Energy. But before feature extraction authors pre-processed the data by doing segmentation, Bandpass filter and Independent Component Analysis . For training and testing the dataset they used Machine learning techniques such as Support vector machine and Linear Discriminant Analysis to classify EEG signals into seven different emotions.After studying the paper [6], they have proposed an approach to classifying human emotions using EEG signals and audio-visual stimulation dependent protocols were designed to obtain EEG signals using 63 sensors.In a neural network the author studied the EEG signals by means of a discrete transformation and grouped them into five frequency sub bands using the 'db4' wavelet function, while two statistical characteristics were derived from the alpha band. They finished with this statement that the audio visual feedback performed better than visual stimulation. For emotional prediction, the studies were compared between auditory and visual elicitors [17] For six selected emotions, tData obtained from 24 participants, including facial electromyography, electroencephalography, skin conductivity and respiratory data, were from various physiological signals. In this paper they employed two methods of data mining. These are, decision rules and K-nearest neighbor based on technique of the rough collection. At the end they obtained an average F1 accuracy of visual scenes and 76.1% accuracy of auditory scenes. They came up with the finding that the auditory stimuli could be best interpreted as emotional elicitors that could affect functional HCI.On the other hand the authors of [8], a functionality collection obtained from the higher order crossing study was used to implement a different emotional identity method based on EEG. They obtained the data collection from 16 participating active volunteers (all subjects were right handed). Six simple EEG signals were defined and four various Classifiers used: QDA, K-NN, MD, SVM. The authors used two different approaches to classify the accuracies. One was a single channel case (received a mean classification rate.From the paper [23] it suggested a novel approach to understanding emotions from electroencephalogram (EEG) signals from users. Eight stable participants were presented with the data collection, the EEG signals were collected by seven EEG amplifier channels. Multi-fractal Detrended Fluctuation Analysis (MFDFA) approach extracted the features from each channel of the collected signals for the entire frequency bands. They implemented some efficient Classifiers called Support Vector Machine, Linear Discriminant Analysis, Quadratic Discriminant Analysis and K Nearest Neighbor.Empirical mode of Decomposition [29] for emotional identification using signals from the electroencephalogram (EEG). The data set of the channelled BIOPAC laboratory device was obtained and EEG signals were gained from a capacity of 13 women and 13 men for 12 nice and 12 frustrating images. The data are now available. Here, EMD has been used to test nonlinear and also the non-stationary time series that have interrupted the signal into IMFs. They used different Classifiers - SVM, LDA, Naive Bayes to classify pleasant and unpleasant emotions and compared their performances. In short, they found that SVM provides a better solution by using EMD methods from EEG signals.A simplistic and very quick process **c8** of data ingestion, feature extraction and feature space formation for epileptic seizure detection is described in this paper The scalp EEG dataset obtained at Boston Children's Hospital from 22 pediatric patients with 192 intractable seizures is used to test this simple approach with a good result. From the obtaining result from the proposed method they indicate that the 10-minute training subsets are as effective as longer training subsets. They also use the ELM classification algorithm works better than the SVM classification technique when classifying multi-channel data at once.So in short they got a very good accuracy which was 99.48% from their report. The emotion detection of EEG brain signals using the Support Vector Machine (SVM) is discussed [21] that they have used emotional stimuli from the International Affective Picture System which is database. They recognized five kinds of feelings, such as happy, relaxed, optimistic, sad and frightened. Implemented a feature extraction process. Via SVM, the combined feature set of all subjects was processed. The findings showed that 70 percent correct in emotion detection in the arousal-valence domain among more than 30 samples. In this research [32], it mentioned about the depression and public health care system. They collected some samples which consisted of 62 depressive individuals and 65 healthy controls. They collected EEG and ECG from each individual. They have used linear and nonlinear methods to derive HRV parameters.

The results were determined on the basis of age, gender. After that the MDD patients had an in-house clinical algorithm for antidepressant treatment. The EEG and ECG reports were again assessed. This provides further understanding of the relation between HRV and depression parameters. However, they faced some problems due to shortages of time and clinical relevance. In addition, small samples are available to show important associations between HRV parameters and measures that index the severity of the symptoms of depression. They needed to do more research on this in order to obtain specific results.In another paper the authors [3] proposed a algorithm of selection for two-stage function.In order to enhance the discriminatory data, the more informative channels are first chosen by them. They used two methods for choosing the channels. One is bidirectional search and another one is plus L minus R techniques. The genetic algorithm (GA) is used after channel selection to pick the best features from the collected signals of channels. Here, the redundant and less discriminating features are also removed, in addition to removing the less comprehensive networks. This productive approach involves the implementation of automobile regressive model parameters, band strength, fractal dimension and wavelet capacity. In order to evaluate the performance of the final subset of functions, classifications such as linear discriminant analyzes and the support vector machine are used to classify the reduced feature set of the two classes. In conclusion, the findings are compared and contrasted with two popular techniques, namely the main component selection (PCA) and the single-stage function selection.The results suggest enhanced classification accuracy in relatively low computational time by choosing the two-stage feature. Brain signal recognition of emotions has become a research subject. The author of [35] applied different machine learning algorithms to describe the emotion as a psychological measure.This paper suggests an approach to describe with better precision the various kinds of estimation in humans. In this analysis, it extracts many features from EEG brain signals and uses the correlation matrix, the data benefit assessment and the recursive process of eliminating features to refine the collection of features. Then the Extreme Gradient Boosting algorithm was used which was adopted for the classification technique. The DEAP data set was used as a checked method, and numerous classification algorithms including Naive Bayes, KNN, C4.5, Decision Tree, Random Forst algorithms were in relation to the proposed solution. The DEAP algorithm is based on the structure of gradient boosting and follows linear model solver and tree learning. In another paper [34] XGBoost has been used in predicting liver disease to gain the well fitted accuracy expected on the particular field. They claimed of getting 99% accuracy on some hyper tuning parameters that was selected through classification boosting technique. The method assembles some pursuant procedure of finding out the accurate prediction using 2nd order derivative where the methodology consists set of component models. These sets were gathered from a bunch of individual weak learners and the process carries forward using iterations. This boosting technique enhances its ability by the use of L1 and L2 as the regularization technique. In research [31], indicates the prediction of outcomes of hypertension in patients who might show subtle reactions but passes through a post traumatic phase using XGBoost,C4.5 decision tree, Random Forest and SVM. Physical condition is evaluated in predicting the after effects. After applying the feature elimination techniques like recursive feature elimination comparatively XGBoost gives the best accuracy among the four methods of about 94.36%. The precision needs to be as neat as possible to avoid any harmful psychological condition on patient and emergency precautions needs to be taken on alarming situation. In paper [37], this research examines the cortical activity recorded through scalp in cardiac patients. They use EEG patterns in patients given anaesthesia in measuring the state caused by the big surgery. It was observed in the tested set of 19 undergone surgery patients that hjorth complexity increases as the artificial ventilation is placed on the support of respiratory activity and on the opposite terms hjorth activity significantly decreases after the surgery on a normal condition. EEG signal deviation is noticed vastly due to the great effect of ventilation, surgery that helps differing the measurements of depth of anaesthesia or its affecting components on body. Pain endurance during a surgery is a difficult task to calculate and this result contributes towards the mechanism of building a machine assessing the spread of pain.

In all the above works, advanced and equipment which are costly or non portable was used for capturing the EEG signals to measure the affects of brain from physiological signals.They have increased capability and signal efficiency, but are costly, non-portable and unsustainable and hence not suitable for casual everyday applications; we may recommend methods for detection of impact with signals obtained by cost-effective and easy-to-use devices, which enhance the interest of parties who try to introduce computer effects on a wide range of applications.

Chapter 3

Background Study

3.1 Brainwaves

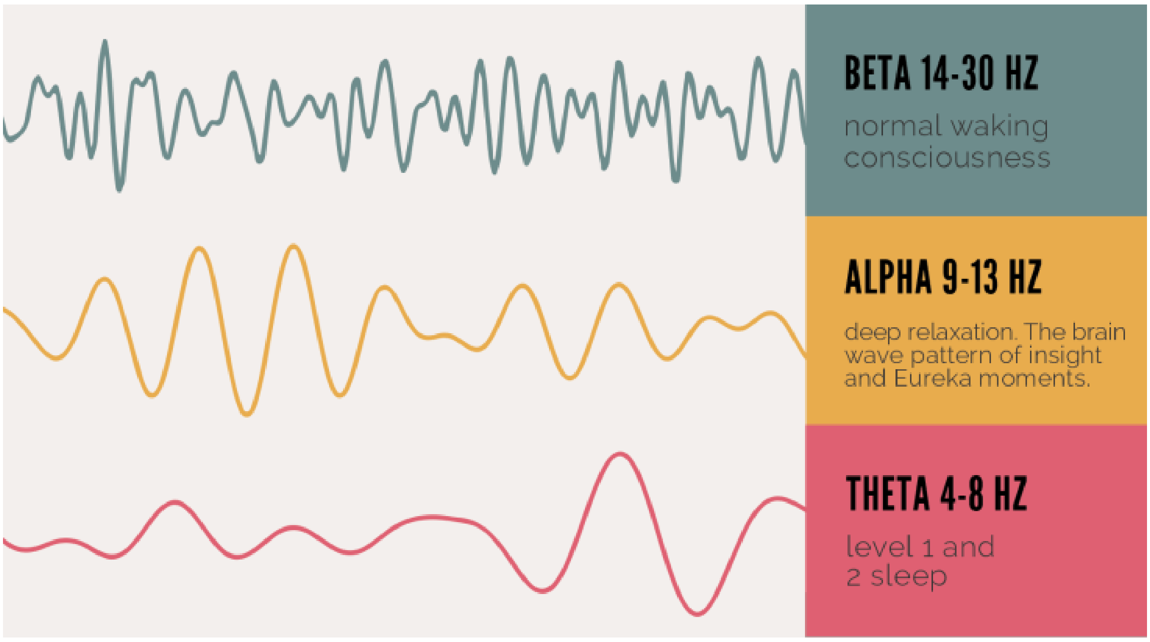


Figure 3.1: Brainwaves

3.1.1 Theta

Theta frequency band pass filter is considered a band pass filter with a frequency range between 4Hz and 8Hz. The brainwaves of theta occur most frequently in sleep, but in deep meditation, they are also dominant. Theta is our path of understanding, remembrance and intuition. Our senses are separated from the external environment in theta and concentrated on signals coming from within. This is the twilight state that we generally feel only fleetingly as we wake up or drift off to sleep.

3.1.2 Alpha

A band pass filter with a frequency range between 8Hz and 13Hz is considered as Alpha frequency bandpass filter of the signal. During softly flowing thoughts, and in certain meditative states, alpha brain waves are dominant. Alpha is the strength of the present. For the brain, Alpha is the resting state. This range identifies a commonly stated rhythm that is most prominent during closed eyes. It provides normal reading when a person is wake. Alpha waves assists to coordinate emotionally, confidently, alertly, incorporate mind or body and instruct.

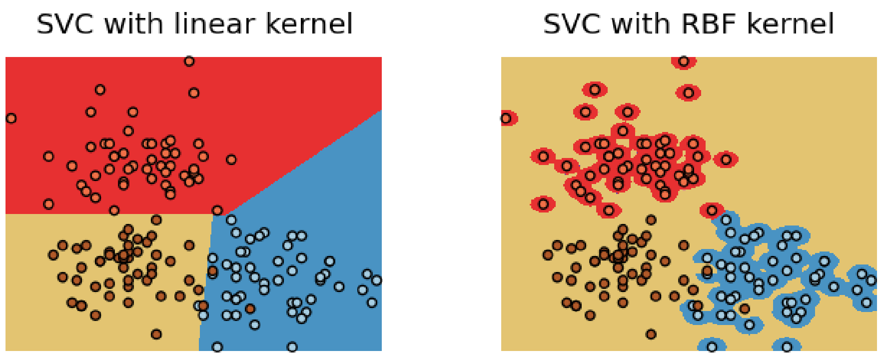
3.1.3 Beta

A band pass filter with a frequency range between 13Hz and 30Hz is usually considered as band pass filter in Beta frequency. But from this we will take between 13-20 Hz When our focus is geared towards cognitive activities and the outside world, beta brainwaves overpower our usual waking state of consciousness. Beta is a 'quick' mechanism when we are diligent and attentive and are engaged in problem solving, assessing, taking decisions and focusing mental tasks

3.2 Algorithms for signal analysis

3.2.1 Support Vector Machine

The support vector machine is a supervised classification and regression analysis model. As well as the linear classification method, SVM effectively performs a non-linear classification, which is performed by its kernal feature. The effectiveness of SVM depends on the selection of kernel parameters. In order, neurology conditions such as epilepsy and sleep disturbances, the support vector method was commonly used to detect Electroencephalogram (EEG) signals. Due to its convex optimization problem, SVM demonstrates good generalization performance for high dimensional data.
There are various types of kernel functions such as linear, nonlinear, polynomial, Gaussian kernel, radial base function (RBF), sigmoid, etc. From these kernels, we initially concentrated on the linear and RBF kernel in our approach.And these are-



Linear kernel: The data will be linearly separated here, it means, it can be divided by a single line, the Linear Kernel is used. It is one of the most prevalent kernels that can be used. It is often used when a particular dataset has large amount of features. It is also easier to train the SVM with a linear kernel than with any other kernel. For the linear kernel, a projection function for the new input is determined according to the dot product between the input (t) and each support vector (t_i).

$$f(t) = b_0 + \sum(a_i * (t, t_i)) \tag{3.1}$$

It consists of computing the internal products of a new vector of input (t) with all supporting vectors in training results. The b_0 and x_i Coefficients for each input should be determined by the learning algorithm from the training results.

Gaussian radial basis function: It is the most used kernel from various kernels of SVM because along the entire x-axis, it has a localized and finite response. It is used when there is not enough prior information regarding the data.

The RBF kernel is defined as:

$$k(m, q) = exp(\frac{\text{mod } (m - q)^2}{2\sigma^2}) \tag{3.2}$$

Also can be written,

$$k(m, q) = exp(-d(x, z)^2\gamma) \tag{3.3}$$

where,

$$\gamma = \frac{1}{2\sigma^2}$$

Here, $\sigma=$ the kernel width parameter
γ = large σ would essentially mean small γ and would imply a smoother fit.

In our analysis, we observe that the RBF kernel as groups the data are best used in a more suitable way in order to better determine effects of the signals.

3.2.2 Extreme Gradient Boosting

Boosting is a process used to solve complex and complicated problems which can be mostly problems of the real world. Here, to improve the used model's precision, boosting strategies combine the poor learners with an efficient learner. Base learners are sequentially generated in Gradient boosting so that the present base learner is more effective than the previous one. It attempts to optimize the previous learner or model's loss function. The advanced variant of gradient boosting is XGBoost. This method is designed to concentrate on speed and computation performance of the model. By building decision trees in parallel, XGBoost supports parallelization and uses cache optimization to optimally utilize resources. It incorporates methods of distributed computing to evaluate large and dynamic concerns.
The principles of the calculation are documented as follows:

- Generate a standard set that is emphatically connected, discretizing component vectors if the element is constant in the set arrangement.
- Choose the principles that result in the forecast label, at which point the new norm set is structured. Determine the lift of each norm and with lift¡1.0.0 evacuate the principles We can get the convincing guidelines set in this way.
- In the effective norm set by most brief duration, greatest raise, and greatest support, sort the concepts.
- In the list of capabilities, print the part and quit if cycle records are more prominent than the number given. Else, in the effective standard package, we get the primary guideline and add its consequence to the list of capabilities, expelling the standard from the set of viable principles. We do not need to include the component that exists on the capability list as of now.
- In the rest of the preparation set, delete examples that meet the standard, ascertain lift and backing.
- Sort the criteria in the viable theory laid down by most extreme lifts, go to step 4 for the greatest support.

XGBoost is a Machine Learning calculation focused on decision tree gathering that utilizes a method of inclination boosting. Artificial neural systems can usually beat all other calculations or structures with forecast issues, including unstructured data (images, content, etc.). Be that as it may, in terms of little to medium organized/forbidden data, calculations based on decision trees are seen as best.

$$f(m) \approx f(k) + f'(k)(m-a) + \frac{1}{2}f^{n}(k)(m-k)^2 \tag{3.4}$$

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n}\left[l\left(q_i, q^{(t-1)}\right) + r_i f_t\left(m_i\right) + \frac{1}{2}s_i f_t^2\left(m_i\right)\right] + \Omega\left(f_t\right) + C \tag{3.5}$$

Here C = Constant

where, the r_i and s_i are defined as,

$$r_i = \partial \hat{z}_i^{(b-1)} . \int \left(z_{i,} \hat{z}_i^{(b-1)} \right)$$
$$s_i = \partial \hat{z}_i^{(b-1)} . \int \left(z_i, \hat{z}_i^{(b-1)} \right)$$

After removing all the constants, the specific objective at step b becomes,

$$\sum_{i=1}^{n} \left[\text{ rift } \ (m_i) + \frac{1}{2} \operatorname{sif}_t^2 (m_i) \right] + \Omega \left(f_t \right)$$

Where,

$$\Omega(f) = \gamma P + \frac{1}{2} \lambda \sum_{j=1}^{P} Z_j^2$$

Here,
m_i = input
z_i = stands for real value known from the training dataset
$\Omega(f)$ = complexity of the tree
P = the number of leaves

3.3 Cross Validation Technique (K-FOLD)

This method is an assessment of machine learning models in a limited data sample. This method has single parameter which corresponds to the number of groups in which the data will be separated. This is why this process is named as cross validation with K amount of folds. Here, if a particular number is chosen for K variable, the data will be divided into K amount of parts. For instance, for K = 10, it will be cross validation with 10 number of folds where the data will be dseperated into 10 parts for further analysis.

Steps:

- Randomly alter the data set.
- Divide the data into K amount of groups.
- Take 1 part of the division as testing and rest for training.
- Store the score after performing testing.
- Repeat the process K times.

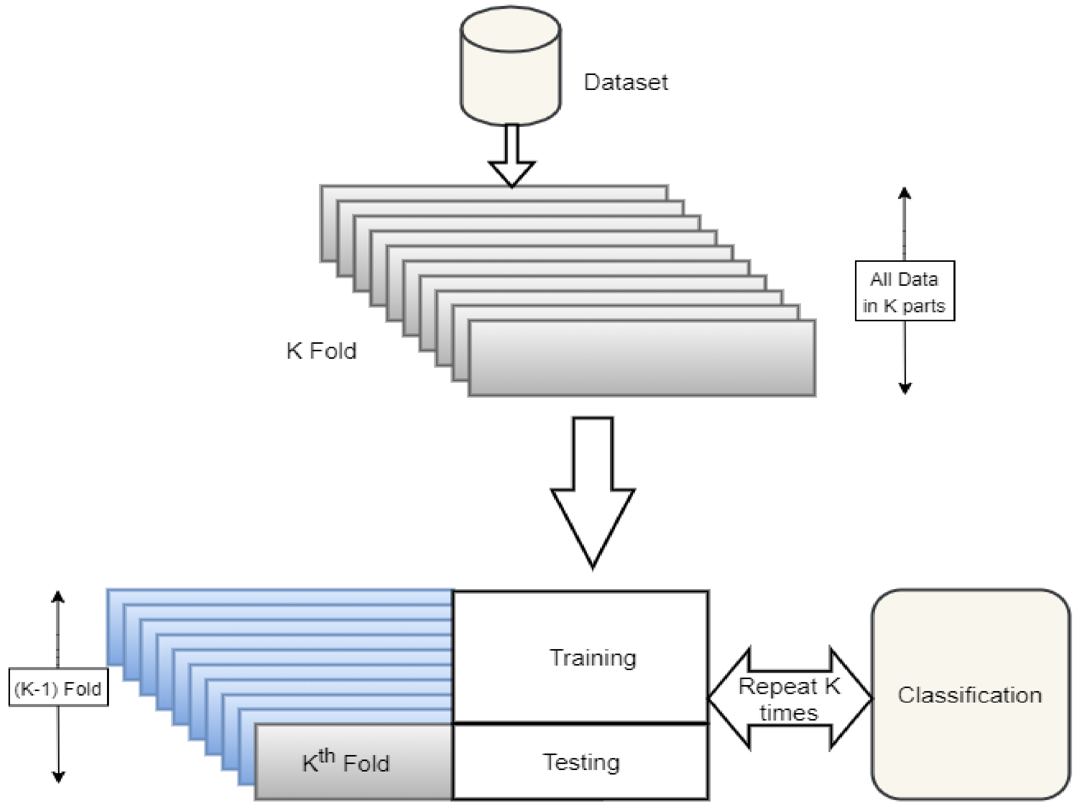


Figure 3.2: Steps of Cross Validation (K-fold)

In order to learn the prediction indicators and validate the function of prediction on the same data is a computational mistake. As here, it is often seen that, by testing the model on the data which is similar to the data of training would have a perfect outcome but will not provide any meaningful results for the data which is still unknown to the model. This is regarded as over-fitting. So to avoid this issue, this cross validation technique with K number of separation for further analysis is a proper technique to use.

Chapter 4

Model Proposal

4.1 Workflow of the proposed model

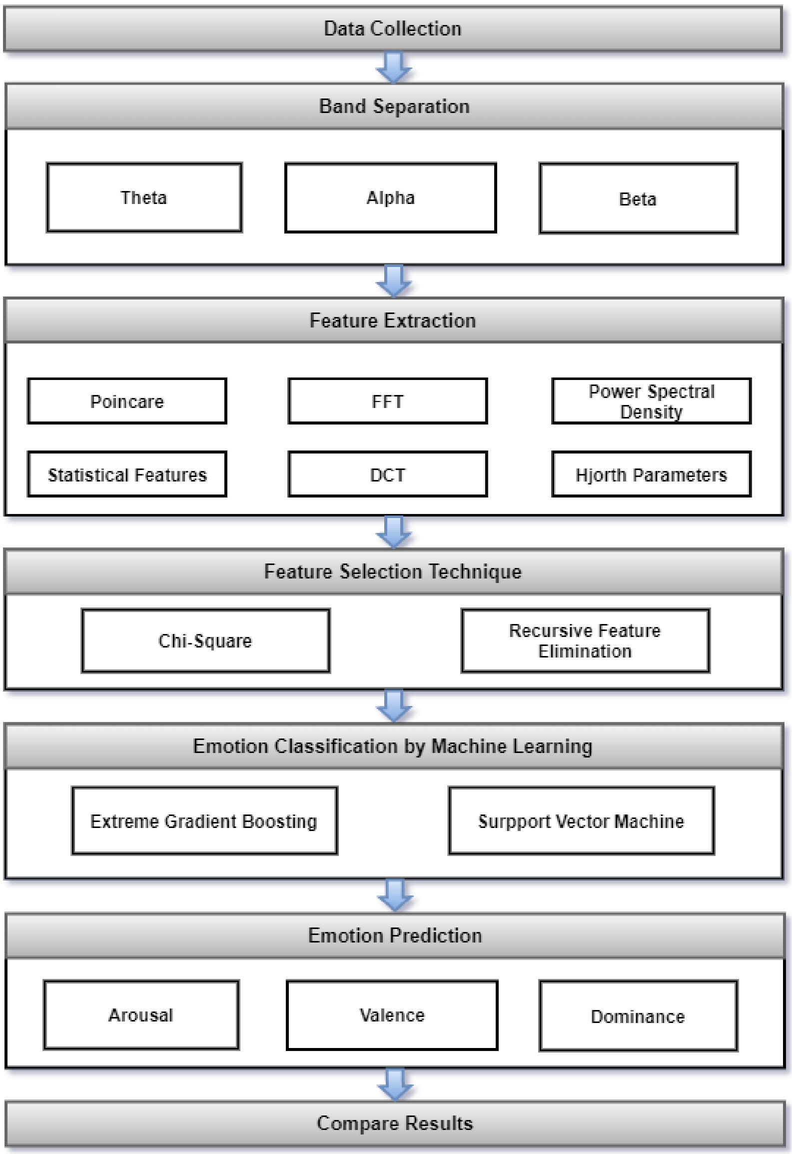


Figure 4.1: Workflow diagram

4.2 Collection of Data and Description

4.2.1 Collected Data

To work with the proposed model, we have collected a data-set called DREAMER from a secondary source, which is proposed in the research [27]. Audio and video stimuli were useful to develop emotional responses to participants of the study in this data set. This data set consists of 18 stimuli tested on participants and Gabert-Quillen et al. [20] have selected and analyzed them to induce emotional sensation. The clips come from many films showing a wide variety of feelings. Two of each film centered on one emotion: amuse, excite, happy, calm, angry, disgusting, fearness, sad and surprising. All the emotional descriptions of each film are shown here and also in the paper [27].

Stimuli	Film Clip	Targeted Emotion
1	Searching for Bobby Fischer	calm
2	D-O-A	surprising
3	The Hangover	amusement
4	The Ring	fear
5	300	excitement
6	National Lampoon's Van Wilder	disgust
7	Wall-E	happiness
8	Crash	anger
9	My Girl	sadness
10	The Fly	disgust
11	Pride and Prejudice	calmness
12	Modern Times	amusement
13	Remember the Titans	happiness
14	Gentleman's Agreement	anger
15	Psycho	fear
16	The Bourne Identity	excitement
17	The Shawshank Redemption	sadness
18	The Departed	surprise

Table 4.1: Stimuli films and their target emotions

The length of all these clips are in between 65 to 393 seconds, which is enough to to use for emotions [24], [10]. However, only the last 60s was considered for next steps of analysis. This experiment was done in a separated dark place and the videos were presented to them with a fourty five inch television monitor and an attached speaker for hearing the audio and test on them. the Electroencephalogram signals were recorded using the device called EMOTIV EPOC which is a wireless device and used as headset, also shown in figure 4.2, which is consists of 14 channels. Using 14 channels the data was collected from sixteen different location. There was also data recorded from the wireless SHIMMER ECG sensor, but in this research we have only focused on the EEG signals.

Figure 4.2: The Emotiv EPOC wireless EEG headset

4.2.2 Data Description

Attribute	***Amount/details***
Number of participants	25(23)
Stimuli Amount	18
Video contents	Audio, Video
Duration of contents	65-393s
Age of participants	22-33
Signals	EEG
Channels	14
Sampling Rate	128hz
Rating of	Valence, Arousal and Dominance
Range of rating	0-5
Total Data	(23*18) = 414

Table 4.2: Data Dimension

Initially the data collection was performed among 25 participants, but due to some technical problems, data collection from 2 of them was incomplete. As a result the data from 23 subjects were considered for further analysis. The data set consists of signals from trail and also pre-trail both and the pre-trail data is collected as baseline for each stimuli test.
This study was also approved by the Ethics Committee of the University of the West of Scotland University, situated in Scotland. [27]

4.3 Signal Processing

The EEG signals typically holds a lot of noises. From that, the vast majority of ocular artifacts are dominating below 4 Hz, muscle motions create over 30 Hz artifacts and power line noise is typically 50-60 Hz Noise. [27]. These noises must be removed before further steps to get improved results. Additionally, in order to work on specific area, we must focus on the range of the frequency, which provides us the signals caused by the stimuli. The frequency band, which contains the information, related to affect recognition task, ranges between 4 to 30 hz band [27]. To get rid of the noise from the signals and to find the band of interest, we have used Band Pass Filtering to get the sample values ranging from 4 – 30 Hz initially.

4.3.1 Band Pass Filter

Band-pass filter is a technique or process that accepts the frequencies within a range of selected frequency bands and discards all the frequencies beyond the frequency of interest. Band pass filter is a technique which is also considered as a combination of low pass filter and also a high pass filter to avoid the frequency we do not need. The main purpose of such a filter is to restrict the bandwidth of the signal, to get the signal from the frequency range we need and to reduce unpleasant noise by blocking the frequency bands that we will not use anyway.

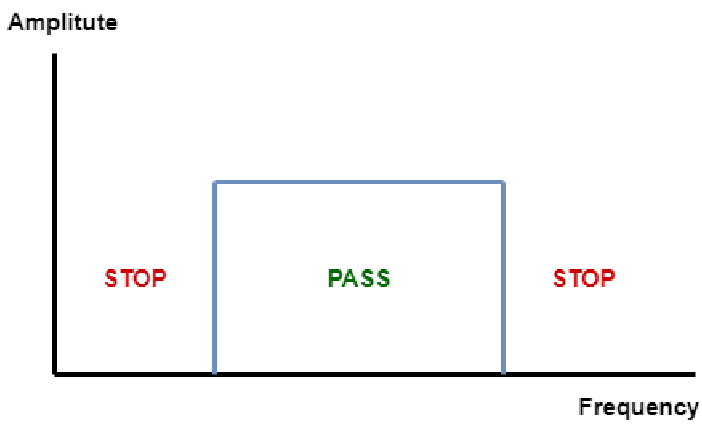


Figure 4.3: Band pass filter

Using this filter technique, we have first filtered the frequency range between 4hz to 30Hz, which contains relevant information which we need. This automatically removes the unpleasant noises that we wanted to remove. As now we have the signals of interest, we have decided to divide them in more 3 bands, which are Theta, Alpha, Beta. We have choosen these bands as these are the most commonly chosen bands for analysis of EEG signals. The definition of the boundaries between bands is rather subjective, but the range that we will use in our case is, Theta which will be ranging between 4hz–8Hz, Alpha ranging between 8hz to 13Hz and Beta ranging between 13hz to 20Hz for our further analysis.

4.4 Feature Extraction

4.4.1 Fast Fourier Transform

Fast Fourier Transform is among the most useful methods for processing various signals which is also performed in many research [12] ,[9], [14], [36], [15]. Here the algorithm of FFT calculates a sequence of Discrete Fourier Transform. The FFT stems are evaluated because they operate in the Frequency Domain, in the time or space, equally computer-feasible. The O(NlogN) result can also be determined by the FFT. Where N is the length of the vector. It functions by splitting a N time domain signal into a N time domain with one single stage. The second stage is an estimate of the N frequency range for the N time domain signals. Lastly, the N spectrum has been synthesized into one frequency continuum to speed up the Convolutional Neural Network training phase.
The equations of FFT are shown below:

$$H(p) = \sum_{t=0}^{N-1} r(t) W_N^{pn} \tag{4.1}$$

$$r(t) = \frac{1}{N} \sum_{p=0}^{N-1} H(p) W_N^{-pn} \tag{4.2}$$

Here H_p = represents the fourier co efficients of r(t)

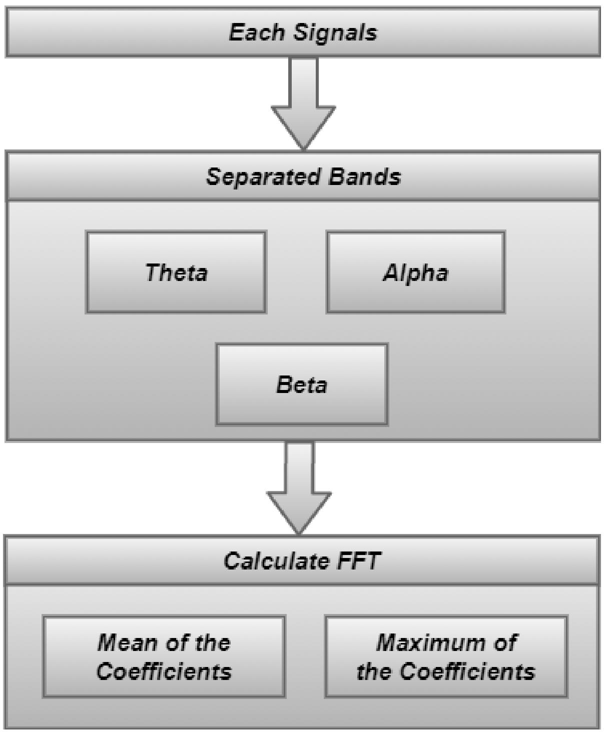


Figure 4.4: FFT feature extraction steps

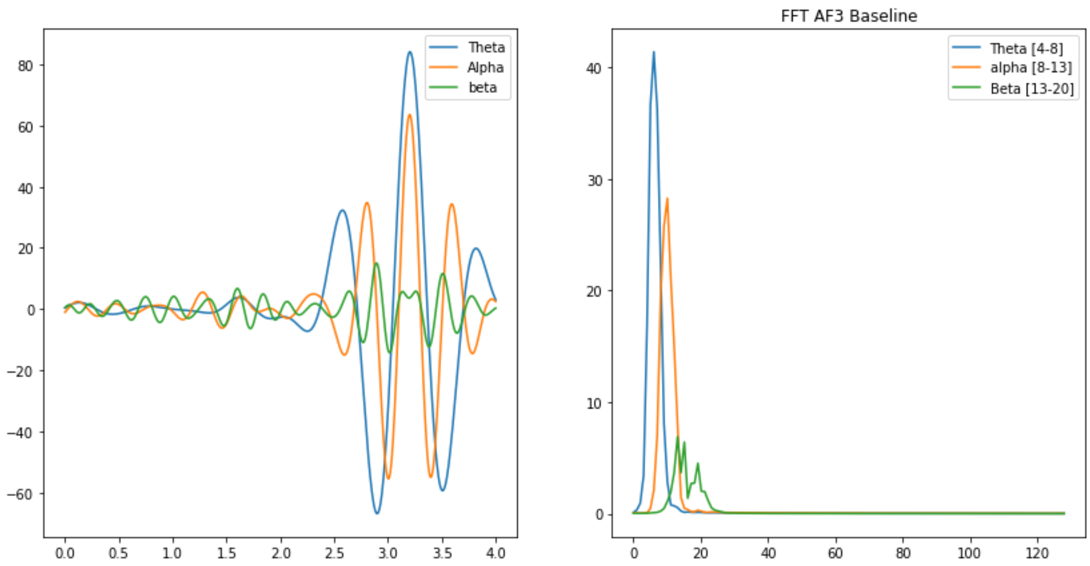


Figure 4.5: A Baseline signal in frequency domain

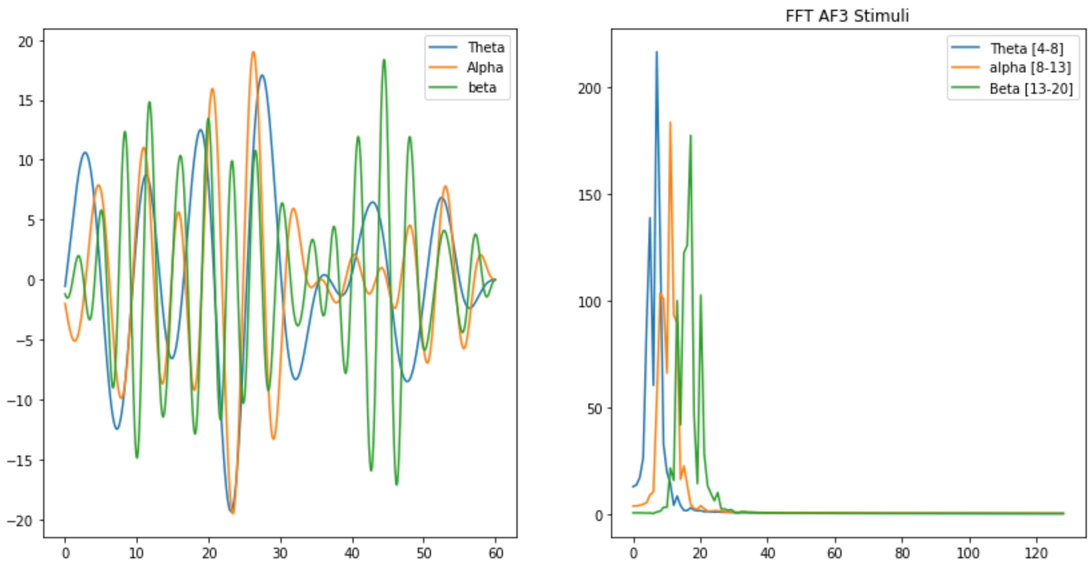


Figure 4.6: A Stimuli signal in frequency domain

We have implemented this FFT(Fast Fourier Transform) to get the Coefficients. From that, we extracted the mean, maximum features resulting total of, [(max, mean)*(3 bands) = 6 Features for each channel. So for 14 channels we got a total [14 channel *6 Features] = 84 features.

4.4.2 Discrete Cosine Transformation

This method exhibits a finite set of data points for all cosine functions at varying frequencies which is used in research [2],[4], [11], [18], [1]. The DCT is usually applied to the coefficients of a periodically and symmetrically extended sequence in the Fourier Series.In signal processing, DCT is the most commonly used transformation method (TDAC).The imaginary part of the signal is zero in the time domain and in the frequency domain the actual part of the spectrum is symmetrical, the imaginary part unusual. With the following equation, we can transform normal frequencies to the mel frequency:

$$X_P = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right) P\right] \tag{4.3}$$

N = the list of real numbers
X_p = a set of N data values

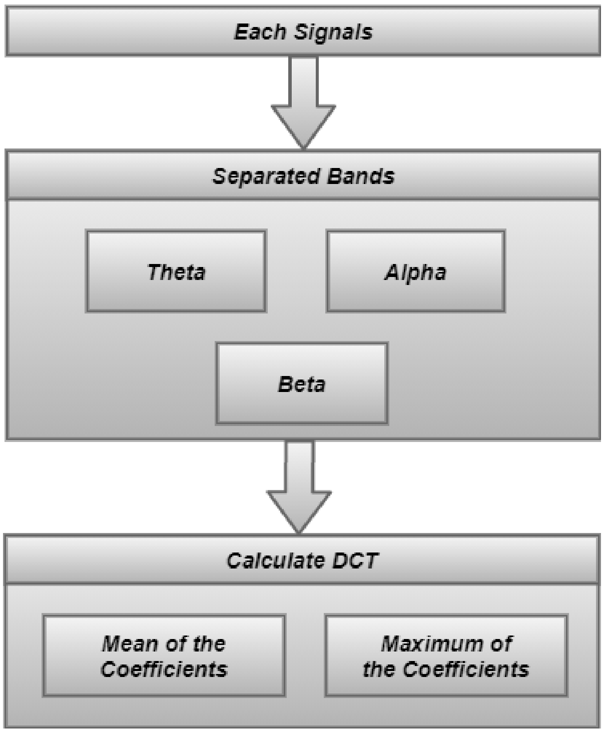


Figure 4.7: DCT feature extraction steps

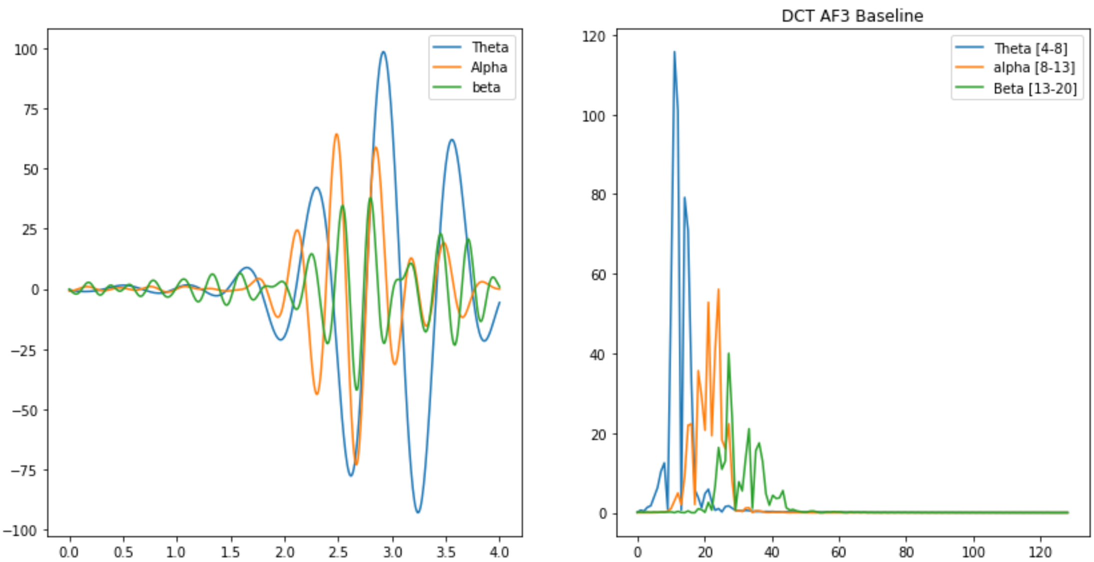


Figure 4.8: A Baseline signal in frequency domain

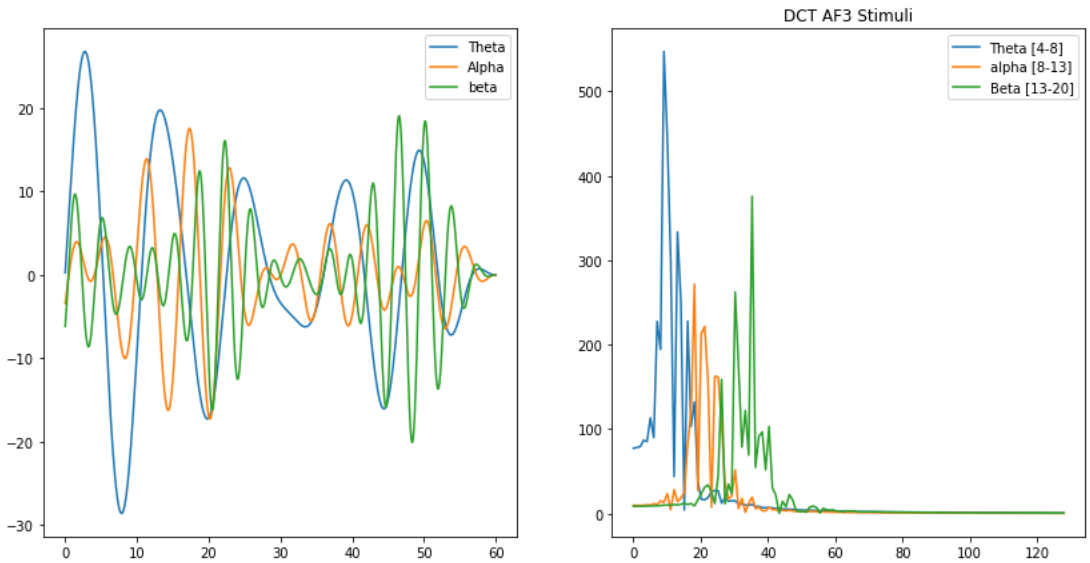


Figure 4.9: A Stimuli signal in frequency domain

We have implemented DCT(Discrete Cosine Transformation) to get the Coefficients. From that, we extracted the mean and maximum as features.In this paper, to obtain a sparse representation in the frequency domain, we suggest the use of the discrete cosine transform (DCT) to pre-process a signal.Resulting, total of, [(max, mean)*(3 bands) = 6] features for each channel. So for 14 channels we got a total of, [14 channel * Features] = 84 features.

4.4.3 Hjorth Parameters

The Hjorth parameter is one of the ways in which a signal's statistical property is indicated in the time domain and has three parameters which are Activity, Mobility, and Complexity. These parameters were calculated in many research [26],[22],[30],[28].

Activity: The parameter of operation describes the power of signal, the variance of a time function. This can suggest the power spectrum surface within the frequency domain.

Mobility: The mobility parameters represent the average frequency or the share of the natural variation of the spectrum. This is defined as the square root of the variance of the first y(t) signal derivative which is divided by the y(t).

Complexity: The parameter Complexity reflects frequency shift. The parameter contrasts the signal resemblance with a pure sinusoidal wave, where the value converges to 1 if the signal is more identical. .

Formula for the Hjorth parameters:

Parameter	Notation
Activity	$\mathrm{var}(y(t))$
Mobility	$\sqrt{\dfrac{\mathrm{var}(y'(t))}{\mathrm{var}(y(t))}}$
Complexity	$\dfrac{\mathrm{mobility}(y'(t))}{\mathrm{mobility}(y(t))}$

Equation: Hjorth parameters

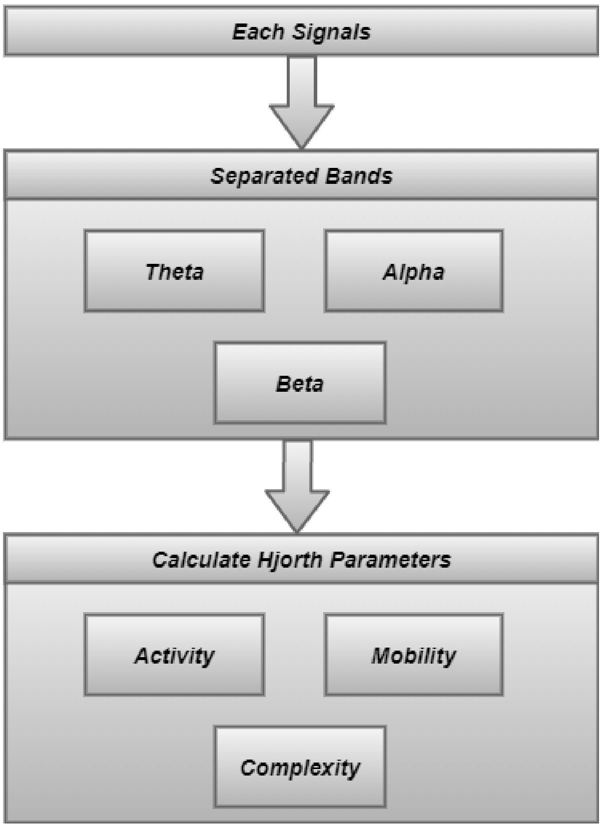


Figure 4.10: Hjorth parameters feature extraction steps

We have calculated all the parameters of Hjorth which are Activity, Mobility, and Complexity as features for our analysise. Resulting,
Total of, [(Activity, Mobility, Complexity) $*$(3 bands)] = 9 Features for each channel. So for 14 channels we got a total,[14 channel $*$9 Features] = 126 features.

4.4.4 Statistical Features

Statistics are the application of applied or scientific data processing using mathematics. We use statistical features to work on information based data and focusing on the mathematical results of these information. We can learn and gain more and more detailed information on how statistics arrange our data in particular and how other data science methods can be optimally used to achieve more accurate and structural solutions.

There are many research [25],[33],[7] on emotion analysis where statistical features were used. The statistical features that we have extracted are as given below:

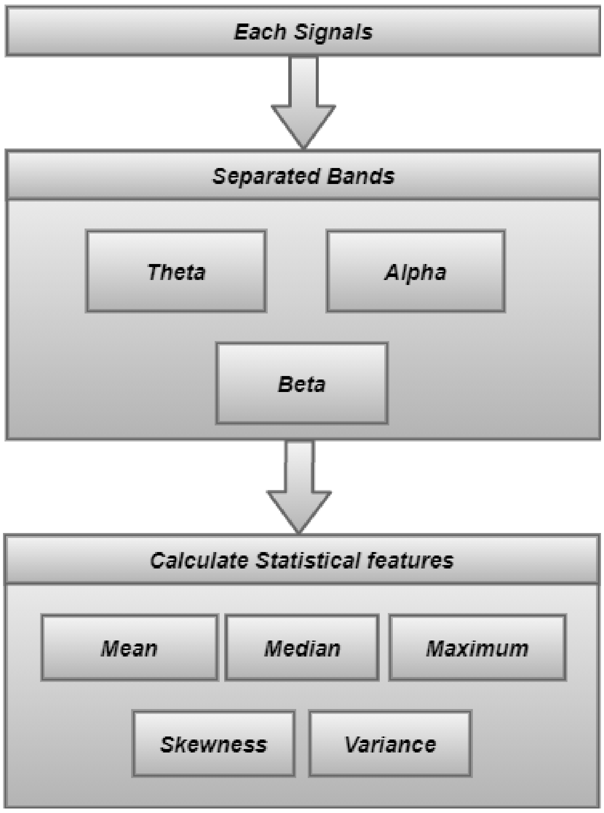


Figure 4.11: Statistical feature extraction steps

Median

Median is the midpoint of a series of values arranged in ascending order.
When the number of values are odd,

$$Median = (\frac{N}{2})^{th} + 1 \tag{4.4}$$

When the number of values are even,

$$Median = \frac{(\frac{N}{2})^{th} + (\frac{N}{2} + 1)^{th}}{2} \tag{4.5}$$

Here our signals are divided into frequency ranges of theta, alpha, beta that then again calculate the median value of each range as a selected feature of state assessment.

Mean

Mean of a signal is calculated to gain the feature of average frequency each participant is providing based on a particular emotion and situation. This feature helps engrave the base point of one range of frequency to train the system for appropriate accuracy in farther alike inputs.

$$(\frac{1}{t} \sum_{i=1}^{t} x_i) \tag{4.6}$$

Here, t=total frequency, $x_i = i^{th}$ term

Max

This value shows the highest ranked frequency of a signal in a state of mind.For each range of frequency max will be non identical.This feature shows the maximum point of brain signal in a particular range of a participant.One can not feel overwhelmed of any emotion after this point according to the system.

Skewness

When the distribution of the data is not in normal distribution and the values are not equally separated by the mid value line in the data curve is considered to be skewed in either side of the midpoint. Whether it is positively skewed where the tail of the curve lies towards the upper value or negatively skewed where tail is on the lower value side. In feature selection of signals skewness detects the majority of the value lies on which side of the midline.

Variance

Variance identifies the dispersion of the data set, more accurately than mean.The smaller the variance the more disperse the data set is.This feature shows us how spread our data points are within the range.

$$\tau^2 = \frac{\sum_{t=1}^{p} (z_i - \mu)^2}{p} \tag{4.7}$$

Here μ is mean amd τ is standard deviation
Here total 5 features are extracted from 14 channels,resulting (5*14) = 70 features in total.

4.4.5 Poincare

The Poincare, which takes a series of intervals and plots each interval against the following interval, is an emerging analysis technique. In clinical settings, the geometry of this plot has been shown to differentiate between safe and unhealthy subjects.It is also used in a time series for visualizing and quantifying the association between two consecutive data points. Since long-term correlation and memory are demonstrated in the dynamics of variations in physiological rhythms, this analysis were meant to expand the plot of Poincare by steps, instead of between two consecutive points, the association between sequential data points in a time sequence. We used two parameters in our paper which are:

SD1: Represent Standard deviation from axis 1 of the distances of points and defines the width from the ellipse (short-term variability).Descriptors SD1 can be defined as:

$$SD1 = \frac{\sqrt{2}}{2} SD(P_n - P_{n+1}) \tag{4.8}$$

SD2: The standard deviations from axis 2 and ellipse length (long-term variability) are equivalent to SD2.Descriptors SD2 can be defined as:

$$SD2 = \sqrt{2SD(P_n)^2 - \frac{1}{2} SD(P_n - P_{n+1})^2} \tag{4.9}$$

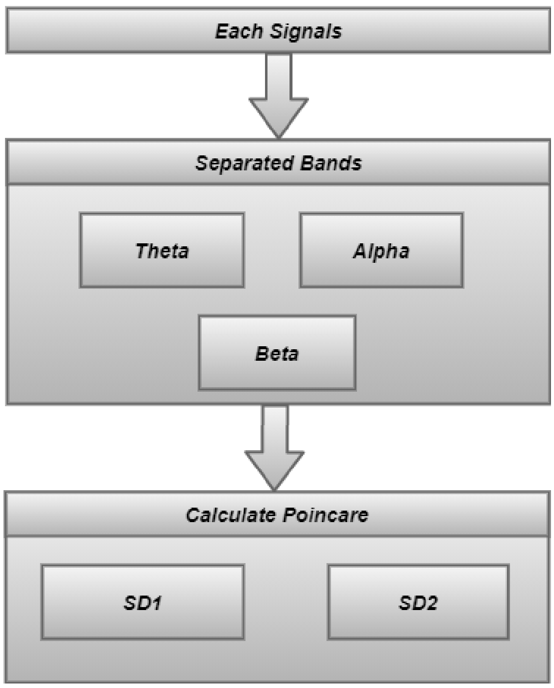


Figure 4.12: Poincare feature extraction steps

In the diagram, it is shown that, we have exacted 2 features which are SD1 and SD2 from each band(theta, alpha, beta) of all the 14 channels,
resulting, (SD1,SD2)*3*14 = 84 features.

4.4.6 Power Spectral Density

The Welch method is a modified segmentation system and is used to assess the average periodogram which is used in paper [5],[15], [27]. Welch method is applied on a time series. For spectral density it is concerned with decreasing the variance in the results.Power spectral density (PSD) informs us which differences in frequency ranges are high and could be very helpful for further study.The Welch method of the PSD can usually be described by the following equations of the power spectra.

$$P(f) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} x_i(n) w(n) e^{-j2\pi f} \right|^2 \tag{4.10}$$

$$P_{\text{welch}}(f) = \frac{1}{L} \sum_{i=0}^{L-1} P(f) \tag{4.11}$$

Here, The equation of density is defined first. Then, Welch Power Spectrum implies that for each interval, the average time is expressed. We have implemented this Welch method to get the PSD(Power Spectral density) of the signal, from that, the mean power has been extracted from each band.

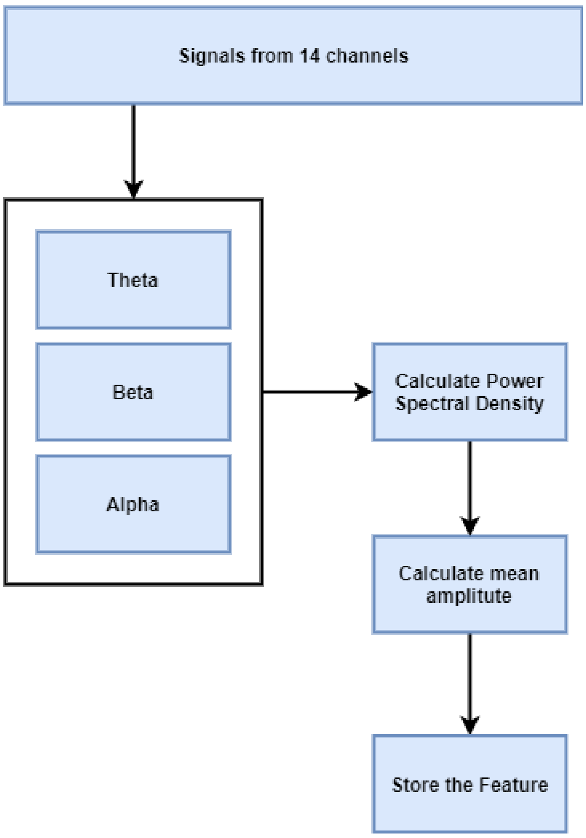


Figure 4.13: PSD feature extraction steps

Total of, [(mean)*(3 bands) = 3 Features for each channel So for 14 channels we got a total, [14 channel *3 Features] = 42 features.

4.4.7 All Extracted Features

Feature technique	Extracted features
Statistical features	Mean, Median, Maximum, Skewness, Variance
Fast Fourier Transform	Mean, Max of coefficients
Discrete cosine transform	Mean, Max of coefficients
Hjorth parameters	Activity, Mobility, Complexity
Poincase	SD1, SD2
Power Spectral density	Mean amplitude

Table 4.3: Extracted features

4.5 Feature Ratio

As we have the signals of pre-trail, from that we have used 4 seconds of pre-trail signals as baseline signals resulting 4*128 = 512 samples for each signal. Then similar to the feature extraction for stimuli, the features from baseline signals were also extracted. Then the stimuli features were divided by the baseline features, in order to get only the differences which can be noticed by the stimuli test only, which is also done in paper [27].

4.6 Feature Scaling

After Extracting all the features and calculating the ratio between stimuli features and baseline features we have added the self assessment ratings of arousal valence and dominance. Now the final data set have 414 data with 630 features for each data. So the data set structure is shown below,

	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Feature9	Feature10	...
0	0.399810	-1.183565	0.800690	0.328107	-4.978389	0.229131	0.177772	0.888415	0.739250	-0.113158	...
1	-0.184464	-0.979553	0.102095	0.028185	-0.202161	-0.245509	0.023804	0.243781	0.124302	-0.174262	...
2	-10.379534	0.374143	0.682109	0.575803	-4.842312	0.017298	-6.372514	1.014596	1.212745	0.411182	...
3	0.166460	1.676886	0.146557	0.043448	0.343049	0.185492	-0.212113	0.212338	0.067883	0.197870	...
4	0.935605	4.578753	0.913949	0.530971	4.916748	1.003016	-0.001346	0.484388	0.253078	34.804499	...
...	...	...	...	...	...	...	...	...	...	...	...
409	5.545656	3.336175	2.061960	4.137637	0.466562	-1.281754	-1.976419	0.929064	1.169510	-1.670856	...
410	-1.241643	0.112553	0.689308	0.517814	-5.058245	1.180783	1.778832	2.183805	2.919873	1.604152	...
411	-2.638844	-2.514022	1.287162	2.850588	-0.966474	-5.171990	-0.366512	2.988450	3.993974	-1.384842	...
412	1.400873	-0.140568	1.055900	1.448882	1.341685	-0.178661	-0.178995	1.312953	1.797478	-0.039673	...
413	0.735351	1.144291	1.696663	3.102877	0.925788	-1.144911	-0.469799	1.650834	2.591888	-2.635898	...

414 rows × 630 columns

Figure 4.14: All Features (Not scaled)

In order to get rid of high variance in our data set, we have scaled the data using Min Max Scaling approach from python sklearn library. In Min-Max Scaling, the estimator scales and translates each value individually so that it is in between 0 and 1, in the specified range.

The formula for Min Max scale is,

$$X_{new} = \frac{X_i - \mathrm{Min}(X)}{\mathrm{Max}(X) - \mathrm{Min}(X)} \tag{4.12}$$

So, We have scaled our data using min-max scaling where the values of the features range between 0 to 1,

Finally the data set we have is,

	0	1	2	3	4	5	6	7	8	9	...
0	0.284999	0.517307	0.009022	0.000070	0.473111	0.315669	0.414742	0.023107	0.000783	0.986850	...
1	0.284343	0.518913	0.001006	0.000006	0.485028	0.314685	0.414378	0.006285	0.000132	0.986847	...
2	0.272907	0.529566	0.007661	0.000124	0.473451	0.315229	0.399250	0.026400	0.001284	0.986877	...
3	0.284737	0.539817	0.001516	0.000009	0.486388	0.315578	0.413820	0.005464	0.000072	0.986866	...
4	0.285600	0.562653	0.010321	0.000114	0.497799	0.317273	0.414319	0.012564	0.000268	0.988621	...
...	...	...	...	...	...	...	...	...	...	...	...
409	0.290771	0.552875	0.023494	0.000889	0.486696	0.312536	0.409647	0.024168	0.001238	0.986771	...
410	0.283157	0.527507	0.007744	0.000111	0.472912	0.317642	0.418529	0.056912	0.003091	0.986937	...
411	0.281590	0.506838	0.014604	0.000613	0.483121	0.304470	0.413455	0.077910	0.004228	0.986786	...
412	0.286122	0.525515	0.011950	0.000311	0.488879	0.314823	0.413899	0.034186	0.001903	0.986854	...
413	0.285375	0.535626	0.019302	0.000667	0.487842	0.312820	0.413211	0.043003	0.002744	0.986722	...

414 rows × 630 columns

Figure 4.15: All Features (scaled)

4.7 Feature Selection Techniques

There are various feature selection techniques which are used by many researchers, to reduce the number of features which are not needed and only the important features which can play a big role in the prediction. So in our paper we used two feature elimination methods. One is Recursive Feature Elimination (RFE) and another one is Chi-square test (chi2).

4.7.1 Recursive feature elimination

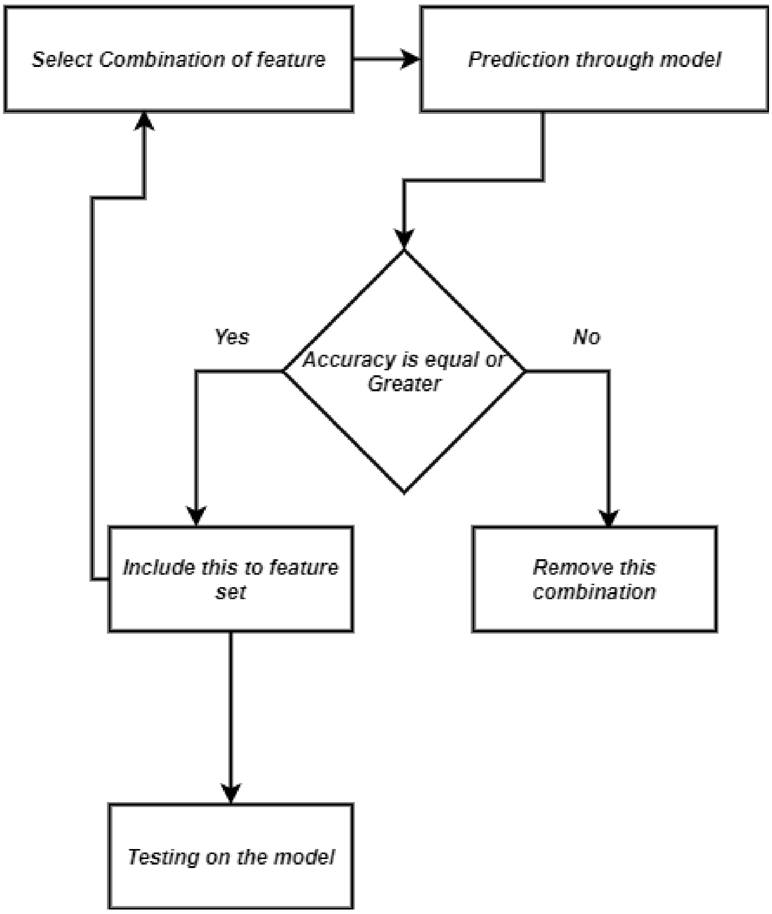


Figure 4.16: Diagram of Recursive Feature Elimination (RFE)

RFE is a wrapper type feature selection technique amongst the vast span of features. Here the term recursive is representative of the loop work of this method that traverses backward on loops to identify the best fitted feature giving each predictor an importance score and later eliminating the least importance scored predictor. Additionally cross-validation is used to find the optimal number of features to rank various feature subcategories and pick the best selection of features for scoring. In this method one attribute is taken and along with the target attribute and this procedure keeps forwarding combining attributes and merging with the target attribute to produce a new model. Thus different subsets of features of different combinations generate models through training. All these models are then strained out to get the maximum accuracy resulting model and its consecutive features. In short ,we remove those features which result in the accuracy to be high or at least equal and return it back if the accuracy gets low after elimination . Here we have used step size of 1 to eliminate one feature at a time at each level which can help to remove the worst features early, keeping the best features in order to improve the already calculated accuracy of the overall model.

4.7.2 Chi-Square

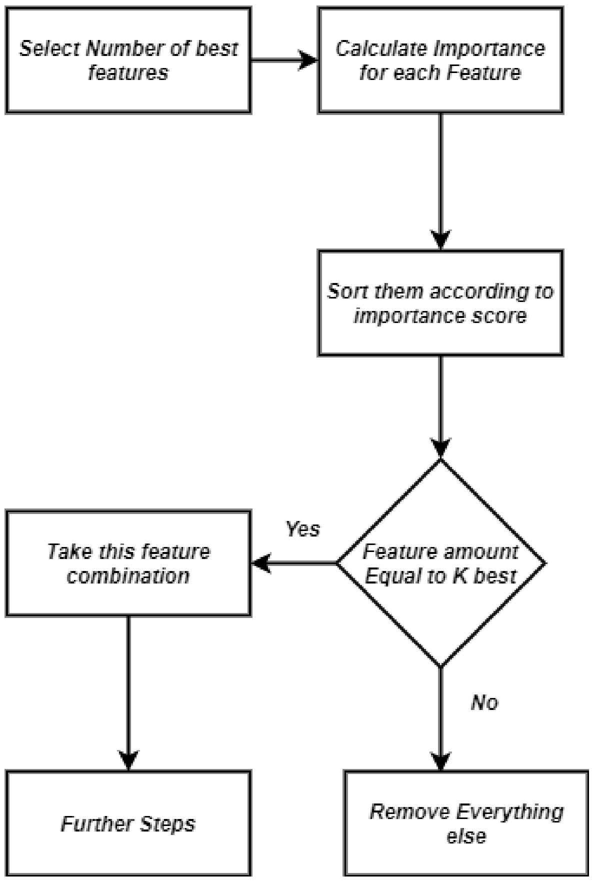


Figure 4.17: Select K best features using Chi-square

Chi-square (chi2) test is a filter method that states the accuracy of a system comparing the predicted data with the observed data based on their importance. It is a test that figures out if there is any feature effective on nominal categorized data or not in order to compare between observed and expected data. In this method one predicted data set is considered as a base point and expected data is calculated from the observed value with respect to the base point.

The Chi-square value is computed by:

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{\left(A_{ij} - \frac{R_i \cdot C_j}{N}\right)^2}{\frac{R_i \cdot C_j}{N}} \tag{4.13}$$

Here, from the above equation,
m = number of intervals,
k = amount of classes,
R_i = amount of patterns in the I^{th} range,
C_j = amount of patterns in the j^{th} range,
$A_i j$= amount of patterns in I^{th} and j^{th} range

After applying RFE and chi2 , from the achieved accuracy we have observed that, chi2 does not incorporate a machine learning (ML) model, while RFE uses a machine learning model and trains it to decide whether it is relevant or not. Moreover, in our research, Chi-square methods failed to choose the best subset of features which can provide better results,but because of the extensive nature, RFE methods give the best subset of features mostly in our research. Therefore we consider RFE over Chi-square for feature elimination.

4.8 Class distribution of Self assessment ratings

In the previous research [27], on this data set, they have calculated the mean and standard deviation for the self assessment of each dimension of emotions. Then they have divided each dimension into two classes which are high or low. As, rating scale of the self assessment was 0 to 5, the boundary between high and low was in the mid point which is 2.5. But this has nothing to do with the categories of emotions. So we had to adjust this boundary based on some of our observation. We have also calculated the mean and standard deviation for each self assessment ratings, shown below:

Stimuli number	Clip Names	Concentrated emotion	Mean,SD Valence	Mean,SD Arousal	Mean,SD Dominance
1	Searching for Bobby Fischer	calm	3.17 ± 0.72	2.26 ± 0.75	2.09 ± 0.73
2	D-O-A	surprising	3.04 ± 0.88	3.00 ± 1.00	2.70 ± 0.88
3	The Hangover	amuse	4.57 ± 0.73	3.83 ± 0.83	3.83 ± 0.72
4	The Ring	fearness	2.04 ± 1.02	4.26 ± 0.69	4.13 ± 0.87
5	300	excite	3.22 ± 1.17	3.70 ± 0.70	3.52 ± 0.95
6	National Lampoon's Van Wilder	disgusting	2.70 ± 1.55	3.83 ± 0.83	4.04 ± 0.98
7	Wall-E	happy	4.52 ± 0.59	3.17 ± 0.98	3.57 ± 0.99
8	Crash	angry	1.35 ± 0.65	3.96 ± 0.77	4.35 ± 0.65
9	My Girl	sad	1.39 ± 0.66	3.00 ± 1.09	3.48 ± 0.95
10	The Fly	disgusting	2.17 ± 1.15	3.30 ± 1.02	3.61 ± 0.89
11	Pride and Prejudice	calm	3.96 ± 0.64	1.96 ± 0.82	2.61 ± 0.89
12	Modern Times	amuse	3.96 ± 0.56	2.61 ± 0.89	2.70 ± 0.82
13	Remember the Titans	happy	4.39 ± 0.66	3.70 ± 0.97	3.74 ± 0.96
14	Gentleman's Agreement	angry	2.35 ± 0.65	2.22 ± 0.85	2.39 ± 0.72
15	Psycho	fearness	2.48 ± 0.85	3.09 ± 1.00	3.22 ± 0.9
16	The Bourne Identity	excite	3.65 ± 0.65	3.35 ± 1.07	3.26 ± 1.14
17	The Shawshank Redemption	sad	1.52 ± 0.59	3.00 ± 0.74	3.96 ± 0.77
18	The Departed	surprising	2.65 ± 0.78	3.91 ± 0.85	3.57 ± 1.04

Table 4.4: Mean and Standard deviation of self assessment ratings of all dimensions

We have calculated this to separate each dimension of emotions into two separate classes, which will be high(1) and low(0) and will be representing two emotional category for each dimension.

Arousal: It is separated into two classes for our research, which are excited/alert and uninterested/bored. First of all we have focused on the average ratings for excitement which co-responds to stimuli number 5 and 16, having 3.70±0.70 and 3.35±1.07 respectively. Additionally for, calmness, we can take stimuli 1 and 11 into consideration where the average ratings are, 2.26±0.75 and 1.96±0.82 respectively. So here, it is clear that, (ratings > 2) can be considered in the class of Excited/Alert and (ratings< 2) can be considered as Uninterested/Bored.

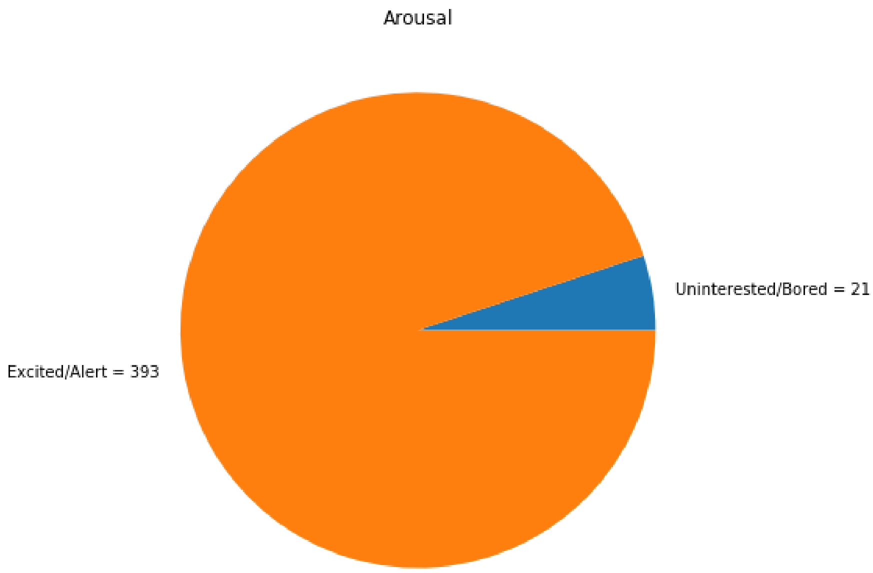


Figure 4.18: Arousal Class Distribution

Here, from the 414 data, 393 was in the excited/alert class and 21 was in the uninterested/bored class. To store this values in the new data set, uninterested/bored is considered as 0 and excited/alert is considered as 1.

Valence: It is separated into two classes for our research, which are unpleasant/stressed and happy/elated. First of all we have focused on the average ratings for happiness which co-responds to stimuli number 7 and 13, having 4.52±0.59 and 4.39±0.66 respectively. Additionally for, fear and disgust, we can take stimuli (4,15) and (6,10) into consideration where the average ratings are, 2.04±1.02, 2.48±0.85, 2.70±1.55 and 2.17±1.15 respectively. So here, it is clear that, ratings > 4 can be considered in the class of happy/elated and ratings< 4 can be considered as unpleasant/stressed.

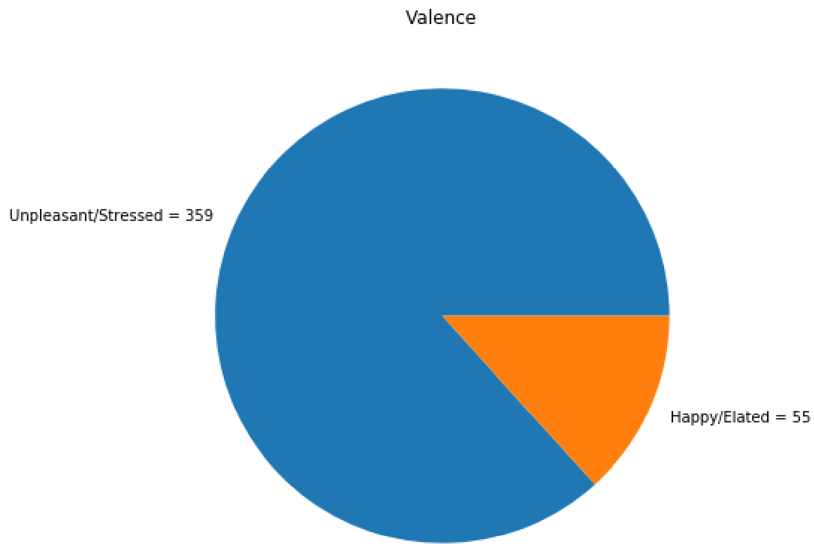


Figure 4.19: Valence Class Distribution

Here, from the 414 data, 359 was in the unpleasant/stressed class and 55 was in happy/elated class. To store this values in the new data set, unpleasant/stressed is considered as 0 and happy/elated is considered as 1.

Dominance: It is separated into two classes for our research, which are helpless/without control and empowered. Here, we have targeted stimuli number, 4,6,8 which has targeted emotions of fear, disgust and anger, having mean rating of 4.13 ± 0.87, 4.04 ± 0.98 and 4.35 ± 0.65 respectively. So we can consider, ratings> 4 in the class of helpless/without control and rest for the class of empowered.

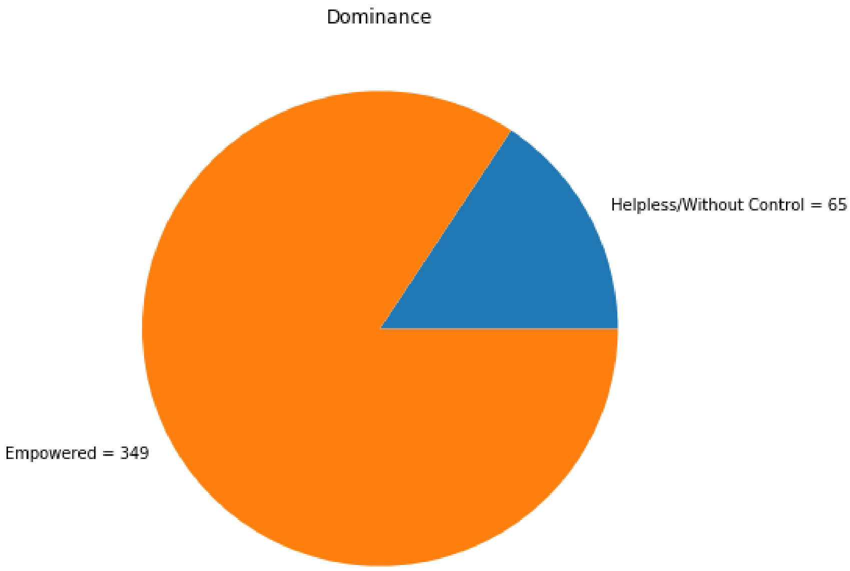


Figure 4.20: Dominance Class Distribution

Here, from the 414 data, 65 was in the helpless/without control class and 349 was in empowered class. To store this values in the new data set, helpless/Without Control is considered as 0 and empowered is considered as 1.

Chapter 5

Results and Discussion

There was multiple machine learning models such as SVM, also used in [3] and XGBoost also used in [35] was used for classification of emotion. In our proposed model, we have used the DREAMER dataset along with various feature extraction techniques mentioned in chapter 4, then have used feature elimination technique to get the best features out of the total extracted features. Here, Recursive feature elimination was our primary choice. Additionally we have done 10 fold cross validation in our accuracy measurement. Finally The results that we have focused on are, minimum, mean and maximum accuracy of all three dimensions of emotion, which are arousal, valence and dominance.

Firstly we have collected the dataset from a secondary source which is proposed in paper [27]. As the dataset was not preprocessed we had to process the data according to our need. In the dataset, as there was 23 subjects tested on 18 stimuli with 14 channel device at 128hz sampling rate, the duration of the recorded signal was 65 to 393, resulting unequal number of sample points. So here, we have only selected the last 60s of the recorded signal resulting total of 60*128 = 7680 points for stimuli. Additionally as there was pre-trail signals in the dataset as baseline for each stimuli, we have selected only 4s of data from that, resulting 4*128 = 512 pre-trail sample points. Then we have divided each channel into 3 separate bands which are theta, alpha and beta. From the extracted band, we have performed feature extraction techniques such as FFT, DCT, Poincare, Power Spectral Density, Statistical features and measuring Hjorth parameters resulting total of 630 features from 14 channels for each stimuli test and also for each baseline.

After extracting the features, we calculated the ratio between the features from stimuli and baseline, to get the actual feature values which is affected by the stimuli only.Then we have used chi-square and feature elimination technique to get the most important features from the total 630 extracted features. From the two mentioned feature selection technique, in our proposed model, the recursive feature elimination technique increased the accuracy of the model.

Finally we have used our proposed machine learning algorithms on the final dataset consisting of 414 data from 23 subjects tested on 18 stimuli. For that, firstly, we have used SVM with RBF(Gaussian radial basis function) kernel to test our model. Here, we have also used 10 fold cross validation for all the dimensions(arousal, valence, dominance) to calculate the minimum, mean and maximum accuracy of the prediction and show the variation between the minimum, mean and maximum. Here, we have achieved mean accuracy of 94.930% for arousal, 86.725% for valence and 84.309% for dominance.

Accuracy	***Arousal***	***Valence***	***Dominance***
Minimum	92.857%	85.366%	82.927%
Mean	94.930%	86.725%	84.309%
Maximum	95.238%	87.805%	85.366%

Table 5.1: SVM Accuracy Results

Then we have also used Extreme gradient boosting(XGBoost) on the final dataset. Here, we have used the following parameters for XGBoost model,

Parameters	***Value***
Max depth	3
Number of estimators	300
Learning rate	0.05
Early stopping rounds	10
Random state	0

Table 5.2: XGB parameters

with all these parameters of the model, we have again performed 10 fold cross validation for all the dimensions(arousal, valence, dominance) to calculate the minimum, mean and maximum accuracy of the prediction and show the variation between the minimum, mean and maximum. Here, we have achieved mean accuracy of 95.174% for arousal, 87.456% for valence and 84.541% for dominance.

Accuracy	***Arousal***	***Valence***	***Dominance***
Minimum	92.857%	83.333%	82.927%
Mean	95.174%	87.456%	84.541%
Maximum	97.561%	90.244%	85.714%

Table 5.3: XGBoost Accuracy Results

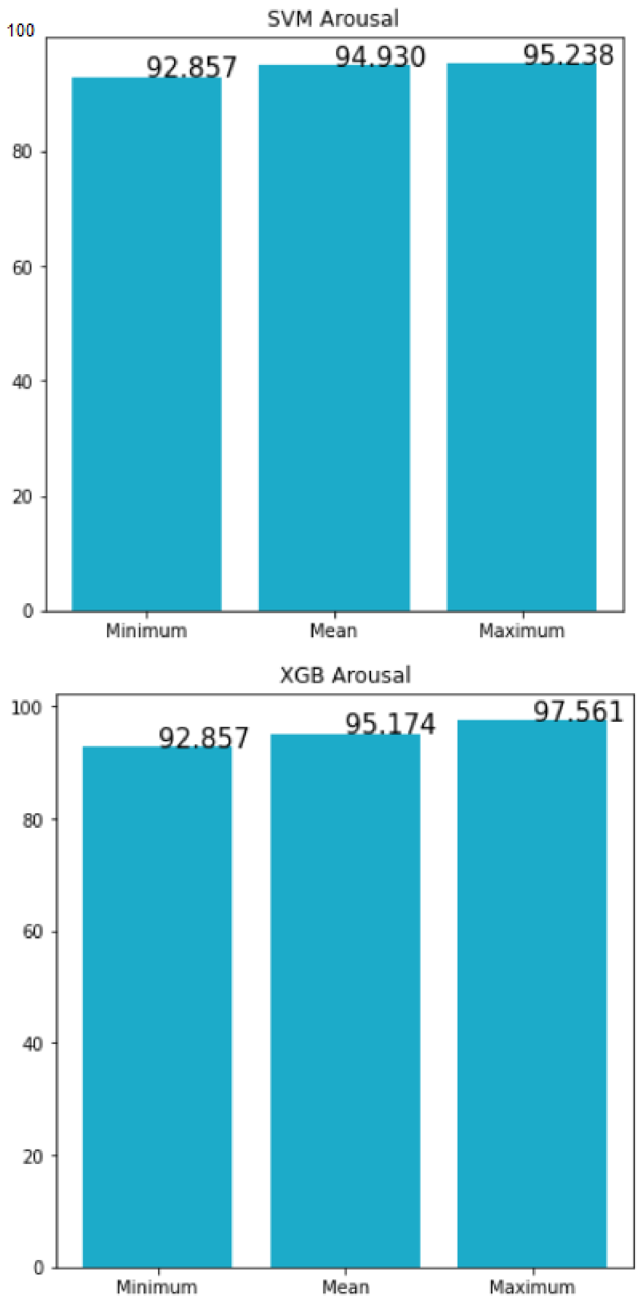


Figure 5.1: Comparison of Arousal Scores

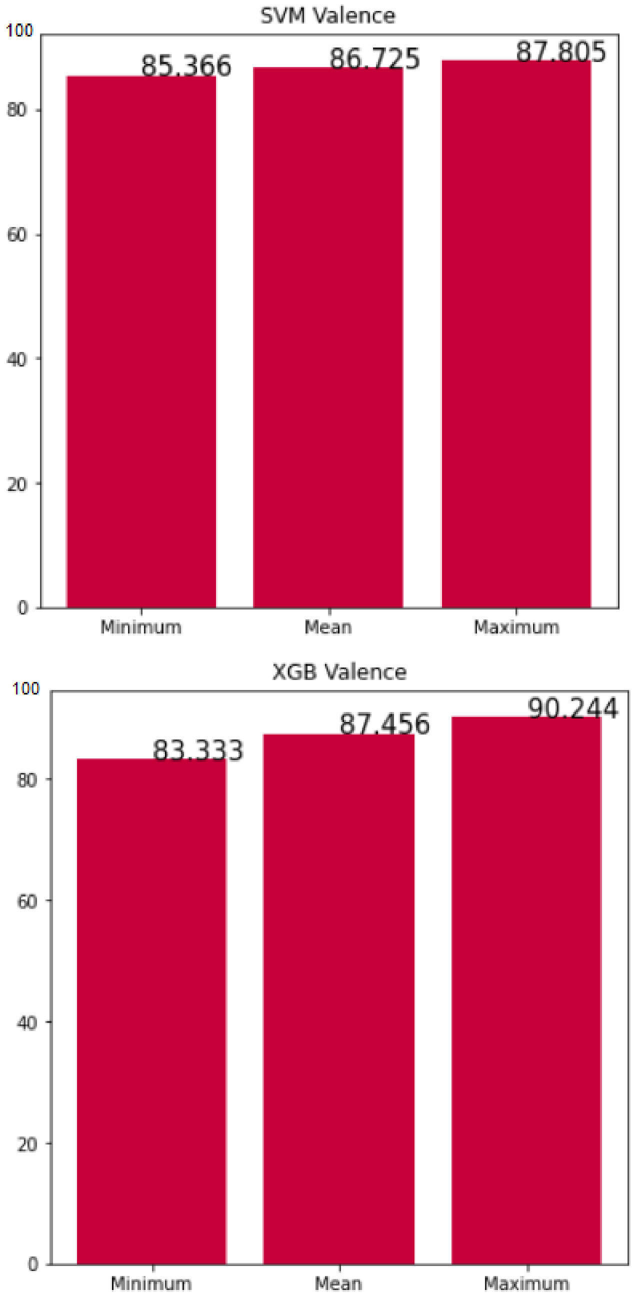


Figure 5.2: Comparison of Valence Scores

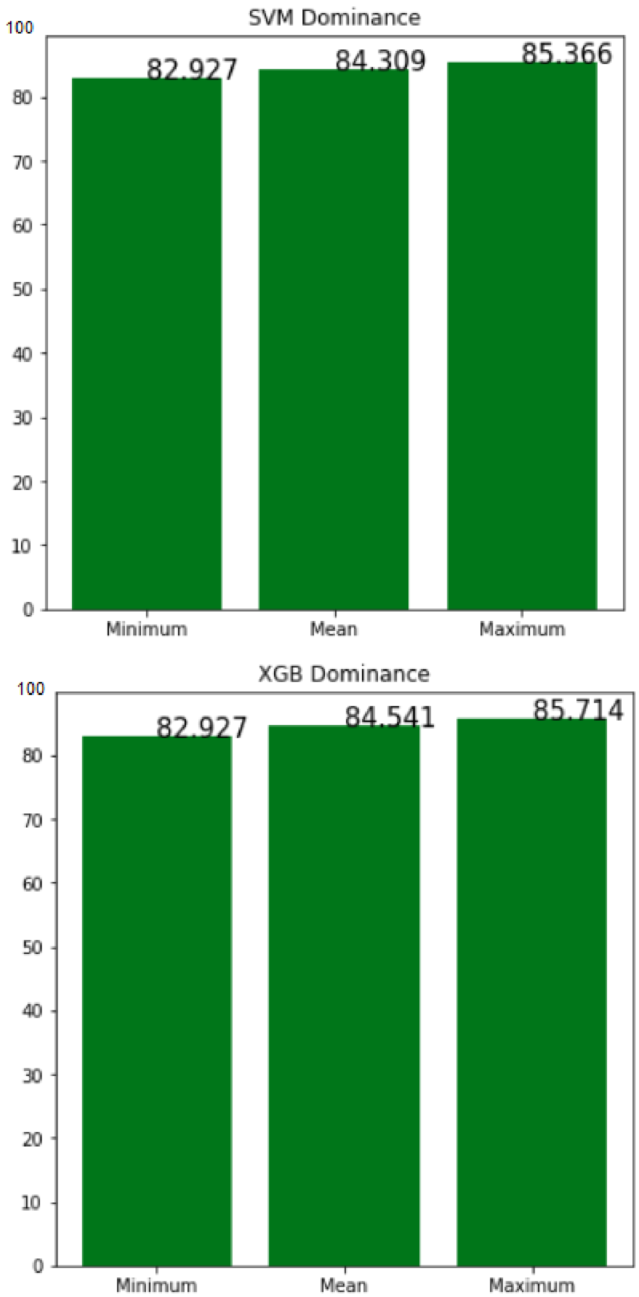


Figure 5.3: Comparison of Dominance Scores

Chapter 6

Conclusion

6.1 Conclusion

In today's world use of technology is becoming our first priority in every task and also a basic need.That is why in this report we have focused on a specific area which is application of brain computer interfacing. As we know, computers can not interact with the people as human do, the inability to understand emotions is a barrier to further innovation. In this research we tried to overcome those limitations, by contributing in the area of analysing the affective state of human. In our proposed model, we have collected dataset called DREAMER, where EEG signals from many participants were recorded for further analysis of emotion. We have used various feature extraction techniques, such as FFT, DCT, Poincare, Statistical features, Hjorth parameters, Power spectral density along with feature elimination technique such as recursive feature elimination. Then we have used SVM and XG-Boost to classify them. From the analysis, we have reached to conclusion that, by using our model, extreme gradient boosting provided us better results with mean accuracy of 95.174% for arousal, 87.456% for valence and 84.541% for dominance, which is significantly higher than the research previously done on this data set.

6.2 Limitation and Future Work

In this research, the main limitation of availability of the data, as we have worked on a dataset consisting of data from 23 subjects tested on 18 stimuli, we have got 414 data from the dataset. It would be great if we could arrange a testing environment and gather more data personally but due to the global pendamic of COVID19 we could not do that. Another challenge is the cost of equipment that may arise, while conducting physical testing. The subject of our research is relatively new, and as such it will not immediately have an impact. For now the applications it has are limited. But that is how every new technology starts off. After further development it can be used various places. If we can use it continuously and on a real time environment it can be used to prevent suicides, or even crimes. If social media platforms used our system then it could provide users with appropriate content based on the emotion of the used. Emotion has a huge impact on our lives, it is what dictates our actions. If we can make machines responsive to emotion then the potentials are limitless.

Bibliography

[1] Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," *International journal of computer vision*, vol. 43, no. 3, pp. 167–188, 2001.

[2] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 2, pp. 458–466, 2006.

[3] M. Sabeti, R. Boostani, S. Katebi, and G. Price, "Selection of relevant features for eeg signal classification of schizophrenic patients," *Biomedical Signal Processing and Control*, vol. 2, no. 2, pp. 122–134, 2007.

[4] G. U. Kharat and S. V. Dudul, "Neural network classifier for human emotion recognition from facial expressions using discrete cosine transform," in *2008 First International Conference on Emerging Trends in Engineering and Technology*, IEEE, 2008, pp. 653–658.

[5] K.-E. Ko, H.-C. Yang, and K.-B. Sim, "Emotion recognition using eeg signals with relative power values and bayesian network," *International Journal of Control, Automation and Systems*, vol. 7, no. 5, p. 865, 2009.

[6] M. Murugappan, M. R. B. M. Juhari, R. Nagarajan, and S. Yaacob, "An investigation on visual and audiovisual stimulus based emotion recognition using eeg," *International Journal of Medical Engineering and Informatics*, vol. 1, no. 3, pp. 342–356, 2009.

[7] ——, "An investigation on visual and audiovisual stimulus based emotion recognition using eeg," *International Journal of Medical Engineering and Informatics*, vol. 1, no. 3, pp. 342–356, 2009.

[8] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," *IEEE Transactions on information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2009.

[9] Y. Chavhan, M. Dhore, and P. Yesaware, "Speech emotion recognition using support vector machine," *International Journal of Computer Applications*, vol. 1, no. 20, pp. 6–9, 2010.

[10] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.

[11] A. Georgogiannis and V. Digalakis, "Speech emotion recognition using non-linear teager energy based features in noisy environments," in *2012 proceedings of the 20th European signal processing conference (EUSIPCO)*, IEEE, 2012, pp. 2045–2049.

[12] M. Murugappan and S. Murugappan, "Human emotion recognition through short time electroencephalogram (eeg) signals using fast fourier transform (fft)," in *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, IEEE, 2013, pp. 289–294.

[13] T. S. Rached and A. Perkusich, "Emotion recognition based on brain-computer interface systems," *Brain-computer interface systems-Recent progress and future prospects*, pp. 253–270, 2013.

[14] S. Jerritta, M. Murugappan, W. Khairunizam, and S. Yaacob, "Electrocardiogram-based emotion recognition system using empirical mode decomposition and discrete fourier transform.," *Expert Syst. J. Knowl. Eng.*, vol. 31, no. 2, pp. 110–120, 2014.

[15] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[16] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "Eeg-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2014, pp. 1–6.

[17] F. Zhou, X. Qu, J. Jiao, and M. G. Helander, "Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors," *Interacting with computers*, vol. 26, no. 3, pp. 285–302, 2014.

[18] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom, "Facial emotion recognition using empirical mode decomposition," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1261–1277, 2015.

[19] A. Bhardwaj, A. Gupta, P. Jain, A. Rani, and J. Yadav, "Classification of human emotions from eeg signals using svm and lda classifiers," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2015, pp. 180–185.

[20] C. A. Gabert-Quillen, E. E. Bartolini, B. T. Abravanel, and C. A. Sanislow, "Ratings for emotion film clips," *Behavior research methods*, vol. 47, no. 3, pp. 773–787, 2015.

[21] R. M. Mehmood and H. J. Lee, "Eeg based emotion recognition from human brain using hjorth parameters and svm," *International Journal of Bio-Science and Bio-Technology*, vol. 7, no. 3, pp. 23–32, 2015.

[22] ——, "Eeg based emotion recognition from human brain using hjorth parameters and svm," *International Journal of Bio-Science and Bio-Technology*, vol. 7, no. 3, pp. 23–32, 2015.

[23] S. Paul, A. Mazumder, P. Ghosh, D. Tibarewala, and G. Vimalarani, "Eeg based emotion recognition system using mfdfa as feature extractor," in *2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE)*, IEEE, 2015, pp. 1–5.

[24] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2015.

[25] A. E. Vijayan, D. Sen, and A. Sudheer, "Eeg-based emotion recognition using statistical measures and auto-regressive modeling," in *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, IEEE, 2015, pp. 587–591.

[26] A. Patil, C. Deshmukh, and A. Panat, "Feature extraction of eeg for emotion recognition using hjorth features and higher order crossings," in *2016 Conference on Advances in Signal Processing (CASP)*, IEEE, 2016, pp. 429–434.

[27] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.

[28] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain eeg sensors," *Ieee Access*, vol. 5, pp. 14 797–14 806, 2017.

[29] M. Degirmenci, M. A. Ozdemir, R. Sadighzadeh, and A. Akan, "Emotion recognition from eeg signals by using empirical mode decomposition," in *2018 Medical Technologies National Congress (TIPTEKNO)*, IEEE, 2018, pp. 1–4.

[30] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring eeg features in cross-subject emotion recognition," *Frontiers in neuroscience*, vol. 12, p. 162, 2018.

[31] W. Chang, Y. Liu, Y. Xiao, X. Yuan, X. Xu, S. Zhang, and S. Zhou, "A machine-learning-based prediction method for hypertension outcomes based on medical data," *Diagnostics*, vol. 9, no. 4, p. 178, 2019.

[32] R. Hartmann, F. M. Schmidt, C. Sander, and U. Hegerl, "Heart rate variability as indicator of clinical state in depression," *Frontiers in psychiatry*, vol. 9, p. 735, 2019.

[33] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10–18, 2019.

[34] S. Murty and R. K. Kumar, "Accurate liver disease prediction with extreme gradient boosting," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 2249–8958, 2019.

[35] S. Parui, A. K. R. Bajiya, D. Samanta, and N. Chakravorty, "Emotion recognition from eeg signal using xgboost algorithm," in *2019 IEEE 16th India Council International Conference (INDICON)*, IEEE, 2019, pp. 1–4.

[36] D. Acharya, A. Billimoria, N. Srivastava, S. Goel, and A. Bhardwaj, "Emotion recognition using fourier transform and genetic programming," *Applied Acoustics*, vol. 164, p. 107 260, 2020.

[37] M. Kaur, N. R. Prakash, P. Kalra, G. D. Puri, T. Samra, and M. Goyal, "Variations in electrocortical activity due to surgical incision in anaesthetized cardiac patients: Electroencephalogram-based quantitative analysis," *Journal of Healthcare Engineering*, vol. 2020, 2020.

Appendix A

CODES

A.1 Feature Extraction Codes

```
# FFT Features

theta_FFT = fft(theta_signal) # Calculating FFT of Theta Band
Ceff1 = 2 / 128 * np.abs(theta_FFT[0:128]) # Values from (0-128)Hz Frequency.

alpha_FFT = fft(alpha_signal)# Calculating FFT of Alpha Band
Ceff2 = 2 / 128 * np.abs(alpha_FFT[0:128]) # Values from (0-128)Hz Frequency.

beta_FFT = fft(beta_signal)# Calculating FFT of Beta Band
Ceff3 = 2 / 128 * np.abs(beta_FFT[0:128]) # Values from (0-128)Hz Frequency.

FFT1 = np.max(Ceff1)
FFT2 = np.mean(Ceff1)

FFT3 = np.max(Ceff2)
FFT4 = np.mean(Ceff2)

FFT5 = np.max(Ceff3)
FFT6 = np.mean(Ceff2)
```

Code: FFT

```
#DCT Features
theta_DCT = dct(theta_signal)# Calculating FFT of Theta Band
Ceff4= 2 / 128 * np.abs(theta_DCT[0:128]) # Values from (0-128)Hz Frequency.

alpha_DCT = dct(alpha_signal)# Calculating FFT of Alpha Band
Ceff5 = 2 / 128 * np.abs(alpha_DCT[0:128]) # Values from (0-128)Hz Frequency.

beta_DCT = dct(beta_signal)# Calculating FFT of Beta Band
Ceff6 = 2 / 128 * np.abs(beta_DCT[0:128]) # Values from (0-128)Hz Frequency.

DCT1 = np.max(Ceff4)
DCT2 = np.mean(Ceff4)

DCT3 = np.max(Ceff5)
DCT4 = np.mean(Ceff5)

DCT5 = np.max(Ceff6)
DCT6 = np.mean(Ceff6)
```

Code: DCT

```
def hjorth(a):

    first_deriv = np.diff(a)
    second_deriv = np.diff(a,2)

    var_zero = np.mean(a ** 2)
    var_d1 = np.mean(first_deriv ** 2)
    var_d2 = np.mean(second_deriv ** 2)

    activity = var_zero
    mobility = np.sqrt(var_d1 / var_zero)
    complexity = np.sqrt(var_d2 / var_d1) / mobility

    return activity, complexity, mobility
```

```
# Hjorth Parameters

activity1, complexity1, mobility1 = hjorth(theta_signal)
activity2, complexity2, mobility2 = hjorth(alpha_signal)
activity3, complexity3, mobility3 = hjorth(beta_signal)
```

Code: Hjorth Parameters

```
# PoinCare - SD1, SD2
theta_data = theta_signal
theta_diff = np.diff(theta_data)
theta_SDSD = np.std(theta_diff)
theta_SDNN = np.std(theta_data)
theta_SD1 = np.sqrt(0.5 * np.power(theta_SDSD, 2))
theta_SD2 = np.sqrt(2 * np.power(theta_SDNN, 2) - np.power(theta_SD1, 2))

alpha_data = alpha_signal
alpha_diff = np.diff(alpha_data)
alpha_SDSD = np.std(alpha_diff)
alpha_SDNN = np.std(alpha_data)
alpha_SD1 = np.sqrt(0.5 * np.power(alpha_SDSD, 2))
alpha_SD2 = np.sqrt(2 * np.power(alpha_SDNN, 2) - np.power(alpha_SD1, 2))

beta_data = beta_signal
beta_diff = np.diff(beta_data)
beta_SDSD = np.std(beta_diff)
beta_SDNN = np.std(beta_data)
beta_SD1 = np.sqrt(0.5 * np.power(beta_SDSD, 2))
beta_SD2 = np.sqrt(2 * np.power(beta_SDNN, 2) - np.power(beta_SD1, 2))
```

Code: Poincare

```
def PSD_Welch(theta_signal, alpha_signal, beta_signal):
    th_frequency, th_power = signal.welch(theta_signal, fs=len(theta_signal), nperseg=256, noverlap=128)
    al_frequency, al_power = signal.welch(alpha_signal, fs=len(alpha_signal), nperseg=256, noverlap=128)
    th_frequency, bt_power = signal.welch(beta_signal, fs=len(beta_signal), nperseg=256, noverlap=128)

    return th_power, al_power, bt_power

```

```
# Finding PSD
power_theta, power_alpha, power_beta = PSD_Welch(theta_signal, alpha_signal, beta_signal)

power_theta = np.max(power_theta)
power_alpha = np.max(power_alpha)
power_beta = np.max(power_beta)
```

Code: Power Spectral density

```
#Statistical Features - (Mean, Mediun, Maximum, Var, Skewness)
theta_mean = np.mean(theta_signal)
theta_median = np.median(theta_signal)
theta_max = np.max(theta_signal)
theta_var = np.var(theta_signal)
theta_skewness = pd.Series(theta_signal).skew()

alpha_mean = np.mean(alpha_signal)
alpha_median = np.median(alpha_signal)
alpha_max = np.max(alpha_signal)
alpha_var = np.var(alpha_signal)
alpha_skewness = pd.Series(alpha_signal).skew()

beta_mean = np.mean(beta_signal)
beta_median = np.median(beta_signal)
beta_max = np.max(beta_signal)
beta_var = np.var(beta_signal)
beta_skewness = pd.Series(beta_signal).skew()
```

Code: Statistical Features

A.2 Class Distribution of Self Assessment Ratings

```python
#Class Distribution of Self assesment ratings for Arousal, Valence and Dominance

ratings = pd.read_csv('/content/drive/My Drive/Colab Notebooks/THESIS/Results/AllRatings.csv')
AllRatings = pd.DataFrame(ratings)

Arousal = []
Valence = []
Dominance = []

values = AllRatings['Arousal']
for i in range(len(values)):
    if values[i] < 2:
        Arousal.extend([0]) # uninterested/bored
    else:
        Arousal.extend([1]) # excited/alert

values = AllRatings['Valence']
for i in range(len(values)):
    if values[i] <= 4:
        Valence.extend([0]) # unpleasant/stressed
    else:
        Valence.extend([1]) # happy/elated

values = AllRatings['Dominance']
for i in range(len(values)):
    if values[i]> 4:
        Dominance.extend([0]) # helpless/Without Control
    else:
        Dominance.extend([1]) # Empowered

FinalFeature['Arousal'] = Arousal
FinalFeature['Valence'] = Valence
FinalFeature['Dominance'] = Dominance
```

Code: Class Distribution of self assessment Ratings

A.3 Classification Codes

```
#Support Vector Machine
model = svm.SVC(kernel='rbf', random_state = 0)
data = dx

for Class_name in ['Arousal','Valence','Dominance']:
    y = data[Class_name]
    X = data.drop(['Arousal','Valence','Dominance'], axis=1)
    X = preprocessing.minmax_scale(X)

    # (Chi Square) It does not Increase the accuracy , rather decreases
    #X = SelectKBest(chi2, k=10).fit_transform(X, y)
    #X = pd.DataFrame(X)

    #Cross Validation Test Score
    cv_score = cross_val_score(model,X,y,cv = 10, scoring='accuracy')



    SVMresult.append(cv_score.min()*100)
    SVMresult.append(cv_score.mean()*100)
    SVMresult.append(cv_score.max()*100)

    print("SVM "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.min()*100), "(Minimum)")
    print("SVM "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.mean()*100), "(Mean)")
    print("SVM "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.max()*100), "(Maximum)")

    from sklearn.metrics import confusion_matrix
    #print(classification_report(y_test, y_predicted))

    print('------------------------------------------------')
```

```
SVM  CV Test Score of Arousal: 92.857 (Minimum)
SVM  CV Test Score of Arousal: 94.930 (Mean)
SVM  CV Test Score of Arousal: 95.238 (Maximum)
------------------------------------------------
SVM  CV Test Score of Valence: 85.366 (Minimum)
SVM  CV Test Score of Valence: 86.725 (Mean)
SVM  CV Test Score of Valence: 87.805 (Maximum)
------------------------------------------------
SVM  CV Test Score of Dominance: 82.927 (Minimum)
SVM  CV Test Score of Dominance: 84.309 (Mean)
SVM  CV Test Score of Dominance: 85.366 (Maximum)
------------------------------------------------
```

Code: SVM Classification

```python
#Extreme Gradiant Boosing Model
from sklearn.feature_selection import SelectKBest, chi2
model = XGBClassifier(max_depth=3,
                      objective='binary:logistic',
                      n_estimators= 300,
                      learning_rate = 0.05,
                      early_stopping_rounds= 10,
                      random_state=0)
data = dx

for Class_name in ['Arousal','Valence','Dominance']:
    y = data[Class_name]
    X = data.drop(['Arousal','Valence','Dominance'], axis=1)
    X = preprocessing.minmax_scale(X)

    X = RFECV(estimator=model, step=1, cv=10, scoring='accuracy').fit_transform(X, y)
    X = pd.DataFrame(X)


    #Cross Validation Test Score
    cv_score = cross_val_score(model,X,y,cv = 10, scoring='accuracy')

    XGBresult.append(cv_score.min()*100)
    XGBresult.append(cv_score.mean()*100)
    XGBresult.append(cv_score.max()*100)

    print("XGB "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.min()*100), "(Minimum)")
    print("XGB "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.mean()*100), "(Mean)" )
    print("XGB "+" CV Test Score of "+Class_name+':',  "%.3f" % (cv_score.max()*100), "(Maximum)")

    print('-----------------------------------------------')
```

```
XGB  CV Test Score of Arousal: 92.857 (Minimum)
XGB  CV Test Score of Arousal: 95.174 (Mean)
XGB  CV Test Score of Arousal: 97.561 (Maximum)
-----------------------------------------------
XGB  CV Test Score of Valence: 83.333 (Minimum)
XGB  CV Test Score of Valence: 87.456 (Mean)
XGB  CV Test Score of Valence: 90.244 (Maximum)
-----------------------------------------------
XGB  CV Test Score of Dominance: 82.927 (Minimum)
XGB  CV Test Score of Dominance: 84.541 (Mean)
XGB  CV Test Score of Dominance: 85.714 (Maximum)
-----------------------------------------------
```

Code: XGBoost Classification

A.4 Comments From the Panel Members

Dr Mohammad Zavid Parvez

- Impressive Results. Good luck for future.
- Have you observed what other researches did? Answered: Yes, we have applied many techniques by gathering knowledge from many researches by focusing on their drawbacks and benefits that can help us.
- Dataset source? Answered: it was from a secondary source and our supervisor helped us in working with this dataset.
- Only Show your work in the Journal version not much definitions. Best of luck.

Dr. Abu Mohammad Khan

- Asked Questions such as, what is Linear kernal in general, why choosen FFT? Answered Those.
- No other specific comments.

A. M. Esfar-E-Alam

- Everything was clear. No comments.

Ms. Mehnaz Seraj

- Why Choosen XGBoost? So, Answered with details and comparison why other approaches are not providing better results and XGBoost it is better.