

Recommendation System for Mood Stabilization Using Content Recommendation

by

Kazi Md. Al-Wakil
23341073

Rifai Rahman
19201013

Nafisa Nawal
20101353

Sababa Rahman Meem
23341074

Sajid Rashid
20101163

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
School of Data and Sciences
Brac University
September 2023

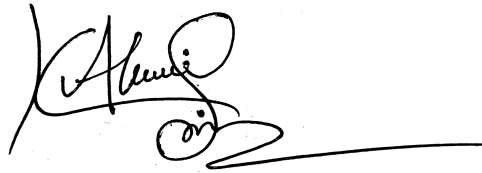
© 2023. Brac University
All rights reserved.

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:



Kazi Md. Al-Wakil
23341073



Rifai Rahman
19201013



Nafisa Nawal
20101353



Sababa Rahman Meem
23341074



Sajid Rashid
20101163

Approval

The thesis titled “Recommendation System for Mood Stabilization Using Content Recommendation” submitted by

1. Kazi Md. Al-Wakil(23341073)
2. Rifai Rahman(19201013)
3. Nafisa Nawal(20101353)
4. Sababa Rahman Meem(23341074)
5. Sajid Rashid(20101163)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on September 21, 2023.

Examining Committee:

Supervisor:
(Member)



Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Co-Supervisor:
(Member)



Mr. Rafeed Rahman
Lecturer
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Program Coordinator: (Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Head of Department: (Chairperson)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Ethics Statement

This research paper is done and contributed by the all the group members and it's plagiarism free.

Abstract

In the era of accelerating technological development, society is confronted with the paradoxical situation of making technological advancements while experiencing a decline in mental health. The importance of mental health seems to be declining significantly. The impact of our daily content intake on emotional well-being is clearly visible. For instance, while a melancholic song can make a person feel sad, an inspirational movie can charge a person's spirit to come up stronger. Hence we intend to employ this concept to propose a system designed to recommend "Feel Good" YouTube videos with the aim of stabilizing an individual's mood when it wavers or becomes low. To do this efficiently, we worked on the SEED Dataset, which is composed of EEG signals and Eye Movement data. We implemented a multifaceted approach, including the extraction of Differential Entropy Features, Wavelet Transform, Shannon Entropy features and Eye movement features. These were further harnessed by Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks to ensure accurate emotion classification. A thorough evaluation of these two deep learning models in the context of emotion classification is presented by focusing on their relevant merits and demerits. Based on the comparisons it is found that CNN is the most suited for our study with an accuracy of 93.01%. Once a mood classification is achieved, our proposed system will curate and suggest trending "Feel Good" content. To tackle this, we implemented a recommendation system based on the fusion of two prevalent techniques. Initially, text classification was employed to extract the emotion associated with the video and later, Pearson Correlation was utilized to obtain accurate correlation between the contents of the videos based on their corresponding ratings from viewers. Furthermore, concepts of Analytic Hierarchy Process (AHP) have been implemented to come up with an efficient algorithm which works in stabilizing an individual's mood gradually. In essence, our innovative system encompasses two primary objectives: the detection of an individual's emotional state through EEG signal analysis and the subsequent stabilization of their mood through targeted content recommendations. By combining these components, we envision a tool that not only comprehends the user's emotional well-being but actively contributes to its enhancement.

Keywords: Emotion Classification, Electroencephalograms (EEGs), Content Recommendation, Mood Stabilization, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Analytic Hierarchy Process (AHP), Text Classification, Pearson Correlation

Dedication

This research is dedicated to all the people who are suffering from depression and feel that their life do not hold any meaning.

Acknowledgement

First of all, all the praise to the Great Allah for whom we could completed our thesis without any major setbacks. Moreover, we want to show our gratitude to our Co-supervisor Mr. Rafeed Rahman sir and respected Supervisor Dr. Md. Golam Rabiul Alam sir. Their guidance and feedbacks were the core of our thesis. Lastly, we are also grateful to our friends, teachers, mentors for their time and support.

Table of Contents

Declaration	i
Approval	ii
Ethics Statement	iv
Abstract	v
Dedication	vi
Acknowledgment	vii
Table of Contents	viii
List of Figures	x
List of Tables	xi
Nomenclature	xiv
1 Introduction	1
1.1 Overview	1
1.2 Problem Statement	3
1.3 Research Contribution	6
2 Literature Review	7
2.1 Background Information	7
2.1.1 Emotion	7
2.1.2 Brain Activity Measurement	7
2.1.3 Electroencephalography (EEG)	8
2.1.4 Recommendation System	11
2.1.5 Content Based Filtering	12
2.1.6 Collaborative Filtering	12
2.1.7 Hybrid Filtering	12
2.2 Related Works	13
2.2.1 Emotion Classification	13
2.2.2 Recommendation System	18

3	Methodology	20
3.1	Overview of the Proposed System	20
3.2	Dataset	22
3.2.1	EEG Data collection	22
3.3	Feature Extraction	24
3.3.1	Differential Entropy (DE) Features	24
3.3.2	Wavelet Energy & Shannon Entropy Features	24
3.3.3	Eye Movement Features	25
3.4	Exploratory Data Analysis (EDA)	26
3.5	Data Pre-processing	33
3.6	Model Specification for Emotion Classification	34
3.6.1	Convolutional neural network (CNN)	35
3.6.2	Long Short-Term Memory (LSTM)	38
3.7	Recommendation System	42
3.7.1	Collaborative Filtering	43
3.7.2	Pearson Correlation	44
3.7.3	Emotion score extraction of Videos	46
3.7.4	Analytic Hierarchy Process (AHP)	47
4	Result analysis	53
4.1	Performance Evaluation Metrics	53
4.2	Experimental Result Analysis	55
5	Conclusion	61
5.1	Challenges	62
5.2	Limitations	62
5.3	Future Work	62
	Bibliography	66

List of Figures

2.1	Valence-Arousal model	8
2.2	Different Parts of brain	8
2.3	Electrode locations of International 10-20 system for EEG recording .	10
2.4	Notations of Placed Electroeds	10
2.5	Content-based Filtering System	12
2.6	Hybrid Filtering	13
3.1	Overview of the Proposed System	21
3.2	Break down of time taken for showing one stimuli/clip	22
3.3	Data Collection from Participants	23
3.4	Early Fusion Technique	26
3.5	Raw EEG Data of one participant from one Session	27
3.6	Emotion graphs from Raw Data	29
3.7	DE Features: How each emotion differs from each other	31
3.8	Eye Movement Features: How each emotion differs from each other .	32
3.9	Label counts for each emotion	33
3.10	Simple CNN Architecture	35
3.11	How Rectified Linear Unit(ReLU) works	36
3.12	Proposed Convolutional Neural Network (CNN) Architecture	37
3.13	RNN vs LSTM	39
3.14	Inside Architecture of LSTM	39
3.15	Proposed Long short-term memory (LSTM) Architecture	41
3.16	Collaborative Filtering: Example 1	43
3.17	Collaborative Filtering: Example 2	44
3.18	Pearson Correlation	45
3.19	Six emotions classified into Positive and Negative emotions	46
3.20	Multi Criteria Decision Making	48
4.1	Comparisons between CNN and LSTM of the score of performance evaluation metrics	55
4.2	Comparisons between CNN and LSTM of Precision, Recall and F1- Score of each label	57
4.3	CNN Accuracy and Loss Curve	58
4.4	LSTM Accuracy and Loss Curve	58
4.5	Confusion Matrix	59
4.6	The order of recommendation	60

List of Tables

2.1	How Brain Activity, Mental state related to Sub-bands and the part of the brain they operate	11
3.1	Extracted Features from Eye-Movement Data	25
3.2	Raw Data Information of one participant from one Session	26
3.3	Data distribution of each emotion of Raw EEG Data	29
3.4	Data distribution of each emotion of DE features	31
3.5	Data distribution of each emotion of Eye Movement Features	32
3.6	Label counts for each emotion of SEED-V dataset	33
3.7	Categorical encoded values of emotions	33
3.8	Proposed Convolutional Neural Network Summary	38
3.9	Proposed Long short-term memory (LSTM) Summary	42
3.10	The 1-9 Fundamental scale	48
3.11	Pairwise comparison matrix of the criteria	49
3.12	For $N = 10$, Random Inconsistency Indices (RI)	51
4.1	Performance evaluation metrics result of CNN & LSTM	55
4.2	CNN: Evaluation Metrics values for each emotion label	56
4.3	LSTM: Evaluation Metrics values for each emotion label	56

List of Algorithms

1	Extraction of Pearson Correlation Scores	45
2	Extraction of emotion scores of videos through subtitles	47
3	Analytic Hierarchy Process (AHP)	50
4	Ranking of the videos and Recommendation	52

Nomenclature

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

- AHP* Analytic Hierarchy Process
- ASP* Asymmetric Spatial Pattern
- BCI* Brain-Human interface
- CBF* Content-Based Filtering
- CF* Collaborative Filtering
- CI* consistency index
- CNN* Convolutional Neural Network
- CNS* Central Nervous System
- CR* Consistency Ratio
- CWT* Continuous wavelet transform
- DDE* Dynamic Differential Entropy
- DE* Differential Entropy
- DT* Decision Tree
- DWT* Discrete wavelet transform
- EDA* Exploratory Data Analysis
- EEG* Electroencephalogram
- EFDMs* Electrode-Frequency Distribution Maps
- EMG* Electromyogram
- EOG* Electrooculogram
- eSM* Enhanced Sentiment Metric
- FCM* Fuzzy C-Means
- FCNN* Fully Connected Neural Network

FFNN Feed-Forward Neural Network
GSR Galvanic skin response
IG Information Gain
IMF intrinsic mode function
KNN K-Nearest Neighbor
LR Logistic Regression
LSTM Long Short-Term Memory
MCDM Multi-Criteria Decision-Making
MLP Multi-layer Perceptron
mV Micro-volts
NB Naive Bayes classifier
OGPCP oriented Gabor phase congruency pattern
PPG Photoplethysmography
PSD Power Spectral Density
RBF Radial Basis Functions
ReLU Rectified Linear Unit
RF Random Forrest
RI Random Index
RS Recommender System
RWE Relative wavelet transform
SAM Self-Assessment Manikin
SEED SJTU Emotion EEG Dataset
SOM Self Organizing Map
STFT Short-Time Fourier Transform
SVM Support vector machine
WT Wavelet transform

Chapter 1

Introduction

1.1 Overview

It is not feasible to describe emotion in one word. Kleinginna and Kleinginna researched on how to define emotion from a literary point of view. They were able to find 92 definitions of emotion in literature till now. Based on the 92 definitions, they concluded that Emotion is a complex system driven by different subjective and objective factors and conciliated by hormonal systems.[1]

In addition, it is an era of digital content. An easy access to all sorts of digital content makes our life more entertaining. Facebook, Instagram, Youtube, Reddit, Twitter etc. are some of the big giants which are providing us with various kinds of content. Despite having so many ways to be entertained, we are not happy because it is also an era of depression. People easily reach a state where he/she finds his/her life meaningless. A lot of variables affect our daily moods and that also affects our daily work flow along with our mental health and everyday progress. The impact is clearly visible in communication, productivity etc.. For example, sometimes we get angry even in silly matters, in which we would not have reacted if our mood were stable. Nowadays, a large part of communication is now done by using electronic devices and its components. Therefore, we want to make a recommendation system which can predict that a person is having a bad day and by detecting that it will recommend some good content to lighten up his/her mood.

According to Picard and Klein, understanding emotions through computers and its ability to pay attention to how humans interact with each other has a great future ahead. [6] The only barrier computers had was that they could not understand human emotions and their needs. The need of detecting a person's emotional well-being through technology is increasing day by day. [6] There has been much research on how a computer should classify emotions. First instinct was to train our computers to identify emotions through facial expressions and voice because humans understand emotions through these 2 mediums mostly. Computers can classify emotions pretty successfully (80-90%) by implementing image processing of facial expression. [10]

Initially emotions can be separated into three parts. Physiological arousal, Expression which is caused by something and the feeling, experience of an emotion. [13] In order to deeply understand emotions, more studies have been conducted on heart rate, skin conductance, dilation of pupils etc.[5], [10] However, this is the time for having brain-human interface (BCI) where interaction with computers happens using brain activity. [13] Everything a person does, it has its origin inside of the brain. Certain changes can be seen in the signals of the brain. Those brain signals contain a lot of information about an action, emotion etc. and by using this information we can extract almost anything about a person. At present, through BCI, we can control almost all the programs of today's world. [20] In order to implement BCI, Electroencephalograms (EEGs) signals are the most suited. These are one of the fast growing brain signals which are being used in many research works. At present, EEG-based human's emotion classifiers, which have been tested on artificial emotions, have a success rate of 60% but it has been proven that to understand human emotions, EEG is more suited.[4], [14]

The conventional method for extracting human emotion is by testing audio and visual data to have an emotional model of a human. This can also be done by examining speech, body movements, gestures, facial expressions. [26] However, compared to the conventional methods, biosignals such as EEG give more accurate and detailed information on human emotions.[21]

EEG data will be processed in such a way so that we can extract features that are most suited. Afterwards, extracted features will be combined with Eye Movement features to create a new angle to the classification. After proper pre-processing, using the deep learning algorithm Long-Short Term Memory (LSTM) and Convolutional neural network (CNN) the data will be trained and tested. Moreover, a detailed comparison on evaluation metrics between the two deep learning models can also be observed in this study.

After successfully detecting a person's mood, now it's time to lighten their mood up. There is a vast ocean of contents out there and from those contents, only those contents will be picked which can stabilize a person's mood and a separate dataset will be created. Afterwards, item-based Collaborative Filtering recommendation system will be built by using Pearson Correlation method and extracting emotion score by classifying the subtitles of the Youtube videos which is targeted to help people with depression or bad mood by recommending them personalized content and maintain their mental as well as physical conditions which will partially divert them and ease up their mind gradually.

1.2 Problem Statement

Emotion can be a tough thing to define. Sometimes it is even hard for humans to understand what the other person is feeling. It's even harder to make it recognizable for machines, computers to learn the way humans feel. Throughout the years, many researchers have tried to find a solution on how to feed our machines enough information so that our machines can successfully identify a person's mood, emotion. Because out of all things, detecting emotion is the crucial part. After mood detection, we can solve many co-related problems.

The problem we came up with is that people get bored with life easily. Life seems meaningless. Everyday people get up from bed, trying to find the meaning of life. If they could not find any meaning, they feel like their life is meaningless hence do extreme things. Our system's purpose is to build a system where the identification of a mentally unstable or depressed person will be automated. In order to ease up the mood, our system will also recommend some contents to stabilize the mood of the person. The contents will be selected through artificial intelligence technology. However the real problem is, how are we going to identify a depressed person. There are many conventional methods. Such as, facial recognition, the tone of the voice while communicating, action, gestures, Physiological signals such as heart rate, skin conductance, galvanic skin resistance, Brain Signals etc.

Emotion detection is nothing new. Researchers have been trying to find a way for years so that a machine can be as close as possible to humans. Therefore, many research works have been done to detect emotion through machines. The most conventional methods to extract emotion from humans were visual based or audio based. As for humans, we identify a person's mood based on his/her facial expressions and all other actions. Therefore, it was a common instinct for the researchers to work with visual or audio dataset. Furthermore, feature extractions in the conventional methods are not that complex if compared to neural signal based methods.

Emotions can be classified using facial expressions, image data, sound samples, voice frequency or gestures. However, we can not necessarily put this practice into our real life and implement it because a number of people can not talk or see, some are physically handicapped as well. [49]

Ekman et al.[2] did research on human emotions and their relation with facial expression. They concluded in their study that all facial expressions have 7 main categories. These categories are the same for the whole world regardless of class, country, caste, age. There are no barricades. But the concerning fact is that these facial expressions can be faked, therefore using facial expressions as test data and concluding an emotion might not be a good idea, the research will not be reliable. The same thing is true for audio data. Humans can mimic any type of pitch of tone regardless of the actual emotion. In order to overcome the situation, Ayata et al.[39] conducted research on Physiological signals. The data were collected by a wearable device which is basically a physiological sensor that works with galvanic skin response (GSR) and photoplethysmography (PPG). However, the results were good but not satisfactory. They obtained 71.53% (arousal) and 71.04% (valence)

for GSR and for PPG the score was 70.92% (arousal) and 70.76% (valence).

Seanglidet et al. [35] conducted a research on patients who have mood disorders, especially elderly persons and monitor them. Their goal was to enhance their patients mood by music therapy. To detect a patient's mood they chose to use facial expression. After careful processing and model implementation, they got an accuracy of about 60%. The reason for low accuracy is mainly because it is difficult to identify the disgust mood. Moreover, their android app version has limited computing resources. Thus, it cannot give predictions for every frame. Thus, thirteen face-distance-ratios are computed for every X frames. Furthermore, the Active Shape Models (ASM) cannot detect the facial feature points properly when the user's face is in motion or the user is showing his/her side view only to the camera. The accuracy also decreases when the face orientation tilts over 20 degrees because the ASMs cannot fit to the new tilted image.

Brain is the core part of our body. Neurons inside the brain contain detailed information about a person. It can also define different emotional states of a person. Analyzing brain signals can be a challenging task because brain signals are dynamic. Each action we do, the thinking, feelings, experience all have different sorts of outcome in signals and they are all unique. The point is, neurons generate different signals in response to an action we perform. Therefore, it is necessary to implement the right algorithms to extract features from raw data, to select optimal features and to classify the states of emotions and achieve higher efficiency.[48] Moreover, while extracting the features from the EEG signal, a few problems can be arised because in raw state the Signals are multi-dimensional. The noise level is also high in EEG signals, to process the signals, extracting features from it can be a hard task to do. Huang et al.[29] tried to keep all the limitations in mind and came up with a new algorithm to extract features called Asymmetric Spatial Pattern (ASP) which extracts spatial filters. Even though the algorithm performed better than other conventional algorithms. However the error rate was still high and that can be minimized.

Chanel et al.[18] conducted a research where they collected user data manually by asking the user to remember any past event of which they are emotionally connected. They split their data into 2 parts. One part was of 3 categories, another part was of 2 categories. They achieved 79% and 76% respectively in those categories. In another research of Chanel et al. [22] used self-learning techniques and extracted features from EEG signals for three classes which resulted in 63%. Furthermore, Pun et al. [16] were able to achieve 72% accuracy for 2 classes and even less accuracy for 3 classes, which is 58%. All these researches came out with results which are good but not sufficient. The results are varied because all those researches have their own purpose and other selection criterias. There are different factors which varied from each of these studies such as the number of subjects participated in the test, the different types of stimuli such as emotional photos, songs, past events, happy memories, funny contents etc., classifier selection, different ways to extract features and many more.

According to a survey conducted on EEG data, for which 41 articles were studied, it has been found that out of those 41 articles only 15% research has been conducted on self-collected data. The other portion of the percentage, has conducted research on an already available dataset that has been open to all to use. Among the 85% of the research, only 7% research has been conducted on the SEED dataset. [49] Therefore, it is required to work more on the SEED dataset so that the true potential or maximum optimality from this dataset can be achieved.

Moreover, most of the work that has been done on the sector of EEG signal, has only worked with 1 or 2 features. For example, Dynamic Differential Entropy (DDE), Electrode-Frequency Distribution Maps(EFDMs), Hjorth Parameter, Power Spectral Density(PSD), etc. However, none of the scholars have tried to classify the emotion based on 3-4 features. [49] Last of all, for a long time researchers have tried to identify what causes a person's mood disability and thus it has become a trend to classify emotions. This area of research is growing rapidly. [49]

Moving on to the recommendation part, Liu et al. [24] used heartbeat to detect the mood and taking the mood into consideration made a music recommendation system. Similarly, Yoon et al.[30] made a music recommendation system which is personalized by using selected features, history of listening and information about context. Rosa et al.[32] made a music recommendation system named Enhanced Sentiment Metric (eSM) which has a correlation with the user's profile and based on lexicon sentiment metric. In order to classify the user's sentiment, authors collected data from user's social media posts. Analyzing the social media posts, personalized songs were suggested to the users. There is one drawback here, identifying a user's sentiment through social networks might not be possible for all cases. There are millions of people using social media and not all of them are active on them. Hence, an inactive person's mood might be unidentified all the time. Furthermore, there are traditional music players which recommend music to users based on their mood. For example, if a person is sad, the system will recommend more sad songs, which will eventually make the situation worse. There are not many mood enhancing recommendation systems available in the market. Moreover, the few mood enhancing recommendation systems are recommending songs only. But a person's mood can be lightened up by various kinds of content such as a cute video of a cat, an inspirational movie, a short video on YouTube etc.

Researchers are either doing emotion classification or building recommendation systems. The idea of classification along with a recommendation system has been addressed by very few scholars. Therefore, the question that has been raised and this research is trying to answer is:

How to process EEG signals optimally so that useful features can be extracted from raw data and which deep learning classification algorithm is more suitable? How to provide mood stabilizing content after classifying a person with a sad mood? What are the best techniques so that the best videos are recommended to the user which will improve the mood gradually?

This study will try to provide an answer to the above question.

1.3 Research Contribution

In this research, we are trying to classify human emotions namely Happy, Sad, Fear, Disgust, Neutral, using EEG Signal and recommend mood enhancing contents from our dataset, stabilizing the mood slowly. In particular, we have tried to extract features from raw EEG Signal and combine it with Eye movement features, a completely different sort of feature in order to add a different dimension to the classification. For classification, we have used Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM). The model that gives the highest accuracy will be chosen. For the recommendation part we are using item based collaborative filtering by using the Pearson correlation method along with text classification, extracting the scores of the subtitle of the videos and combining the scores with the Analytic Hierarchy Approach (AHP) .

The main contribution of this study is summarized below:

- We extracted 2 different features from raw EEG signals namely Shannon Entropy, Wavelet Energy features. These features were extracted using Wavelet Transform technique.
- We combined 4 different features namely Differential Entropy, Shannon Entropy, Wavelet Energy and Eye movement Features. After that, we classified the five emotions (Happy, Sad, Fear, Disgust, Neutral) using the Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM).
- We showed a comparison between Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) based on their Accuracy, F1-Score, Recall, Precision. We also showed why CNN works better with EEG signals in comparison to LSTM.
- We created our own datasets consisting of different sorts of Youtube videos. Most of the videos are rated by a survey we have conducted on google form. 10 people have participated and contributed their valuable time by giving a rating on a scale of 1 to 5, to a particular video from our dataset.
- We combined 2 different techniques, Pearson Correlation and emotion score after classification of a text in order to recommend the best videos to the user.
- We developed a system where after classifying a ‘Sad’ person, we will be recommending some of the highly rated Youtube videos from our own dataset in order to stabilize the mood of the person gradually.

In chapter 2, some of the previous works that have been done on our topic have been discussed as literature review. The workflow and the methodology behind our work have been provided in chapter 3. In chapter 4 we showed our findings after the experiment and analyzed it. Lastly, in chapter 5 we concluded our study by giving a brief summary of the research, acknowledging our limitations and discussing the scopes of future works.

Chapter 2

Literature Review

2.1 Background Information

2.1.1 Emotion

Emotion itself does not have any physical appearance or value. It is an abstract thing, a feeling that shows how we behave for specific instances. We act differently depending on the situation, we are driven by our emotions. As it is not palpable, we can not measure it accurately. According to various researches, there are many types of emotions. For example, satisfaction, nostalgia, fear, empathetic, disgust, confusion, envy, craving and many more. [49]

In order to classify emotions, many researchers have suggested many models. One of the most popular models among them is Russel's Circumplex 2D model. Emotional models can be further categorized into 2 parts. [49]

- 2D (Two-Dimensional) Model
- 3D (Three-Dimensional) Model

The Figure 2.1 below shows the representation of Valence Arousal Model Two-dimensional model contains values of Valence and Arousal. Where the amount of pleasure is indicated by Valence and the amount of excitement is indicated by Arousal. [49]

2.1.2 Brain Activity Measurement

We can tell a lot by analyzing brain signals because we humans are driven by our brains. Each and every work is regulated by our brain. We can compare our brain to the Storage Device of our computer. Only difference here is that the storage device has a limited amount of space but our brain has the capacity to store an infinite amount of information. There are several ways to measure brain activity. Such as, Functional Resonance Imaging (fMRI), Positron Emission Tomography (PET), Electroencephalography (EEG).[20]

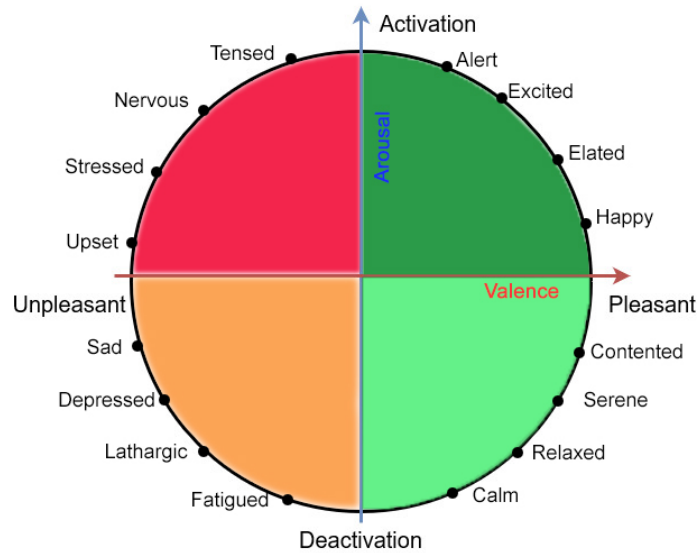


Figure 2.1: Valence-Arousal model

2.1.3 Electroencephalography (EEG)

Brain is one of the most crucial body parts of a human being. It is the core of the nervous system. It is also called the Central Nervous System (CNS). It is made up of 3 major parts namely Cerebrum, Cerebellum and Brainstem. Cerebrum takes up the most space of the brain and is divided into two parts, Right and Left hemispheres. The hemispheres are also divided into four different lobes. The names of these lobes are Frontal, Parietal, Temporal and Occipital. The figure Figure 2.2 shows a visual representation of these lobes and their location. [49]

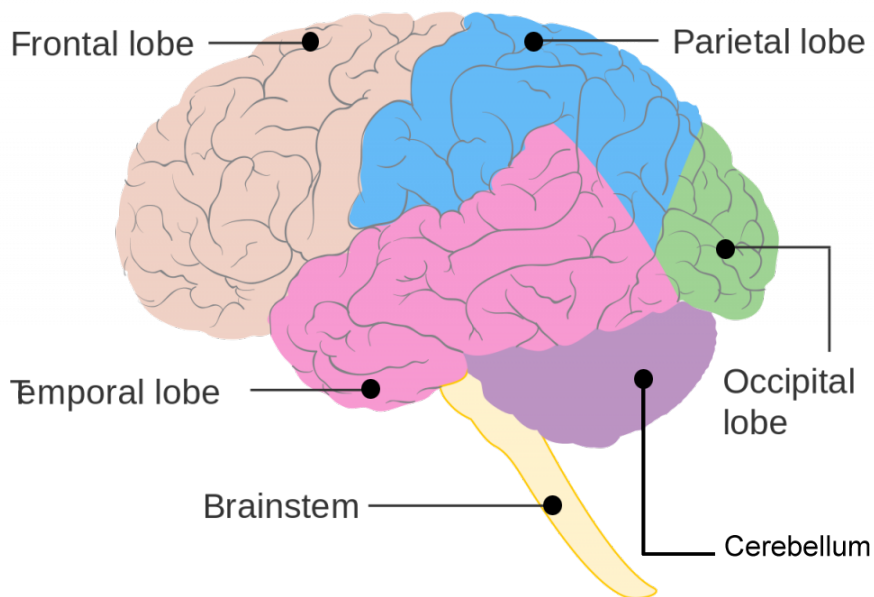


Figure 2.2: Different Parts of brain

The neurons of our brain generate electrical potential in an active state. A group of neurons' electrical activity is measured from outside of the skull through EEG. However, as there are many tissues involved in the brain, also there is the skull, connected electrodes can not detect the exact location of the signal. In order to measure the electrical activity electrodes are positioned on the scalp. Standard caps have 19 electrodes. Although the number of electrodes is increasing day by day in order to get more accurate brain signals. We can see pulsed, rhythmic signals are produced by the neurons. Based on the frequency the signals can be divided into 4 bands.[20]

1. Alpha Band
2. Theta Band
3. Beta Band
4. Delta Band

Moreover, there is another band called *Gamma Band*. Usually, frequency range from 30-100 hz falls into this band.

EEG signals are collected by a non-invasive and painless method. In order to record macroscopic electrical activity, electrodes are placed on the scalp. EEG recordings are the activity of the surface layer of the brain which is underneath the scalp. We get graphs as our output from the EEG machine representing electrical activity of the brain. Small sensors are attached to the scalp to pick up the electrical signals produced by the brain. In the cortical layer, neurons which are active underneath the scalp can be recorded using the EEG machine. Intensity of the signal is very small, usually measured in Micro-volts (mV). EEG machines do not detect the activity of a single neuron, rather detects the population level of neural activity. Electrodes are placed using the 10/20 rule. Based on the place of the electrodes, the names are given. There are 4 areas where electrodes can be placed.

1. Frontal
2. Temporal
3. Central
4. Parietal

All of the information regarding the EEG signals, their sub-bands, the way each sub-bands work and the part of the brain they are related to is summarized and shown in the table Table 2.1.

The Figure 2.3 & Figure 2.4 below shows how electrodes are placed

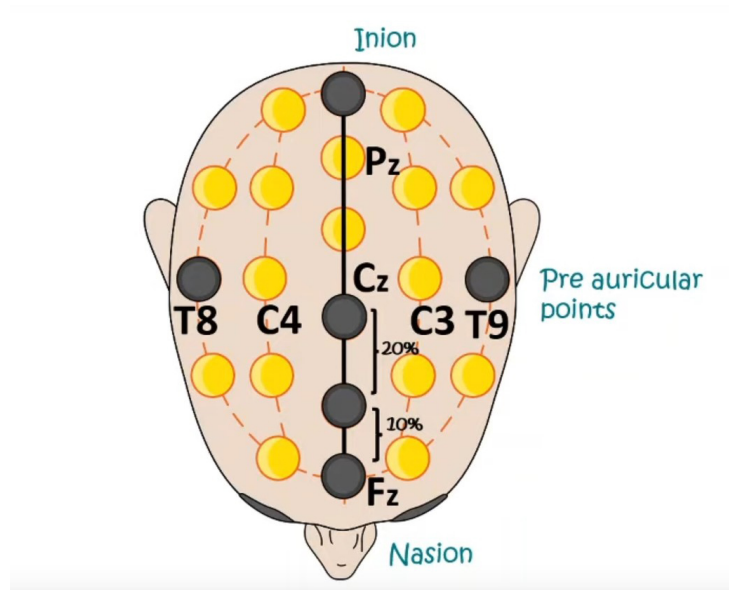


Figure 2.3: Electrode locations of International 10-20 system for EEG recording

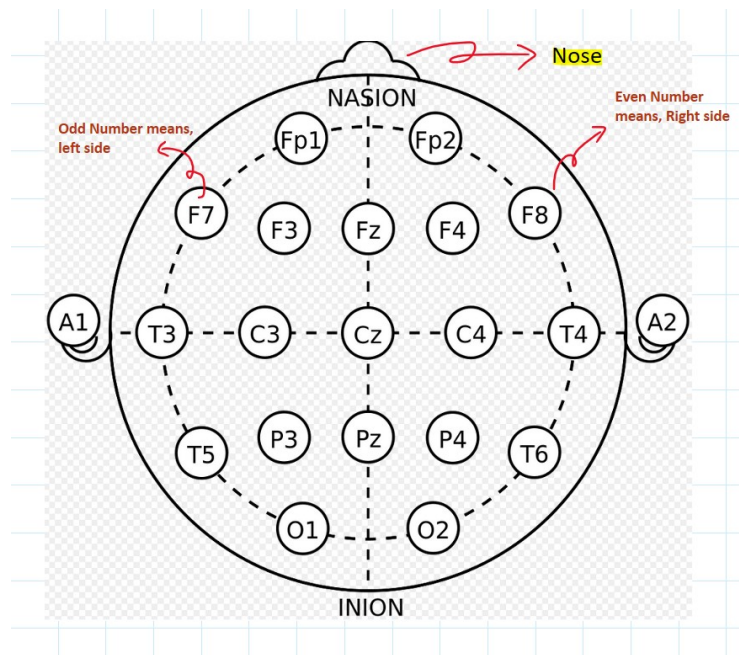


Figure 2.4: Notations of Placed Electrodes

Table 2.1: How Brain Activity, Mental state related to Sub-bands and the part of the brain they operate

Sub-Bands	Range of Frequency	Location on Brain	Mental state and activity
Delta wave	(0-4) Hz	Frontal	Deep Sleep, Continuous Attention (for babies), Unconscious
Theta wave	(4-12) Hz	Midline, Temporal	Drowsiness, Imaginary, Enthusiastic, Fantasy
Alpha wave	(8-12) Hz	Frontal, Occipital	Closing the eye, Relaxed, Calm
Beta wave	(12-30) Hz	Frontal, Distributed on sides symmetrically	Calm to intense to stressed, Aware of surroundings, Anxious, Thinking, Start to alert
Gamma wave	>30 Hz	Frontal, Central, Somatosensory Cortex	Cross-modal sensory processing, Alertness, Agitation, Short term memory for matching objects

2.1.4 Recommendation System

In this era of technology, the industries like content and product are growing rapidly. It has become a genuine problem to find the relevant products, contents or media that suits our preference. From the vast ocean of products, contents it is very necessary to build such systems where users can get their own preferable contents, products on the screen of their electronic devices. Recommender System works by taking information of the user, about the likings and dislikings then recommend accordingly. It is being used in every website, e-commerce site, content based websites like Facebook, Instagram, Youtube etc. [41]

The core of the recommendation system can be categorized into 3 parts based on how the contents, products are being recommended. These are:

- Content-based Filtering
- Collaborative filtering
- Hybrid Filtering

Among these Collaborative Filtering is the most popular and widely used. [41]

2.1.5 Content Based Filtering

Content Based Filtering works by recommending the product or content by seeing the previous record or rating of the user. The relationship between the recommendation and user is solely dependent on the user himself. The system creates a user's profile with the content type that the user has liked previously. This sort of filtering is widely used in Publications, News websites. The Figure 2.5 summarized how content-based filtering works.[41]

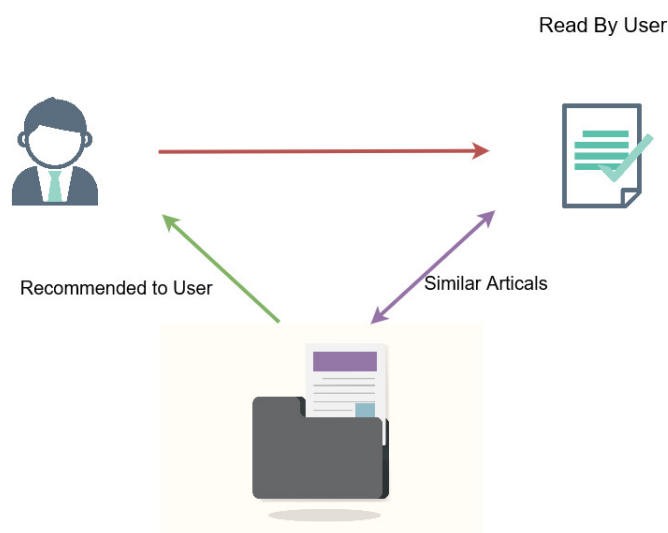


Figure 2.5: Content-based Filtering System

2.1.6 Collaborative Filtering

Collaborative filtering is a technique where users get recommendations of those which are liked by other similar users. In this type of recommendation system there will be a database where the products or contents are rated and then the system finds the similar users based on their profile and likings of the product or content. [41]

Collaborative filtering can perform prediction of rating of the user and can also recommend N number of contents which the user may like. This filtering technique can be categorized into two parts. Those are (i) Memory-based Filtering and (ii) Model-based filtering. Memory-based filtering measures the similarity between multiple users using Cosine Similarity or Pearson correlation or Jaccard Coefficient etc. [41]

2.1.7 Hybrid Filtering

As the name suggests, Hybrid filtering is a technique where multiple recommendation systems get ensemble together and the goal is to have a better performance ratio over traditional filtering methods. This process is categorized into 7 parts. [41]

1. Weighted
2. Switching
3. Mixed approach
4. Feature combination
5. Feature augmentation
6. Cascade
7. Meta-level

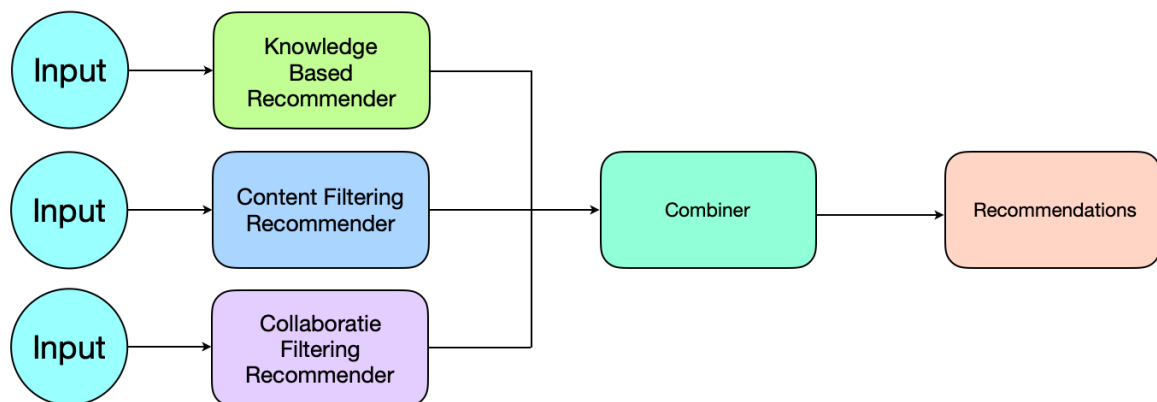


Figure 2.6: Hybrid Filtering

2.2 Related Works

2.2.1 Emotion Classification

Implementation of BCI can be a mouthful task if there are too many electrodes to connect and the connection with the computer is complex. In the research of Bos and his fellow researchers, a BraInquiry EEG PET(Personal Efficiency Trainer) device has been used to implement BCI. The device has 5 electrodes and the connection is also simple. The brain activity is measured using two channels. One channel for two dipole electrodes and a ground channel.[13]

Study shows that our emotions are caused by the amount of actions in our two frontal lobes (Left, Right). If the left frontal lobe is more active than we can say a person is happy. Similarly, if the right frontal lobe is more active then he/she is having a bad day.[9]

In the year 2000, Chopin classified six emotions with a success rate of 64% using EEG signals and with the help of neural networks. Those emotions were classified based on emotional valence and arousal.[4] This model is a bi-directional model

and has four emotional states which are: high arousal high valence (HAHV), high arousal low valence (HALV), low arousal high valence (LAHV), and low arousal low valence (LALV)[46]. Thus, emotional state is predicted based on valence-arousal model. The emotions can be represented as a 2D map, where valence is in the ‘Y’ axis and arousal will be in the ‘X’ axis. In order to label valence and arousal, Bos used the International Affective Picture System (IAPS) and International Affective Digitized Sounds (IADS) library to conduct his research.[3], [12]

Lin et al.[26] conducted a research where data were collected from 26 users. Among them 16 were male and 10 female. All aged around 24 years old. Each user’s EEG signal or data were collected during the phase of listening to some music. The researcher wanted to see the reaction in brain activity during different kinds of music and how music affects the brain signals. Parietal and frontal lobes are two lobes that provide detailed information which are related to human emotions thus easy to process the signals. That is why feature extraction was done from these two lobes for this research. More related works have been done later on, such as, EEG feature based work has been done by Ishino and Hagiwara [8] where they used neural networks in order to classify four emotional states. They got an average accuracy of 54% to 68%. On the other hand, Takahashi et al. [10] wanted to recognize emotions by using EEG, Skin Conductance and Pulse. Together they are called Multimodal Signals. Takahashi got an accuracy of almost 42% for five emotions by implementing Support Vector Machine (SVM). Furthermore, another research by Chanel et al. [14] conducted on three emotion classes achieved an accuracy of 58% by implementing Naive Bayes Classifier. Moreover, one of the best accuracy of 82.27% was achieved to identify eight emotional states. by Heraz et al.[19] by implementing the K-nearest neighbors classifier. Later on, Chanel et al. [22] conducted another research using SVM classifier and as features used EEG time-frequency information. 63% of average accuracy was recorded in this research. Ko et al. [23] worked with relative power changes of EEG signal and feasibility of it and assumed the emotional states of a user by using Bayesian networks. Lastly, Zhan and Lee [25] achieved 72.67% to 73.33% by implementing an SVM classifier which uses the frontal lobe’s unbalanced attributes as its features.

Bazgir et al. conducted a research focused on valence-arousal model. Data is taken from the DEAP database. Here, the EEG signals are labeled based on valence-arousal-dominance emotion model [40]. According to Coan et al. [7], positive and negative emotions are respectively associated with left and right frontal brain regions. The brain activity decreases more in the frontal region compared to any other regions of the brain. Hence, the channels chosen to investigate this study are: F-3, F-4, F-7, F-8, FC-1, FC-2, FC-5, FC-6, FP-1 and FP-2. Average mean reference (AVR) method is used to reduce the electronic amplifier, power line and external interference noise. The mean is calculated for each channel and subtracted from every single sample of that channel. The sample values are normalized between [0,1]. Using the mother wavelet function, the EEG signals are decomposed based on frequency called bands. They are Theta (4-8 Hz), Alpha (8-16 Hz), Beta (16-32 Hz), Gamma (32-64 Hz) and noises for anything above 64 Hz. Now, entropy and en-

ergy are computed from each window of every frequency band. Afterwards, PCA is applied to the extracted features to generate mutually uncorrelated features, known as principal components or PCs [27]. Bazgir et al. achieved an accuracy of 91.3% for arousal and slightly less percentage (91.1%) for valence using the SVM classifier with beta frequency band. [40]

Alhalaseh and Alasasfeh [48] processed their EEG signal data by using Intrinsic Mode Functions (IMF) or Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD). Generally, biomedical or disease related sectors mostly use these two methods. Later on, Higuchi's Fractal Dimension (HFD) and entropy technologies were used to extract features from the EEG Signal Data. Some of the known classifiers are used for the classification of emotion, such as Convolutional Neural Network (CNN), K-Nearest Neighbor (K-NN), Decision Tree (DT), Naive Bayes. For performance evaluation, DEAP dataset has been used and the proposed system achieved an accuracy of 95.20% while CNN model was implemented.

Moreover, Alhagry et al. [36] conducted their research on emotion recognition with a deep learning approach. They retrieved features from raw EEG signals by applying Long Short-Term Memory (LSTM). The features were broken into parts: high/low arousal, high/low valence and liking. To verify the system Alhagry et al. used DEAP dataset and achieved an accuracy of 86.65% for arousal, 85.45% for valence and 88% for liking class. Santamaria-Granados et al. [44] conducted a similar research but on AMIGOS dataset [42]. They used physiological signals such as electrocardiogram, galvanic skin response with a combination of machine learning approaches to extract signals in time, frequency and non-linear domain. After successfully analyzing the signals, they accomplished a great efficiency and precision in terms of classifying the states of emotions.

In addition, Mehmood et al. [37] did not use any available dataset. They used a sensor to collect EEG data from 21 healthy cases which were based on 14-channel. Data were collected while the user was stimulated with 4 different types of images triggering 4 different emotions, which are happy, sad, calm, scared. By using a statistical approach features were extracted and varieties of classifiers such as Naive Bayes, SVM, K-NN, Linear discriminant Analysis, Random Forest, deep learning and many more methods were used in order to identify the efficient algorithm. The end result was positive in terms of classifying emotions. Furthermore, Al-Nafjan et al. [38] used the DEAP dataset and implemented deep neural networks (DNN) for the emotion classification. Later on, it was found out that the method has similar approaches to State-of-art-emotion detection techniques, hence both were compared. In short, the study implemented DNN to a large dataset and classified different states of emotions and got good results.

It can be seen from the above discussion that most of the studies were based on DEAP dataset. Many classifiers have been used to conduct the research but among them CNN and K-NN, SVM were most commonly used. Moreover, the criteria and

purpose of each of the research was different and varied from one another. Lastly, after carefully analyzing the related works in this field, it can be seen that there is still a chance to improve the accuracy of emotion classification if available resources are properly utilized.

Zamanian, and Farsi, (2018), [45] conducted a research where they used the DEAP dataset to implement their algorithm in order to detect emotion. The Deap dataset consists of records of 32 participants ranging from the age of 19 to 32. These people were shown 40 different kinds of musical videos while their EEG signals were recorded with the help of a 32 channel BioSemi acquisition system. The amount of data generated here was huge as 32 channels were used. It takes 63 seconds with a frequency of 128 Hz to work with the data. Thus, it takes a lot of memory to store the data, and a long time to process it too. To overcome the problem, the researchers decided to work only with the data that was generated in the first 7.5 seconds. The algorithm that they used could successfully extract the required features from the less amount of data that they got in that time frame.

Furthermore, they limited the number of channels from which data was generated in order to reduce the size of the data which in turn would speed up the process. Two groups of channels were compared. One group consisted of channels P7, P3 and PZ. The other group consisted of channels P7, P3, PZ, PO3, O1, CP2 and C4. The purpose of this modification was to speed up the process and use fewer electrodes so that the user could feel comfortable while detecting their mood in real time. Thus, the algorithm was developed in such a way that could be user friendly besides having a high accuracy in the produced output. [45]

The researchers used Gabor wavelength features and intrinsic mode functions features in their research. For Gabor wavelength features, they performed convolution on Gabor filters with a 2D matrix. The matrix was developed by the data of each of the channels that they selected in each row. After the convolution was performed, they extracted three features which were energy, mean amplitude and oriented Gabor phase congruency pattern (OGPCP). For intrinsic mode functions features, they took the empirical mode decomposition approach to create the intrinsic mode functions (IMFs) first. Then they extracted 5 features from the IMFs which were maximum frequency, central frequency, entropy, root mean square, and variance. [45]

In this study, the researchers incorporated the genetic optimization algorithm with Support Vector Machine (SVM) and used it to find optimized hyper planes for the classification of the features. The researchers used Radial Basis Functions (RBF) as the kernel function to determine the parameters gamma and C. These parameters had a direct influence on the level of accuracy of the classification. Finally, they used the 10-fold cross validation technique on the data to train and test the algorithm. The results showed that they got the highest accuracy of 93.86% when they used only the three channels which were P7, P3, PZ and the Gabor filter with four scales

and six orientations (4x6). The entire model was designed based on the valence arousal model and four emotions were considered which were happiness, sadness, excitement and hatred. [45]

Wang et al. [28] conducted a research to detect human emotions using EEG signals where they classified the emotions into four categories- joy, relax, sad and fear. They used several movie clips of different categories to test the emotions of the participants. After watching each movie clip, the participants were asked to fill-up a Self-Assessment Manikin (SAM) form where they rated valence, arousal, and the specific emotion that they felt while watching the clip. This information was later used to verify EEG-based emotion classification.

They used a 128-channel electrical signal imaging system, SCAN 4.2 software, and a modified 64-channel QuickCap with embedded Ag/AgCl electrode to record EEG signals from 62 active scalp sites referenced to vertex (Cz) for the cap layout. On the center of the forehead they attached the ground electrodes. They kept a 16-bit quantization level at the sampling rate of 1000 Hz for recording the EEG data. At first, the researchers down-sampled the EEG signals to a sampling rate of 200 Hz. Then, they checked the time waves of the EEG data and removed the recordings which were contaminated by Electromyogram (EMG) and Electrooculogram (EOG). Moreover, the researchers divided each channel of the EEG data into 1000-point epochs with 400-point overlap. Finally, they considered each and every epoch of all the channels of the EEG data to calculate the features. [28]

For feature extraction, the researchers used the time domain to obtain statistical features and the frequency domain to obtain features based on the power spectrum. In the time domain, they extracted six features which were the mean of the raw signal, the standard deviation of the raw signal, the mean of the absolute values of the first differences of the raw signal, the mean of the absolute values of the second differences of the raw signal, the means of the absolute values of the first differences of the normalized signals, and means of the absolute values of the second differences of the normalized signals. In the frequency domain, they got five features which were delta rhythm, theta rhythm, alpha rhythm, beta rhythm, and gamma rhythm.[28]

Finally, the researchers used three different classification algorithms which were KNN, SVMs, and MLPs. In the case of KNN, they used the Euclidean distance method. In SVMs, they used a radial basis function kernel. In MLPs, they used a neural network which had three layers. The results showed that on average they got the highest accuracy of 66.51% when they used EEG frequency domain features and support vector machine classifiers. [28]

2.2.2 Recommendation System

Rodrigues et al.[34] have done some work on a framework which combines the user's geographic information and item-based collaborative filtering. This system was destined to solve the data sparsity issues and cold start. However, this recommendation system worked well only with new users, giving them a boost to experience the best contents that matches with the profile of the user. In addition, some other researchers have proposed a method based on clustering. They used coherent clusters which are semantic. Moreover, for recommendation they have extracted keywords from the contents which are available on the web. This is called Domain Ontology. [31]

Furthermore, Baoyao Zhou et al.[11] has tried to predict the next web page by using a sequential pattern mining method. They have used model-based collaborative filtering where they have their own database to store the web pages that have been accessed sequentially, from which the pattern to access the web of the user can be extracted.

Another group of scholars tried to make a recommendation system using the widely used Collaborative Filtering and combined it with the Fuzzy C-Means (FCM) algorithm. These two methods work very well in their respective sector. Collaborative filtering works better with rating prediction and recommendation, whereas FCM works well with clustering the items. In many cases FCM works better than K-means clustering. [41] Koochi et al. [33] worked with Similar type of method also. Where they will use collaborative filtering but for clustering they will use 3 algorithms, FCM, K-Means Clustering and Self Organizing Map (SOM) clustering. The idea was to show a comparison between these 3 clustering algorithms. In the study it has been seen that FCM works better than the other two clustering algorithms.

A scalable collaborative filtering algorithm was introduced by some scholars which was purely based on matrix factorization. Rather than relying on one User-item rating matrix, this system uses two matrices. These are called decision matrices, where one matrix represents User-keywords and the other matrix represents User-category. This algorithm was implemented using real-world data and also worked well with new items proving that Collaborative filtering method can be scalable. [41] In addition, one of the main problems of collaborative filtering was addressed by a group of researchers, which is data sparsity. Users and the number of items are gradually increasing every day. The graph of this growth is exponential. Therefore to solve this problem Dynamic Weighted Collaborative Filtering was introduced. This method works after the similarity between the user and item is found then it finds the similarity impact of the user and item by a method which controls the weights. It has been experimented and seen that the model works very well under various cases of data sparsity. [41]

In short, the problems that has been addressed by researchers are as follows:

1. Cold-start Problem: The problem occurs with new users and items. The new users failed to see contents that are suited for him/her. Same goes for contents. The new contents get ignored because it has not been rated or seen much by the users. [41]
2. Data sparsity problem: The problem occurs with the increasing number of users and contents. Most of the contents are not being rated by users creating the data-sparsity problem. [41]
3. Scalability Problem: The problem occurs with the increasing number of users and contents too. However, here the problem is at the core of the algorithm that has been used. The used algorithm may have worked well when the dataset was small but failed to give better results when it has to work with larger datasets. [41]
4. Privacy Issues: The problem is related to the privacy of the users. Most of the users are concerned about their privacy and might not want to rate a content which will be seen or recorded elsewhere. Therefore, most of the users do not rate, hence because of this privacy issue recommendation systems do not perform well. [41]
5. Synonyms: The problem occurs when similar contents or items have different or synonymous names. For example: “Action Movie” and “Action Film” are the same thing. The increased amount of synonymous words decreases the quality of the recommendation system as the system can not understand the depth of two similar words. [41]

Chapter 3

Methodology

3.1 Overview of the Proposed System

The proposed system is a combination of effective feature extraction from 2 different datasets, fusing the features to classify emotion with deep learning models Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) and a recommendation system to provide quality content to stabilize the mood. The whole system can be divided into 6 distinct phases. (1) Feature extraction from raw data (2) Feature Fusion (3) Exploratory Data Analysis (4) Data processing (5) Classification (6) Recommendation System. The overview of the system can be visualized from the Figure 3.1.

In our proposed system we have used two standard datasets (explained in section 3.2) in which we performed feature extraction techniques and got 4 different sets of features namely, (1) Differential Entropy Features (2) Wavelet Energy Features (3) Shannon Entropy Features (4) Eye Movement Features. After using the early fusion technique of these features we got the dataset for Emotion Classification.

After Exploratory Data Analysis (EDA) and pre-processing of the data, we applied the deep learning models CNN and LSTM, showing a detailed comparison of the two models and how one is better than the other. Then we passed the classified emotions through our recommendation system and enabled it to stabilize a person's mood by recommending good contents.

In this work, we tried to show through experimentation that the proposed CNN and LSTM model performs well with the features that we have extracted and also identified the better model for emotion classification by comparison. Moreover, the recommendation system has also proved to be an effective solution in order to stabilize the mood.

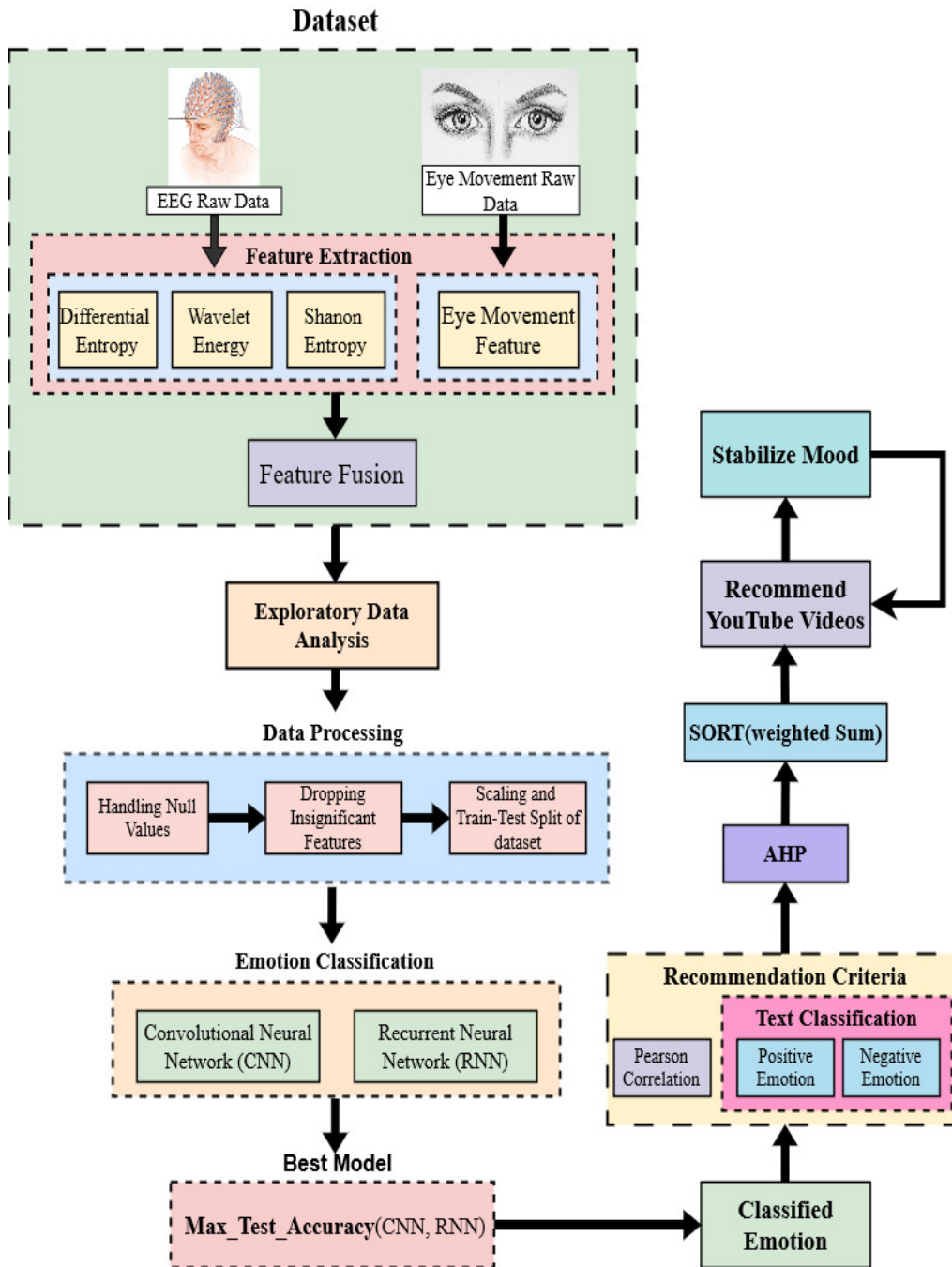


Figure 3.1: Overview of the Proposed System

3.2 Dataset

3.2.1 EEG Data collection

The recommendation system that we built is dependent on Emotion Classification. Hence, EEG and Eye movement based emotion classification has been conducted on SEED Dataset. Wei Liu et al. developed this dataset by experimenting with a few subjects in different sessions.[50] There are 5 types of SEED dataset namely (1) SEED-IV (2) SEED-VIG (3) SEED-V (4) SEED-GER (5) SEED-FRA. For our study, we worked with SEED-V dataset which has five states of emotions. They are (1) Happy (2) Sad (3) Disgust (4) Neutral (5) Fear

Wei Liu et al. tried to build the dataset by experimenting with 16 subjects. Among them 6 were males and 10 were females. Participants were all volunteers and students of Shanghai Jiao Tong University. Certain criteria were checked such as, usual vision and hearing capability, volunteers using their right hand as their dominant hand, stable state of mind etc.. Participants were picked through an online personality test posted on social media. The personality test is called the Eysenck EPQ personality test. [50]

Emotions were extracted or recorded by applying a stimulus material method. In simple words, participants were asked to watch specific video clips, generated emotional states were recorded corresponding to that clip. In order to have proper validation participants were asked to participate in the same experiment three times, three different days. There was at least a 3 day gap between the experiment days. Each participant was asked to watch 15 video clips as stimulus material on each experimenting day. Every video clip was unique to remove the boredom of the participant. Each session was approximately 50 minutes long. After each stimulus inducing materials were played, there was rest time, either 15 or 30 seconds for the participants, depending on the stimulus they were being shown. In the resting session, when a participant gets 15 to 30 seconds, they were asked about their feelings about the clip by rating the effectiveness of the clip on a scale of 0 to 5, where 5 being the most effective and 0 means no effect at all. The breakdown of time of showing one movie clip from starting to resting session is shown in the Figure 3.2 and the overall data collection process that is divided into 3 sessions can be visualized from Figure 3.3. Lastly, the 5 emotional states were numbered from 0 to 4 where 0 is 'Disgust', 1 is 'Fear', 2 is 'Sad', 3 is 'Neutral', 4 is 'Happy'.

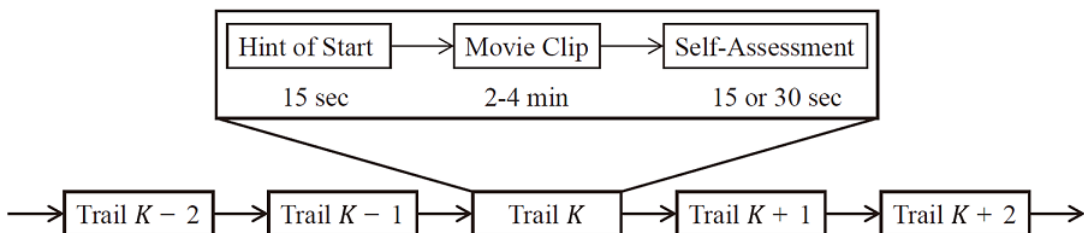


Figure 3.2: Break down of time taken for showing one stimuli/clip

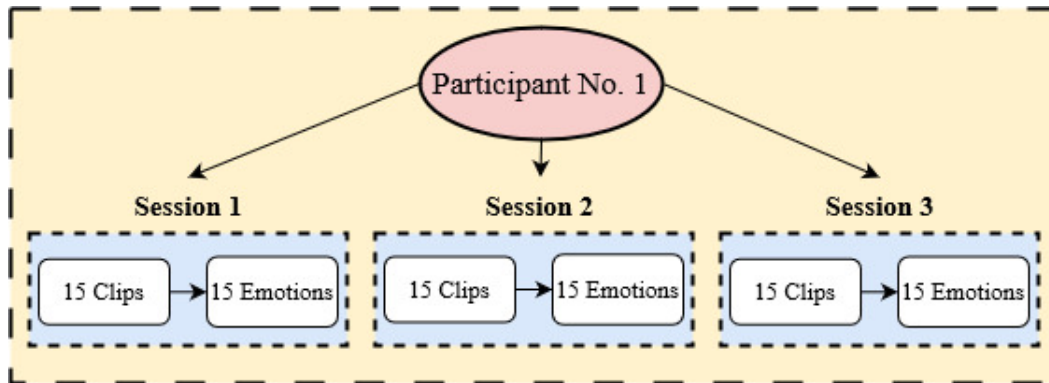


Figure 3.3: Data Collection from Participants

While the stimulus was being shown, a 62 channel ESI NeuroScan System was recording the EEG data and SMI eye-tracking glass was collecting Eye Movement Data. [50]. The raw data collected from the EEG machine was added with additional noise. The frequency range was far from workable, making the data noisy. Therefore, the raw data was brought under 200 Hz sampling rate to remove the noise. In order to be more precise, to remove more noise and artifacts, the data were passed through a bandpass filter of 1 Hz to 75 Hz.

In short, there were 16 participants, each had 3 sessions and in each session they were shown 15 movie clips as stimuli to generate certain emotions. A 62 channel ESI NeuroScan System was collecting all the EEG signals while they were being shown the movie clips. Therefore, from all the sessions of all the participants, there were 156 Million data points ready to be processed.

3.3 Feature Extraction

3.3.1 Differential Entropy (DE) Features

Differential entropy (DE) is a measurement technique which can calculate the complexity and entropy of continuous random variables. [47] In this study, Differential entropy (DE) features were extracted with a 4-second hanging window and without overlapping using Short-Term Fourier Transforms (STFT). Extracted features were within the segment of 5 frequency bands: (1) Delta : 1~4 Hz (2) Theta : 4~8 Hz (3) Alpha : 8~14 Hz (4) Beta : 14~31 Hz (5) Gamma : 31~50 Hz. [50]

The Differential Entropy equation by which features were extracted:

$$DE = - \int_{-\infty}^{\infty} P(x) \ln(P(x)) dx \quad (3.1)$$

If we assume that, Gaussian distribution $x \sim N(\mu, \sigma^2)$ is maintained by the EEG signals then the calculation can be simpler,

$$DE = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx = \frac{1}{2} \ln 2\pi e\sigma^2 \quad (3.2)$$

The device 'ESI NeuroScan System' responsible for capturing EEG signals has 62 channels and each channel has 5 bands, therefore, $(62 \times 5) = 310$ features were extracted from the raw EEG data.

3.3.2 Wavelet Energy & Shannon Entropy Features

We have used the Discrete wavelet transform (DWT) which gives us a non-redundant representation of the wavelet. We get it by using a simple recursive filter scheme, and we may also apply an inverse filtering operation in order to get the original signal back. The wavelet coefficients give us the full information in a simple way, and also provide us with an estimation of local energies at the different scales. The information is arranged in a hierarchical scheme of nested subspaces in $L^2(R)$ which is known as multiresolution analysis. [17]

In order to extract features from the raw data we have used Wavelet Filter Bank Technique where the signal is divided into 5 frequency sub-bands (alpha, beta, gamma, delta, theta). A filter in the lower level is responsible for separating the frequency bands in half and gives us High Pass which is detailed coefficient and low pass which is approximation coefficient. Until the desired frequency ranges were achieved, the approximation coefficients were further passed through filters. As the filters were being applied one by one, the technique is called Filter Banks. The whole process is applied for each channel and each sub-bands. As there were 62 Channels and each channel had 5 sub-bands we got $(62 \times 5) = 310$ features as Wavelet Energy and 310 features as Shannon Entropy. Later on we took, mean of Wavelet Energy and Shannon entropy then appended to our main dataset. Lastly, we chose Daubechies wavelet of order 6 (db6) because its soothing feature was appropriate

for detecting changes of the EEG signals.

Wavelet Energy and Shannon Entropy were calculated using the Equation 3.4 & Equation 3.5

$$P_i = Data_i^2 \quad (3.3)$$

$$Wavelet_Energy = \sum_{i=0}^n \log_2(P_i) \quad (3.4)$$

$$Shannon_Entropy = - \sum_{i=0}^n P_i \times \log_2(P_i) \quad (3.5)$$

3.3.3 Eye Movement Features

From the raw data collected from SMI eye-tracking glass, both statistical and computational features were extracted. The parameters that were considered are (1) Pupil Diameter (X and Y) (2) Dispersion (X and Y) (3) Fixation Duration (ms) (4) Blink Duration (ms) (5) Saccade (6) Event Statistics. Among these parameters, 33 features were extracted in which Mean, Standard Deviation and many other statistical and computational features were included.[50] The 33 features that were extracted from the raw data are summarized in Table 3.1.

Table 3.1: Extracted Features from Eye-Movement Data

Eye movement parameters	Extracted Features	Total Features
Pupil Diameter (X and Y)	Mean, Standard Deviation, DE in four bands (0–0.2Hz,0.2–0.4Hz, 0.4–0.6Hz,0.6–1Hz)	12
Dispersion (X and Y)	Mean, Standard deviation	4
Fixation duration (ms)	Mean, Standard deviation	2
Blink duration (ms)	Mean, Standard deviation	2
Saccade	Mean and standard deviation of saccade duration(ms) and saccade amplitude($^{\circ}$)	4
Event statistics	Blink frequency, fixation frequency, fixation duration maximum, fixation dispersion total, fixation dispersion maximum, saccade frequency, saccade duration average, saccade amplitude average, saccade latency average	9
Total Features		33

After extracting features successfully, we now have 33 eye movement features, 310 Differential Entropy features and 2 features of mean value of 310 data points extracted as Wavelet Energy and Shannon Entropy. Therefore, a fusion would be a great approach to combine all the features accordingly. For our study we have used early fusion technique which can be visualized from the Figure 3.4. Therefore, after merging all these features, in total we have $(310 + 33 + 2) = 345$ features representing one emotion.

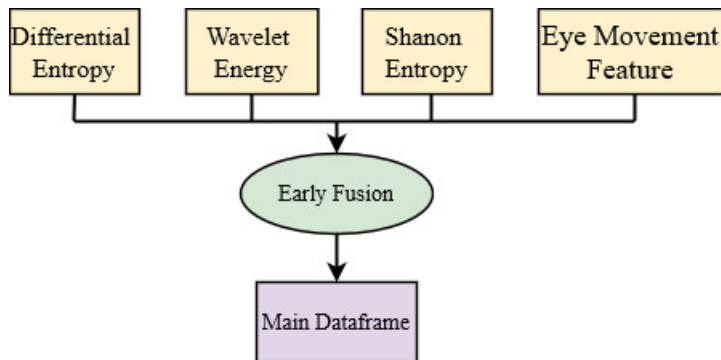


Figure 3.4: Early Fusion Technique

3.4 Exploratory Data Analysis (EDA)

After analyzing the raw data of one participant, it was found that there were some unwanted channels which were present in the data. So, the raw data was with 66 channels at first. Therefore, the unwanted channels namely 'M1', 'M2', 'VEO', 'HEO' were removed and presented with only 62 channels. Raw Data information can be visualized from Table 3.2.

Table 3.2: Raw Data Information of one participant from one Session

Channel Names	FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8, ...
Number of Channels	66
Custom Reference Applied	False
Highpass	0.0 Hz
Lowpass	500.0 Hz
Meas Date	2018-04-08 05:35:05 UTC
Signal Frequency	1000.0 Hz
Subject Information	5 Items(dict)

In the Figure 3.5 the signal of raw EEG data of one participant from one session can be observed. Moreover, from the Figure 3.6 it can be seen that how each emotion look like from raw EEG data.

Furthermore, a comparison of each emotions, how they differ from each other in terms of differential entropy features is shown in the Figure 3.7. Similarly, comparison based on eye movement features of each emotions can be observed in the

Figure 3.8. In these graphs, blue, orange, green, red and purple lines represent the happy, sad, disgust, neutral, fear emotions respectively.

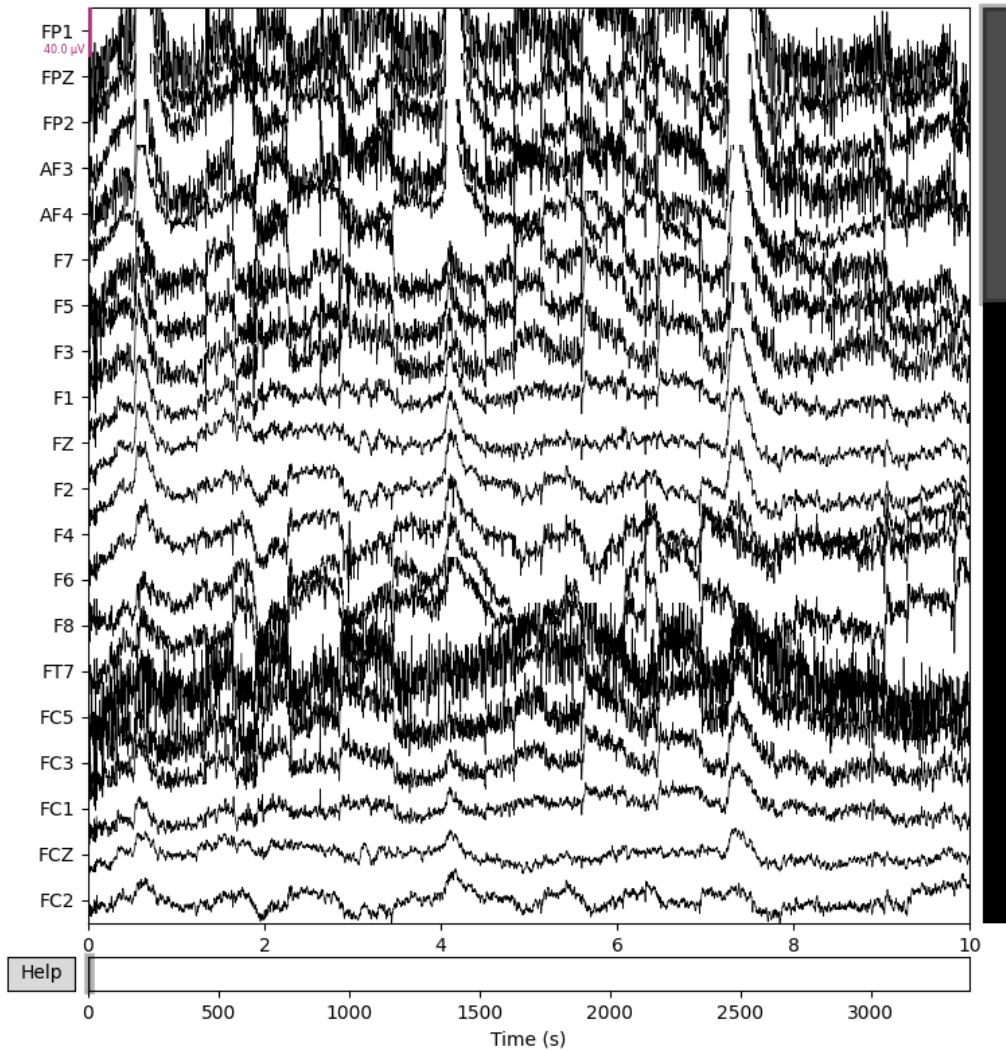
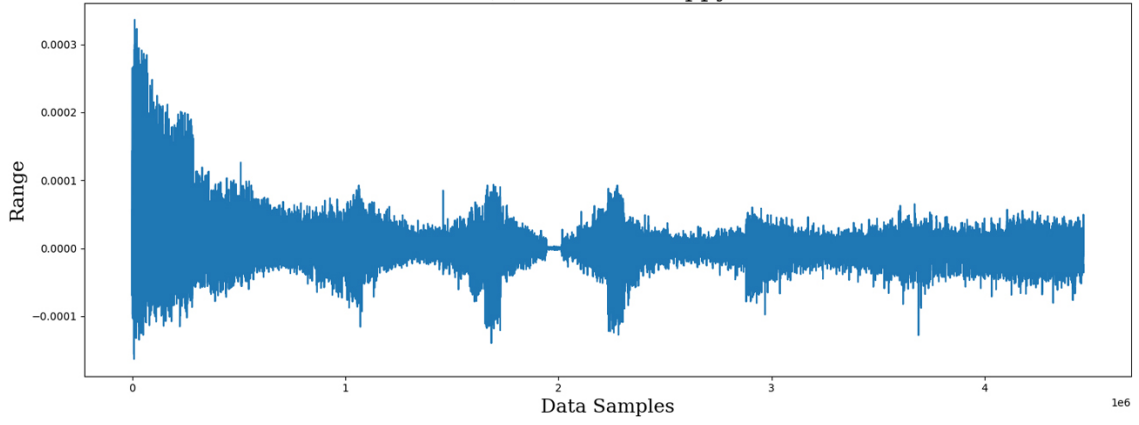
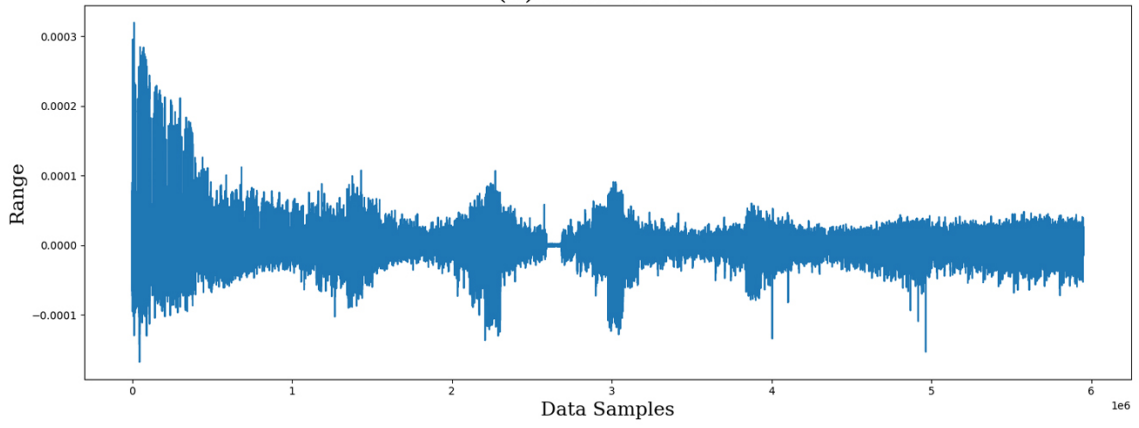


Figure 3.5: Raw EEG Data of one participant from one Session

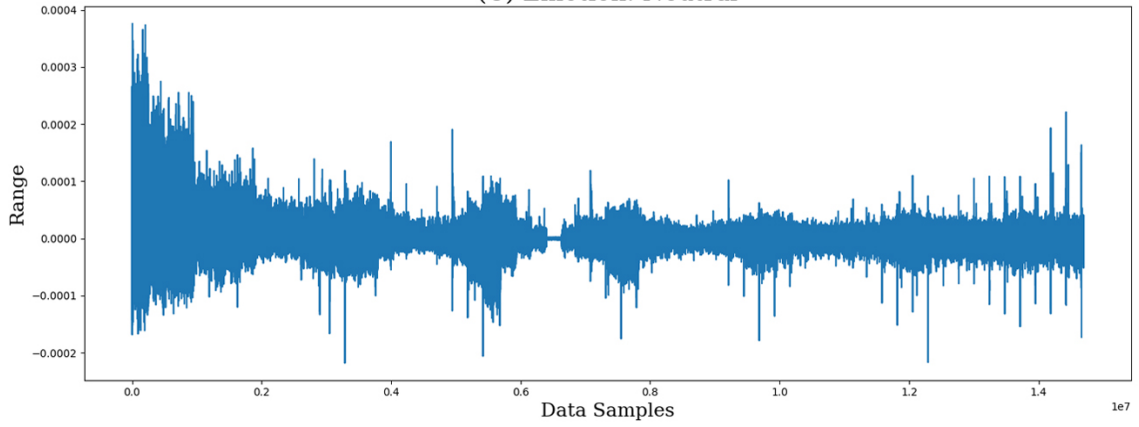
(A) Emotion: Happy



(B) Emotion: Fear



(C) Emotion: Neutral



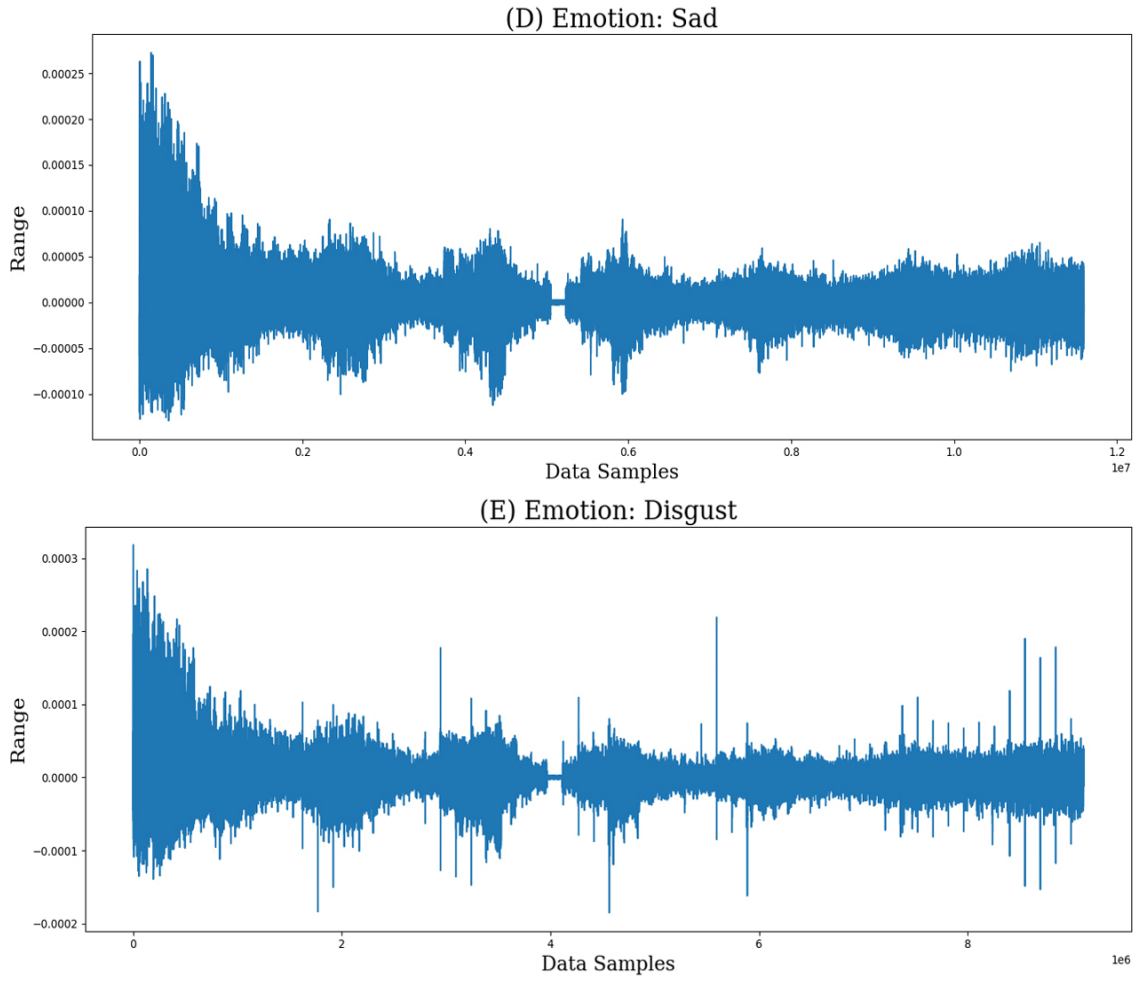


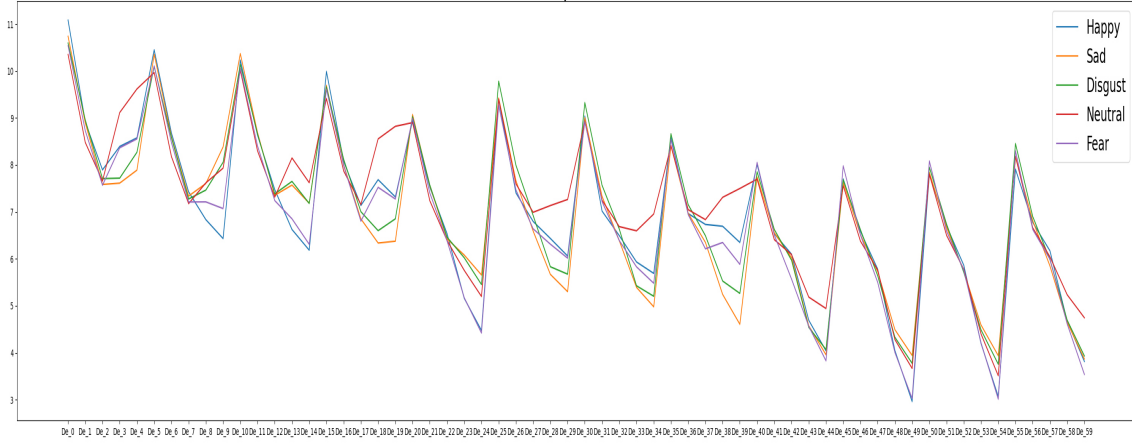
Figure 3.6: Emotion graphs from Raw Data

The mean, maximum, minimum, standard deviation of the values of Raw EEG data of each emotion can be found in the Table 3.3.

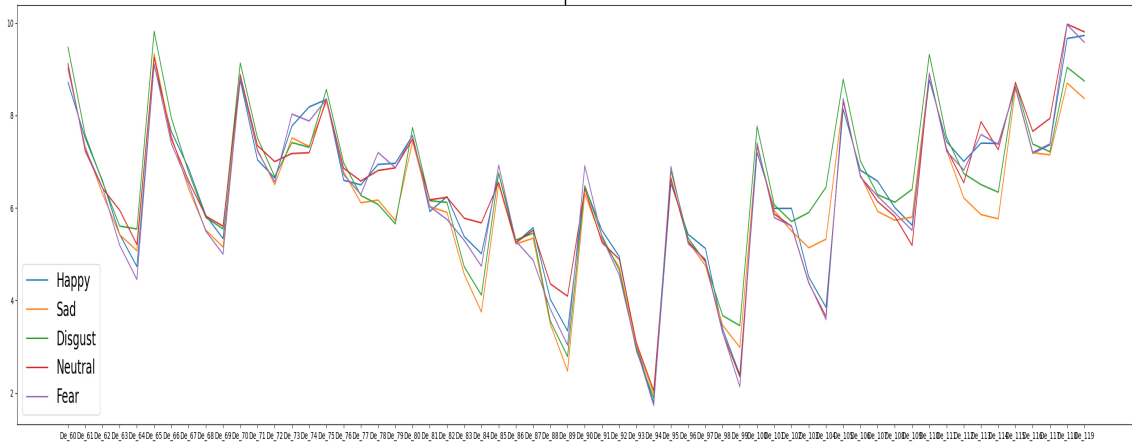
Table 3.3: Data distribution of each emotion of Raw EEG Data

Parameters	Happy	Fear	Neutral	Sad	Disgust
Mean	1.614e-07	-1.223e-07	2.494e-08	1.138e-07	1.226e-07
Maximum Value	3.36e-04	3.20e-04	3.76e-04	2.72e-04	3.18e-04
Minimum Value	-1.63e-04	-1.68e-04	-2.18e-04	-1.29e-04	-1.85e-04
Standard Deviation	1.813e-05	1.675e-05	1.699e-05	1.584e-05	1.597e-05

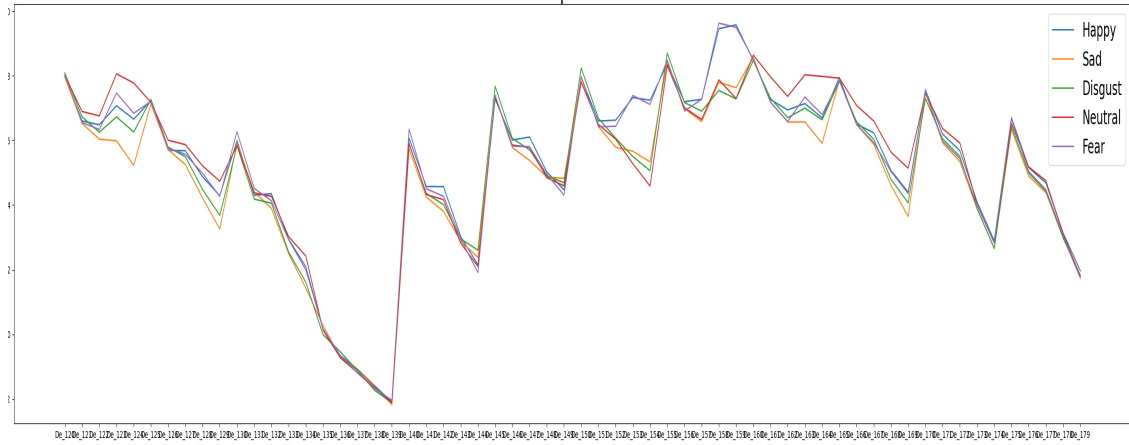
A: Emotion Graph of DE Features



B: Emotion Graph of DE Features



C: Emotion Graph of DE Features



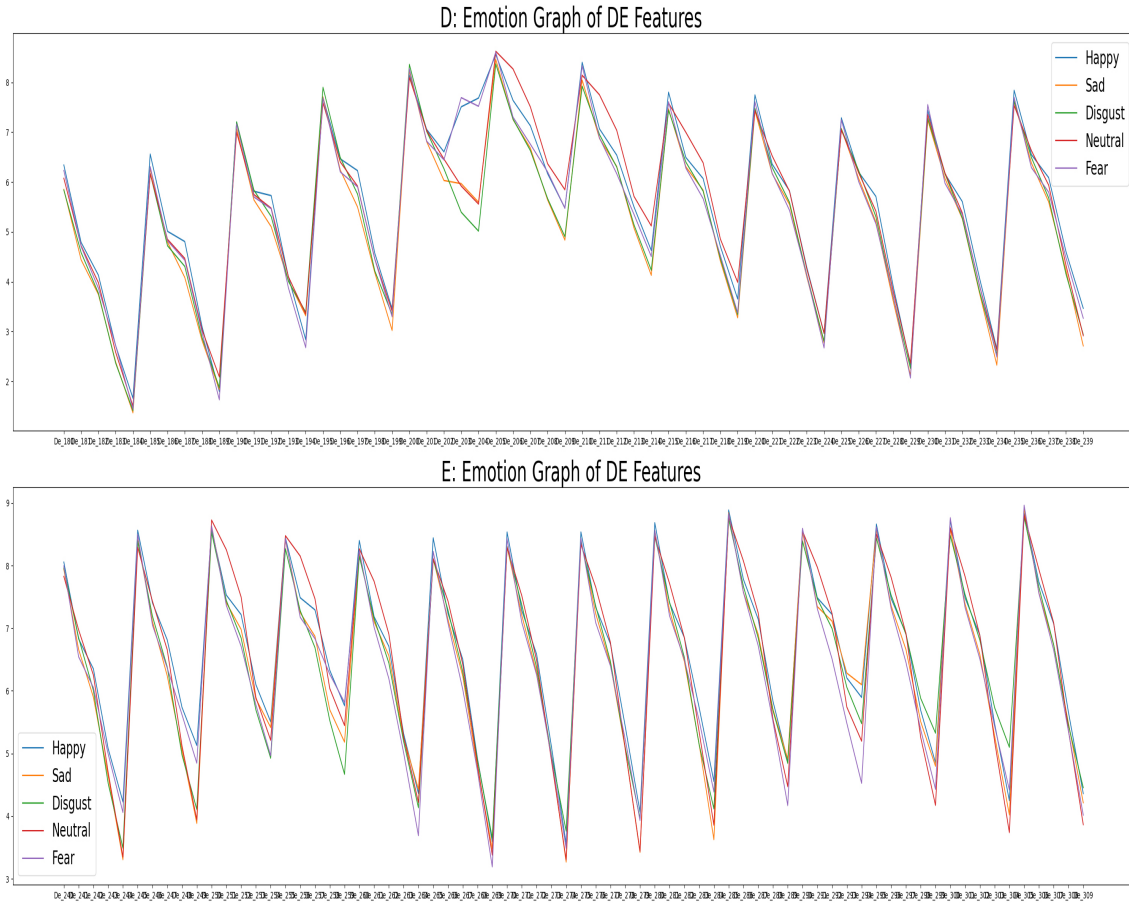


Figure 3.7: DE Features: How each emotion differs from each other

The mean, maximum, minimum, standard deviation of the values of Differential Entropy features of each emotion can be found in the Table 3.4.

Table 3.4: Data distribution of each emotion of DE features

Parameters	Happy	Fear	Neutral	Sad	Disgust
Mean	6.176	6.038	6.211	5.912	6.039
Maximum Value	11.0924	10.540	10.529	10.739	10.596
Minimum Value	-2.100	-2.045	-2.118	-2.176	-2.122
Standard Deviation	1.969	2.001	1.990	1.961	1.970

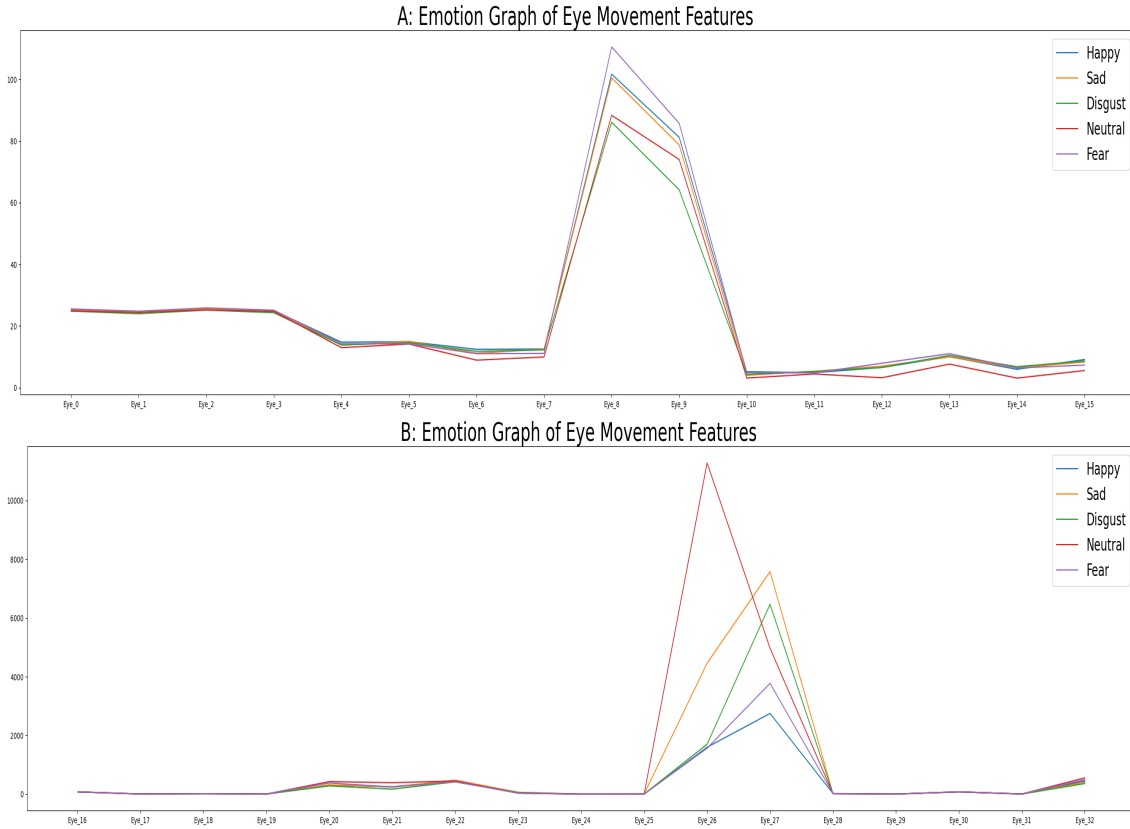


Figure 3.8: Eye Movement Features: How each emotion differs from each other

The mean, maximum, minimum, standard deviation of the values of Eye Movement features of each emotion can be found in the Table 3.5.

Table 3.5: Data distribution of each emotion of Eye Movement Features

Parameters	Happy	Fear	Neutral	Sad	Disgust
Mean	194.300	227.209	564.087	427.783	302.827
Maximum Value	2747.3	3773.4	11280.1	7577.1	6464.6
Minimum Value	0.4	0.3	0.3	0.3	0.3
Standard Deviation	534.936	688.365	2076.434	1475.489	1129.317

The number of label counts for each emotion of the SEED-V dataset are shown in the Table 3.6 and can be visualized from the Figure 3.9. It can be observed that Happy and Disgust emotions have relatively less label count than the other emotions.

Table 3.6: Label counts for each emotion of SEED-V dataset

Emotion	Label Counts
Disgust	4896
Fear	5968
Sad	7616
Neutral	5872
Happy	4816

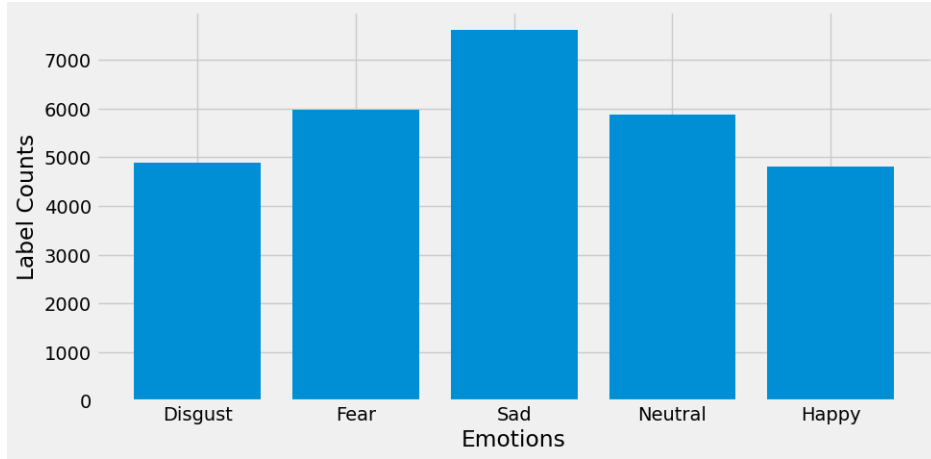


Figure 3.9: Label counts for each emotion

3.5 Data Pre-processing

After merging all the features, there were few cells with null values, therefore those values needed to be handled. The cells with the missing values are replaced with the values of the same emotions of the same person. Later on, categorical encoding was used to encode the emotion labels which are shown in the Table 3.7.

Table 3.7: Categorical encoded values of emotions

Emotion	Encoded Value
Disgust	0
Fear	1
Sad	2
Neutral	3
Happy	4

The shape of the dataset was 29168×345 excluding the label or target column. Therefore, two types of reshaping techniques were used to apply the CNN and LSTM. CNN is popularly used to classify images, therefore the model’s input shape requires a 3D shaped data which means in the first layer, the input_shape takes 3 parameters where the first parameter represents channel numbers, the last two input parameters represent the height and width of the input image.

However, the height and width of the image have to be the same. That is why, some of the features needed to be dropped in order to reshape the dataset into the same height and width dimension. The features were dropped by analyzing the Figure 3.7 & Figure 3.8 where the features are not so correlated with the label and the value of the features is almost the same for each of the emotions.

The dropped features were De_20, De_7, De_16, De_35, De_51, De_52, De_86, Eye_0, Eye_1, Eye_2, Eye_3, Eye_5, Eye_17, Eye_18, Eye_19, Eye_22, Eye_23, Eye_24, Eye_29, Eye_30, Eye_31. After dropping these features, the new shape of the dataset was 29168 x 324 and reshaped into (29168, 1, 18, 18) where the values represent total number of inputs, number of channels, height and width of the input respectively. In other words, the signal is represented as a gray scale image.

Sample values contained a range of values which needed to be scaled in order to get good performance results. Therefore, the sample values were scaled in between 0 to 1 using Standard Scaler which uses the Equation 3.6.

$$z = \frac{x - u}{s} \tag{3.6}$$

Here,

x → Sample Value

z → Scaled Value

u → Mean of the Samples

s → Standard Deviation of the Samples

Now, label and desired features are splitted by the ratio of 75:25. In short, 75% data is for training and 25% data is for testing. Finally, our data is ready to be fitted into the models.

3.6 Model Specification for Emotion Classification

In order to classify our data, we relied on some of the most popular deep learning models, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network, a variety of Recurrent Neural Network (RNN). We also tried machine learning models, such as Support Vector Machine (SVM), Random Forest (RF), however machine learning models did not perform well. Hence we discarded those models.

Both CNN and LSTM worked pretty well on our dataset. Although we have used both the models to evaluate the performance of our test data, we only used the model to predict the emotion that has performed better than the other in terms of evaluation metrics. The comparison between the two models will be addressed in the result analysis part.

3.6.1 Convolutional neural network (CNN)

Convolutional Neural Network (CNN) or ConvNet is one the most popular neural networks when it comes to Computer Vision. It is the number one choice if the assigned task is to detect objects, image processing or any other sectors of Computer Vision. CNN works better with data that are in grid-like format, for example: an image. Moreover, in recent years, it has also been proven that CNN also works well with Signal data because of its feature recognition ability.

CNN works in a similar fashion as our brain. The brain receives the stimuli and processes the information by connecting neurons in such a way that by looking at a thing once, we manage to identify the object. By identifying patterns, features from complex data CNN also processes in the same way. A simple CNN consists of a convolutional layer, pooling layer and a dense layer where neurons are fully connected. The number of these layers depends on the task that has been assigned. Figure 3.10 represents the simplest form of CNN architecture.

At first, the input shape has to be formatted properly where the shape will be

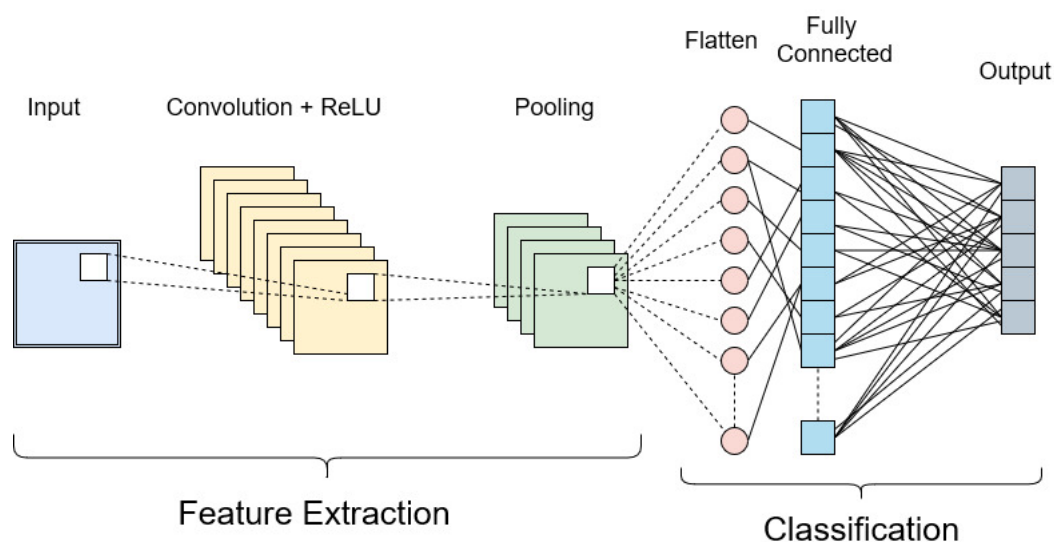


Figure 3.10: Simple CNN Architecture

3-D (Three Dimensional). As, CNN is a special deep learning model which was improvised for image classifications or object detection, the model works better with 3-D data where the height and width of the image will be specified and the other parameter will be the number of channels. For example $\text{input_shape} = (32,32,3)$ means the image's height and width are respectively 32 and there are 3 channels (Red, Green, Blue). By looking at the number of channels, we can say that the image is a colorful image.

In the first convolutional layer, the input shape has to be specified. The next task is to specify the number of filters or kernels we want and the size of the kernel. The kernels will help find the pattern throughout the data. Two of the most important parameters of the convolutional layer are "Stride" and "Padding". 'Valid Padding' means there will be no extra layer outside of the main layer and 'Same

'padding' means there will be even layers that will be added to the main layer and Stride means how many steps the filters will take while shifting. Afterwards, dot product between the input layer and kernel takes place and Rectified linear unit (ReLU) works on the product Matrix making the model non-linear and also solves the vanishing gradient problem. The function of ReLU is defined at Equation 3.7.

$$f(x) = \max(0, x) \tag{3.7}$$

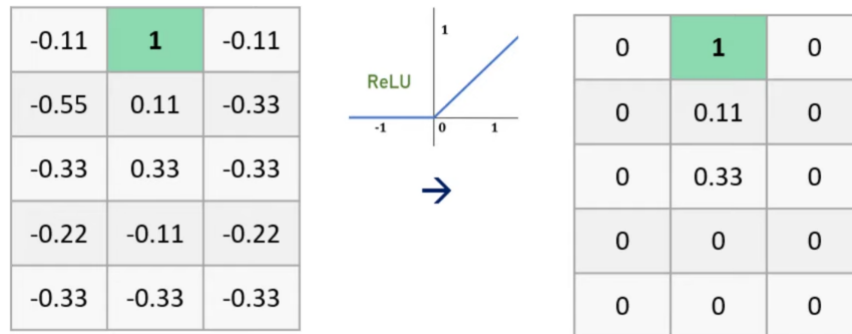


Figure 3.11: How Rectified Linear Unit(ReLU) works

After the convolutional layer, we do pooling to reduce the size of the matrix as well as to decrease the amount of trainable parameters. There are two types of pooling (i) Average pooling and (ii) Max pooling. Then, the trained parameters are flattened and passed through a dense, fully connected hidden layer to output. The dimension of the output layer after each convolutional layer and pooling layer can be calculated using the Equation 3.8

$$output = \frac{input_size - kernel_size + 2 \times padding}{stride} + 1 \tag{3.8}$$

In order to construct our CNN model we decided to have three convolutional layers with 'Valid Padding' and strides = (1,1) along with the three max_pooling layers with pool_size = (2,2), 'Valid padding' and strides is set to default which is None, this will eventually takes the value specified in pool_size. The convolutional layers have 114, 100, 32 filters respectively. The first two convolutional layers have the kernel_size =(3 x 3) and the last layer has the kernel_size =(2 x 2).

After each convolutional layer and max_pooling layer we have added dropout layers of 50% so that we can avoid overfitting problems. After the last max_pooling layer the data has been flattened and gets connected with two dense layers having 60, 32 neurons respectively. In each dense layer, we have used ReLU as our activation function. After each dense layer, a dropout layer of 50% and 30% has been added respectively. Lastly, the output layer is also densely connected with the previous hidden layer and the activation function here is 'Softmax' which returns an array

having probabilistic values of target labels.

We have trained our model with the loss function ‘Categorical Crossentropy’ as our labels are one hot encoded and ‘Adam’ as the optimizer. Moreover, we considered 45 epochs as our threshold to get to the optimal result having the batch size of 128. In each backpropagation, the weights of the filters are updated and has a strong mathematical background.

In the convolutional layers, we got 1140, 51400, 6432 parameters respectively which can be calculated using the Equation 3.9. After that, we flattened our data and received $(13 \times 1 \times 16) = 208$ neurons followed by two dense layers having 60, 32 neurons accordingly. After each of the dense layers 50% and 30% data has been dropped out, solving the overfitting problem. Hence, in total we got 73,629 trainable parameters.

$$TotalParameters = (k \times k \times c \times n) + n \quad (3.9)$$

Where,

k = Kernel Size

c = Number of Channels or number of filters in the previous layer

n = Number of filters in the layer

The overall architecture of the model can be visualized from the Figure 3.12.

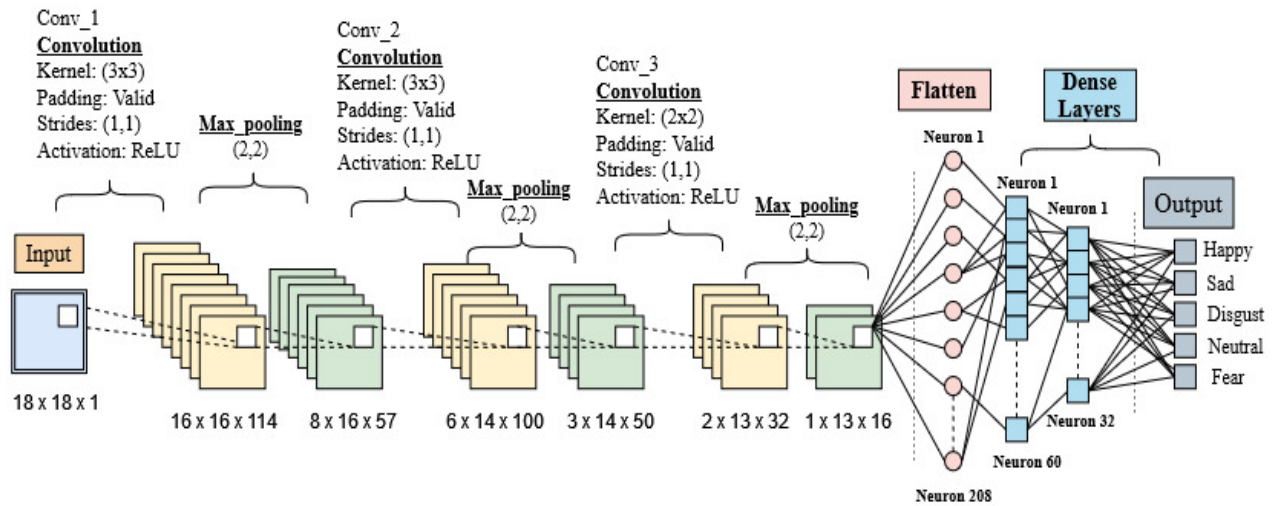


Figure 3.12: Proposed Convolutional Neural Network (CNN) Architecture

Table 3.8: Proposed Convolutional Neural Network Summary

Layers	Output Shape	Parameters
Conv2d Layer_1	(None, 16, 16, 114)	1140
Max_pooling Layer_1	(None, 8, 16, 57)	0
Dropout: 50%	(None, 8, 16, 57)	0
Conv2d Layer_2	(None, 6, 14, 100)	51400
Max_pooling Layer_2	(None, 3, 14, 50)	0
Dropout: 50%	(None, 3, 14, 50)	0
Conv2d Layer_3	(None, 2, 13, 32)	6432
Max_pooling Layer_3	(None, 1, 13, 16)	0
Dropout: 50%	(None, 1, 13, 16)	0
Flatten	(None, 208)	0
Dense: hidden_layer_1	(None, 60)	12540
Dropout: 50%	(None, 60)	0
Dense: hidden_layer_2	(None, 32)	1952
Dropout: 30%	(None, 32)	0
Output Layer	(None, 5)	165
Total parameters		73,629

3.6.2 Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNN) is one of the most widely used deep learning models that is used to predict sequential or time series data. Which means, whenever we have to classify some data that is related to its previous parts, we can use RNN. However, RNN has some short-comings. In sequential data, if the model does not have to memorize too much information, RNN performs well. This is called the “Long term Dependency” problem. Moreover, the vanishing gradient problem is also prominent in this deep learning mode. Gradient descent algorithm is used to update weights. As the model goes deeper into the lower layers, the weights hardly change and the model can not learn to its full potential. In order to solve this problem Long Short Term Memory networks (LSTM) was introduced. It is a type of RNN that has a memory cell to store words carefully, input gate, output gate and forget gate.

In the Figure 3.13, for the RNN part, a_{t-1} represents the activation function which gets updated in every iteration, x_t represents the input data, and o_t represents the output. For the LSTM part, c_{t-1} represents memory cell responsible for storing long term memory and consists of the three gates that have discussed before. h_{t-1} represents short term memory same as the one that is present in the RNN.

Forget gate (f_t) decides whether to retain an information or forget an information that is present in the memory cell. Input gate(i_t) works with Candidate value (\tilde{C}_t). Candidate value is responsible for adding new information into the memory cell. Both the forget gate and input gate uses ‘Sigmoid’ activation function, therefore the value of the gates stays between either close to 0 or close to 1. Hidden state(h_t) value works as an output and its calculated by multiplying the values of Output gate(o_t)

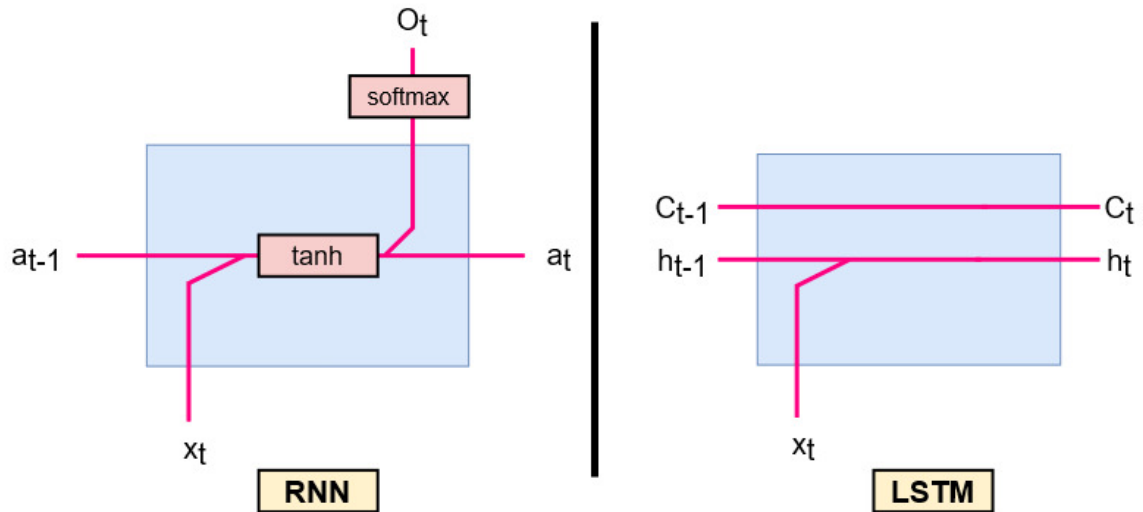


Figure 3.13: RNN vs LSTM

and \tanh of current Memory Cell State (C_t). From the Figure 3.14 we can visualize the architecture of LSTM. Also, all the formulas related to LSTM architecture can be found from Equation 3.10 to Equation 3.15.

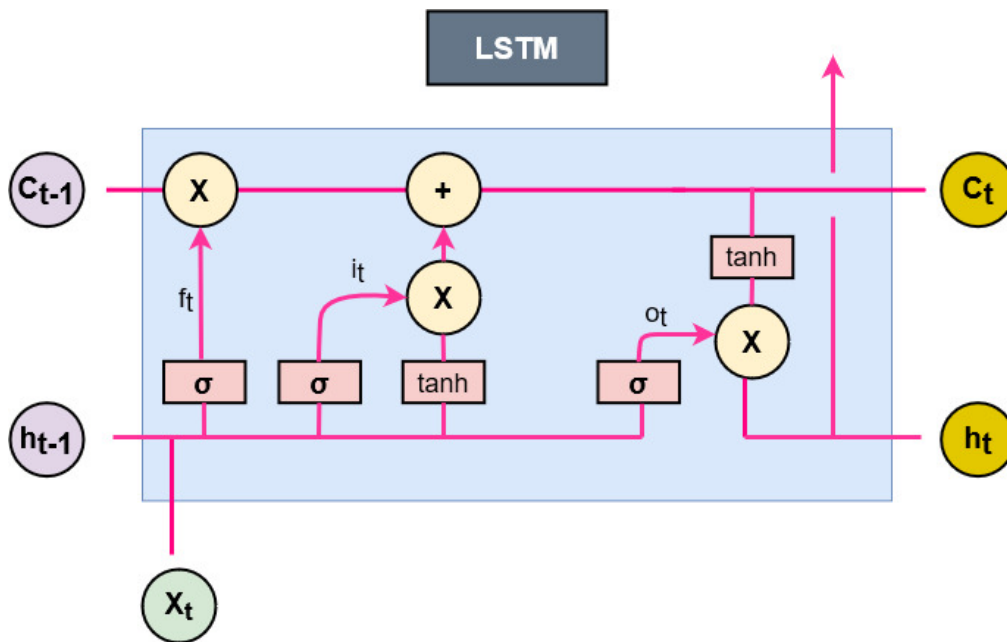


Figure 3.14: Inside Architecture of LSTM

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (3.10)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.11)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.12)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.13)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.14)$$

$$h_t = o_t \times \tanh(C_t) \quad (3.15)$$

Here,

C_t →Memory Cell State

C_{t-1} →Previous Memory cell state

\tilde{C}_t →Candidate Value

h_{t-1} →Previous Hidden State

x_t →Input Vector

f_t →Forget Gate

i_t →Input Gate

o_t →Output Gate

W →Associated Weight

b →Bias

In our case, for each emotion, the features we have extracted follows a particular sequence. That is why we are using LSTM which can store previously occurred important input value and can classify based on that. We have built our LSTM model using two LSTM layers having 64 and 32 hidden units respectively with the `input_shape = (345,1)` where the first parameter denotes the number of neurons as inputs or the timesteps and the second parameter denotes the number of features needed to represent the data of one timestep. The “`return_sequence`” parameter has been kept as “True” in order to return the full output sequence and we are using the “tanh” activation function.

After each LSTM layer we have added dropout layers of 40% so that we can mitigate overfitting problems. After the LSTM layers, there are three dense layers or hidden layers using ReLU as their activation function and having 32, 16, 8 neurons respectively. A dropout layer of 40%, 40% and 30% has been added after each dense layer. Last of all, the last hidden layer is densely connected with the output layer which is predicting the emotions (Happy, Sad, Disgust, Neutral, Fear). Activation function “Softmax ” has been used in the output layer. Moreover, in order to train our model we used the ‘Sparse Categorical Crossentropy’ loss function, ‘Adam’ as the optimizer, batch size of 128 and we considered 41 epochs so that we can get an optimal result.

In the LSTM layers, we got 16896, 12416, 1056 parameters respectively which can be calculated using the Equation 3.16. After that we have three strongly connected dense layers of which parameters can be calculated by using the Equation 3.17. Therefore, in the first hidden layer we have $((32 \times 32) + 32) = 1056$ parameters. By using the similar Equation 3.17, second and third hidden layers have 528, 136 neurons each in order. For that reason, we have a total of 31,077 trainable parameters.

$$N_p = 4 \times \{(h + f) \times h\} + h \quad (3.16)$$

Here,

N_p → Total number of parameters

h → Number of hidden units in the present layer

f → Number of features, which is defined in the 2nd parameter of `input_shape` (for the first LSTM layer). Afterwards, f denotes the number of hidden units present in the previous LSTM layer.

$$d_p = (n_{prev} \times n_{present}) + bias \quad (3.17)$$

Here,

d_p → Dense layer parameters

n_{prev} → Number of neurons present in the previous layer

$n_{present}$ → Number of neurons present in the current layer

$bias$ → Equals to the number of neurons that is present in the current layer

The overall architecture of the LSTM model can be visualized from the Figure 3.15.

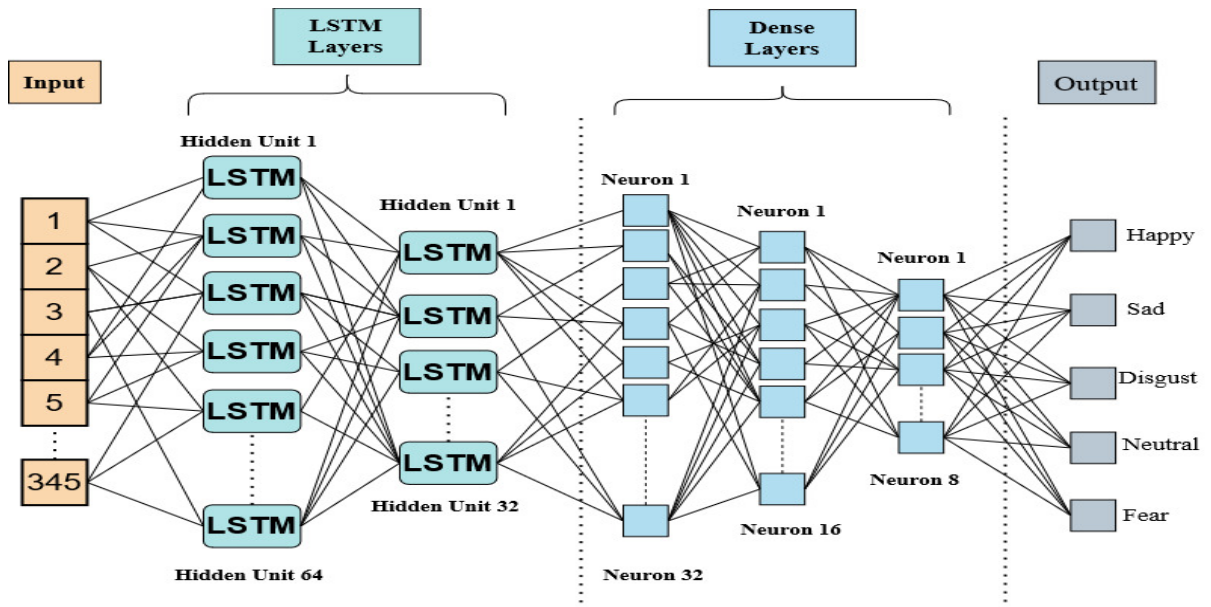


Figure 3.15: Proposed Long short-term memory (LSTM) Architecture

Table 3.9: Proposed Long short-term memory (LSTM) Summary

Layers	Output Shape	Parameters
LSTM Layer_1	(None, 345, 64)	16896
Dropout: 40%	(None, 345, 64)	0
LSTM Layer_2	(None, 32)	12416
Dropout: 40%	(None, 32)	0
Dense: hidden_layer_1	(None, 32)	1056
Dropout: 50%	(None, 32)	0
Dense: hidden_layer_2	(None, 16)	528
Dropout: 40%	(None, 16)	0
Dense: hidden_layer_3	(None, 8)	136
Dropout: 30%	(None, 8)	0
Output Layer	(None, 5)	45
Total parameters		31,077

3.7 Recommendation System

The key challenge for the recommendation system is to understand a particular individual's choice as it varies significantly from individual to individual. For this, recommendation models are proposed, which prioritizes user's previous experience, particularly likes and interests, to understand their choice of interest based on these. This understanding and implementation has to be done accurately. Hence, we need to build the recommendation system in such a way so that it can achieve accurate understanding and implementation by the metrics of similarity. However, finding just correlation will not suffice, that must also be based on interrelations between certain factors.

There are two types of recommendation systems. (i) Content Based Filtering (CBF) (ii) Collaborative Filtering (CF). However, CBF does not work with user related data. It does the job by looking at the data, recommending a user based on their past experience only. Other users of similar interest do not play any role here. Furthermore, this model works better only if the user already has an interest recorded somewhere in the database. On the other hand, CF works with the data of test users as well as other similar users to find the best content for the user. Moreover, it can be recommended far better than CBF. Hence, in this study, for the recommendation part we used Collaborative Filtering and to find the relation between the items we have used Pearson Correlation method.

Furthermore, recommending a content based on Correlation is not enough as the Pearson correlation scores can be biased as it is purely constructed on the ratings of the users. Therefore, it is necessary to understand the emotion of the video by classifying the subtitles of the videos and extract the emotion scores such as Joy, Love, Anger, Sad, Fear, Surprise of the video.

3.7.1 Collaborative Filtering

Collaborative Filtering is a technique which focuses on the relationship between the users and items. It recommends contents by finding the similarity between users by tracking their ratings, likes and interest. Hence, collaborative filtering works way better than content based filtering. This technique makes use of explicit ratings given by users or those inferred from log-archives to make accurate predictions of user's favorable content. User's interaction history such as, ratings, likes and views etc. plays an important role in measuring the similarity between users.

This final rating of an item is determined by merging predictions from three origins. This includes predictions based on ratings (i) of the same item provided by other users (ii) of different items by the same user (iii) of other similar users. Let's say we have two users, User A and User B. User A prefers 'item a', 'item b' and 'item c'. User B prefers 'item a' and 'item b'. As both the users like 'item a' and 'item b' it can be inferred that the users have similar kind of taste. Hence, 'item a' , 'item b' & 'item c' will have high correlation. Therefore, 'item c' will be recommended to user B, considering the fact that the items are identical. The scenario that has been mentioned above can be visualized from the Figure 3.16.

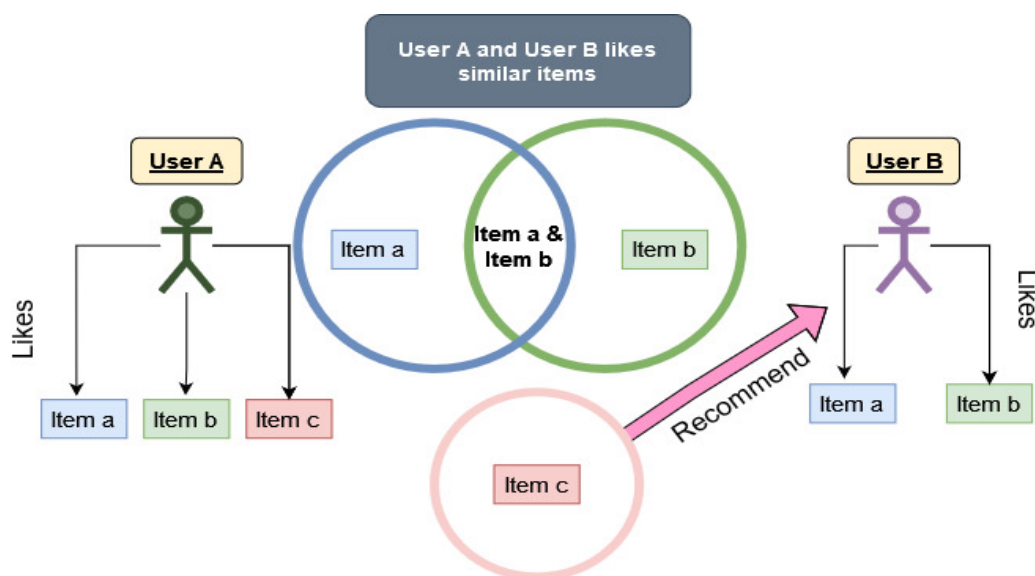


Figure 3.16: Collaborative Filtering: Example 1

There are two types of Collaborative Filtering Techniques. They are (i) Memory Based Approach and (ii) Model Based Approach. Memory based approach can be further categorized into (i) Item-item filtering (ii) User-item filtering. As we do not have a big chunk of data we can not apply model based filtering on our dataset. Therefore, we are using a Memory based approach, more specifically, Item-item filtering.

In item-based collaborative filtering, if an item is given, the system finds out the user who liked the item and fetches the top most correlated items from their history. Here, the correlation between items is prioritized.

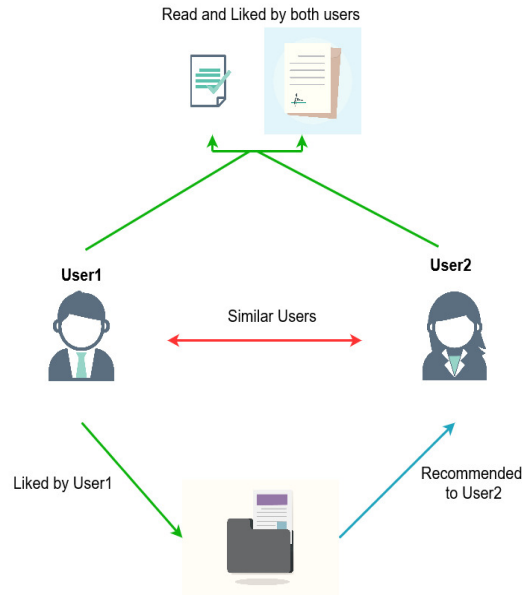


Figure 3.17: Collaborative Filtering: Example 2

3.7.2 Pearson Correlation

Cosine similarity and Pearson Correlation Coefficient are two of the most commonly used techniques to measure the similarity or correlation between the items. For this study, the similarity is calculated using the Pearson correlation coefficient to determine the degree of correlation between two items.

Pearson correlation coefficient works in a simple way in which it finds the linear changes between two items. The correlation formula ranges from -1 to 1 where '1' represents the most correlated and '-1' represents most inversely correlated. We can visualize this clearly from the Figure 3.18

We can calculate Pearson Correlation between two items by using the Equation 3.18.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \times \sum x^2 - (\sum x)^2][n \times \sum y^2 - (\sum y)^2]}} \quad (3.18)$$

Here,
 n →Datapoints
 x →Item 1
 y →Item 2

The whole process of finding out the Pearson Correlation of the Youtube videos is described in the Algorithm 1.

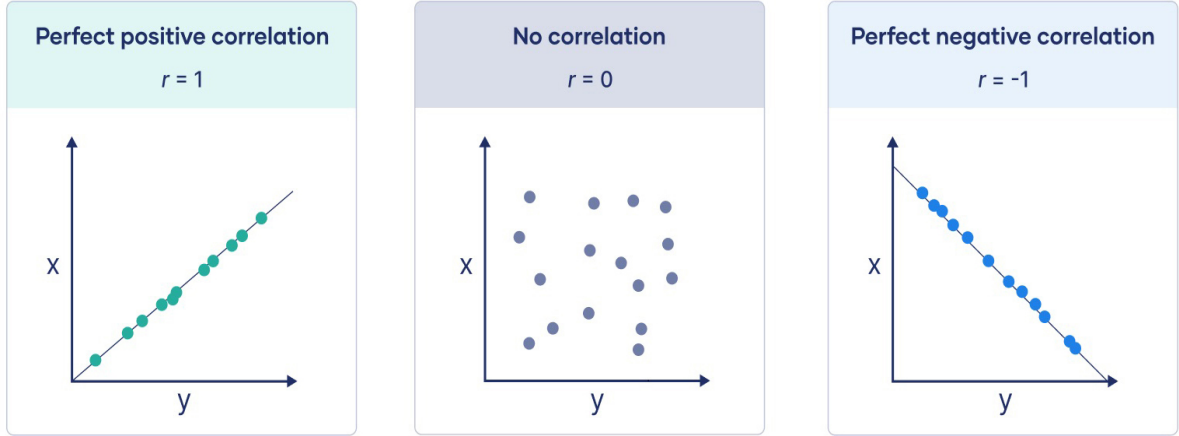


Figure 3.18: Pearson Correlation

Algorithm 1 Extraction of Pearson Correlation Scores

Input:

R_f : A Dataframe having ratings of various users

V_f : A Dataframe having the video names, type and links

Output: $Pearson_s$: A Dataframe that hold Pearson Correlation Scores

- 1: Initialize a dataframe to hold merged dataframe as M_f
 - 2: $M_f \leftarrow Merge(R_f, V_f)$
 - 3: $U_r \leftarrow$ Pivot table of M_f
 - 4: $U_r \leftarrow PreProcessing(U_r)$ ▷ Dropping columns, filling Null values with 0
 - 5: Initialize a Dataframe to hold Pearson Correlation values as I_{sd}
 - 6: $I_{sd} \leftarrow$ Correlation among videos using *Pearson* method
 - 7: Initialize a list that holds tuple of video names along with their ratings as W_v
 - 8: Initialize a Dataframe to hold the correlated videos with scores as $Pearson_s$
 - 9: **for** each value of W_v **do**
 - 10: $Pearson_{s_i} \leftarrow GetSimilarVideos(value)$ ▷ Returns the correlation score
 - 11: $Pearson_{s_i} \leftarrow \sum Pearson_s$ ▷ Column wise summation
 - 12: $Pearson_{s_i} \leftarrow Normalization(Pearson_s)$
 - 13: **End for**
 - 14: **Return** $Pearson_s$
-

3.7.3 Emotion score extraction of Videos

In order to understand the context of a text, text classification is very much needed. The goal is to recommend content that can bring joyfulness to a user. As we have an unlabeled text-classification problem, we can not train the dataset with Machine Learning or Deep Learning models. To solve this issue, pre-trained models came in handy. Therefore, in this study a transformer library of Hugging Face is used to classify the texts.

The model ‘distilbert-base-uncased-emotion’ has been used which is trained on an emotion dataset of Twitter. The model is faster and smaller than any other Bert based models with an accuracy and F1 score of 93.8%. The model takes a text as an input and returns a list of dictionaries where each dictionary holds an emotion and corresponding score. We get six different emotions namely Joy, Love, Fear, Anger, Sad, Surprise.

To begin with, the subtitles of the videos were extracted from the Youtube videos and kept on a text file. Basic pre-processing was done by using the by removing the line breaks, punctuations and stop words with the help of Spacy Library and using ‘en_core_web_lg’ package. Later on, the texts are feeded to the model in order to extract the scores of the texts. Finally, the whole process of classifying the subtitles and extracting the score of emotions is described in the Algorithm 2 and can be visualized from Figure 3.19.

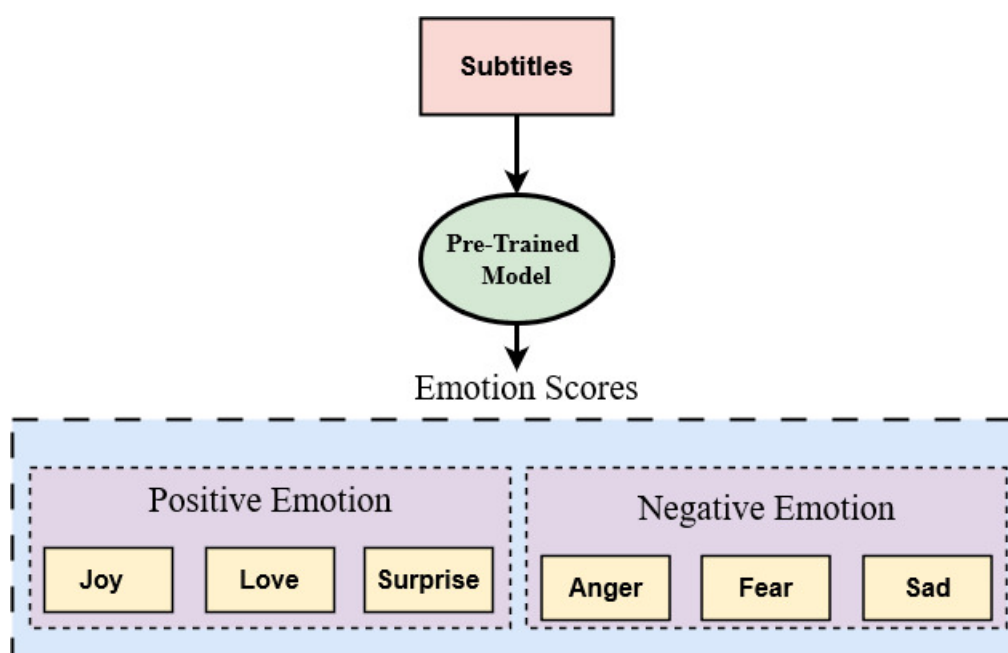


Figure 3.19: Six emotions classified into Positive and Negative emotions

Algorithm 2 Extraction of emotion scores of videos through subtitles

Input: F_t : A column of subtitles of the videos $Subtitle_{df}$: A Dataframe having the video names and subtitles**Output:** NLP_s : A Dataframe that hold Positive and Negative emotion scores of the videos

- 1: Initialize a new column in the Dataframe, P_t as Pre-processed text
- 2: **for** each value of F_t **do**
- 3: $P_{t_i} \leftarrow PreprocessAndVectorize(value)$ \triangleright PreprocessAndVectorize() takes a string and removes the line breaks, stop words and white spaces
- 4: **End for**
- 5: Initialize text classification model as $Model$
- 6: Initialize a temporary list to store all scores of emotions of a video as L_1
- 7: Initialize a list to store all the temporary lists as L_2
- 8: **for** each value of P_t **do**
- 9: $L_1 \leftarrow Model(value)$ \triangleright L_1 holds the scores of Joy, Love, Surprise, Fear, Anger, Sad emotions
- 10: $Pos_emotion \leftarrow (\sum \text{scores of Joy, Love, Surprise emotions})$
- 11: $Neg_emotion \leftarrow (\sum \text{scores of Sad, Fear, Anger emotions})$
- 12: $L_1.append(Pos_emotion)$
- 13: $L_1.append(Neg_emotion)$
- 14: $L_2.append(L_1)$
- 15: **End for**
- 16: Initialize a dataframe to hold only scores of the videos as NLP_s
- 17: $NLP_s \leftarrow DataFrame(L_2)$
- 18: $NLP_s \leftarrow Concat(Subtitle_{df}, NLP_s)$
- 19: **Return** NLP_s

3.7.4 Analytic Hierarchy Process (AHP)

The 2 dataframes (i) Text Classification Dataframe containing the scores of Happy/Positive and Sad/Negative emotion (ii) Pearson Correlation score of videos Dataframe is merged to one single dataframe for the ease of further calculations. So, the merged dataframe has the name of all the videos and their corresponding positive, negative and Pearson Correlation scores.

In the merged dataframe, Youtube videos can be considered as candidates from which we have to recommend to users. The recommendation of videos can be done by ranking the videos in a certain way. In order to rank the videos, there are 3 variables that need to be taken into consideration, which are Positive Emotion, Negative Emotion and Pearson Correlation score.

There are multiple criteria from which a list of recommended videos has to be listed. This Multi-Criteria Decision-Making (MCDM) problem can be solved with Analytic Hierarchy Process (AHP). The problem which needs to be solved can be visualized from Figure 3.20.

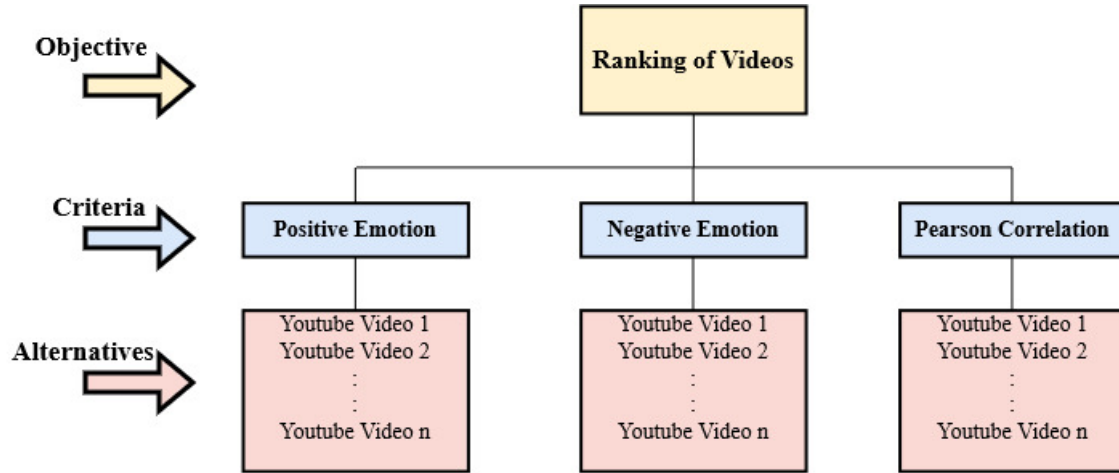


Figure 3.20: Multi Criteria Decision Making

In order to address the problem, the importance level of each criteria needed to be set by following the Table 3.10.[43] Thus we get a pairwise comparison matrix $A = [a_{ij}]_{n \times n}$ which is a square matrix of (n x n) dimension where 'n' is the number of criteria. Moreover, the matrix follows reciprocal properties as stated in the Equation 3.19.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

$$a_{ji} = \frac{1}{a_{ij}} \quad (3.19)$$

The pairwise comparison matrix for our study can be found in Table 3.11. The rating "8" between Positive emotion and Negative emotion states that the preference of Positive emotion is very strong than Negative emotion. The other comparison ratings can be interpreted in a similar manner.

Table 3.10: The 1-9 Fundamental scale

Intensity of importance	Definition
1	Equal importance
2	Weak
3	Moderate importance
4	Moderate plus
5	Strong importance
6	Strong plus
7	Very strong or demonstrated importance
8	Very, very strong
9	Extreme importance

Table 3.11: Pairwise comparison matrix of the criteria

	Positive Emotion	Negative Emotion	Pearson Correlation
Positive Emotion	1	8	4
Negative Emotion	1/8	1	1/5
Pearson Correlation	1/4	5	1

Now that the pairwise comparison matrix is formed, few consecutive steps will be performed.

Firstly, the matrix is normalized using the Equation 3.20

$$a_{ij}^* = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}} \quad (3.20)$$

Where, $j = 1, 2, \dots, n$

Secondly, the criteria weights, $w = [w_1, w_2, \dots, w_n]$ are calculated using the Equation 3.21

$$w_i = \frac{\sum_{j=1}^n a_{ij}^*}{n} \quad (3.21)$$

Where, $i = 1, 2, \dots, n$

Now that the weights of corresponding criteria are found, the validity of the importance rating that has been assigned in Table 3.11 needed to be checked. For this, we need to find the maximum eigenvalue of matrix λ_{max} by using the Equation 3.22 and Equation 3.23. λ_{max} is a validating parameter that holds much importance in the AHP process. [15]

$$M_w = A \times w_i = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{13} \\ a_{21} & a_{22} & \dots & a_{23} \\ \cdot & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ w_n \end{bmatrix} \quad (3.22)$$

Where,

$A \rightarrow$ Pairwise Comparison Matrix

$w_i \rightarrow$ Criteria Weights

In our study, the pairwise comparison matrix A is a (3 x 3) square matrix and the Criteria Weights matrix has a dimension of (3 x 1). Therefore, the resultant matrix M_w after the multiplication has a dimension of (3 x 1).

After that, using the Equation 3.23 λ_{max} can be calculated.

$$\lambda_{max} = \frac{\sum_{i=1}^n \frac{M_{w_{i1}}}{w_{i1}}}{n} \quad (3.23)$$

Using the value of λ_{max} Consistency Index (CI) can be calculated using the Equation 3.24

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (3.24)$$

With the value of CI now we can calculation Consistency Ratio (CR) using the Equation 3.25 . If $CR < 0.1$ the values that are assigned in pairwise comparison matrix in Table 3.11 are said to be valid or acceptable. Otherwise, the pairwise comparison matrix have to be reconstructed.

$$CR = \frac{CI}{RI} \quad (3.25)$$

Where,
 $RI \rightarrow$ Random index

The values of Random Index (RI) can be found in Table 3.12. As we have 3 criteria, the value 0.58 has been used in our study. [15] Furthermore, the value of CR is 0.09 which less than 0.1. Therefore, the pairwise comparison matrix is valid and we can use the criteria weights, w . The whole process of finding the criteria weights, w and validating the pairwise comparison matrix is described in the Algorithm 3.

Algorithm 3 Analytic Hierarchy Process (AHP)

Input:

I_t : A pairwise comparison matrix of Positive, Negative emotions and Pearson Scores
 n : Number of criteria

Output: w : Criteria Weights

- 1: Initialize $n = 3$
 - 2: $a_{ij} \leftarrow$ a cell from I_t
 - 3: Initialize a matrix where each cell a_{ij}^* will be normalized
 - 4: $a_{ij}^* \leftarrow \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}$ where $j = 1, 2 \dots n$
 - 5: Initialize Criteria Weights, w
 - 6: $w_i \leftarrow \frac{\sum_{j=1}^n a_{ij}^*}{n}$ where $i = 1, 2 \dots n$
 - 7: Initialize a matrix, M_w
 - 8: $M_w \leftarrow I_t \times w_i$
 - 9: Initialize maximum eigenvalue of matrix as λ_{max}
 - 10: $\lambda_{max} \leftarrow \frac{\sum_{i=1}^n \frac{M_{wi1}}{w_{i1}}}{n}$
 - 11: Initialize Consistency Index as CI
 - 12: Initialize Consistency Ratio as CR
 - 13: Initialize Random Index, $RI = 0.58$
 - 14: $CI \leftarrow \frac{\lambda_{max} - n}{n - 1}$
 - 15: $CR \leftarrow \frac{CI}{RI}$
 - 16: **IF** $CR < 0.1$ **then**
 - 17: Pairwise comparison matrix I_t is valid
 - 18: **Return** Criteria Weights, w
 - 19: **else**
 - 20: Change the values of Pairwise comparison matrix
-

Table 3.12: For N = 10, Random Inconsistency Indices (RI)

N	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.9	1.12	1.24	1.32	1.41	1.46	1.49

The Criteria weights, w we got are 0.688, 0.068, 0.244 which respectively represents the weights of Positive emotion, Negative emotion and Pearson Correlation score. To elaborate, the positive emotion and negative emotion have respectively 68.8% and 6.8% of total weights. Pearson correlation score has 24.4% of total weights.

Now the scores of each criteria is multiplied with their respective weights that has been found from to AHP algorithm by using the Equation 3.26 and we will get the weighted sum, W_{sum} .

$$W_{sum} = (H_{s_i} \times w_1) + (S_{s_i} \times w_2) + (P_{s_i} \times w_3) \quad (3.26)$$

where,

$i = 1,2,3 \dots m$ ($m =$ Number of videos)

$H_s \rightarrow$ Positive emotion score

$S_s \rightarrow$ Negative emotion score

$P_s \rightarrow$ Pearson correlation score

$w_1 \rightarrow$ Weight of Positive emotion score

$w_2 \rightarrow$ Weight of Negative emotion score

$w_3 \rightarrow$ Weight of Pearson Correlation score

Lastly, the process of ranking the videos based on the weighted sum, W_{sum} and the recommendation process is described in the Algorithm 4.

Algorithm 4 Ranking of the videos and Recommendation

Input:

AHP_{df} : A Dataframe having the importance values of Positive, Negative emotions and Pearson Scores

$Pearson_s$: A Dataframe that holds Pearson Correlation Scores

NLP_s : A Dataframe that holds Positive and Negative emotion scores of the videos

Output: V_{nl} : List of Recommended videos

- 1: Initialize a Dataframe to hold the merged dataframe as M_d
 - 2: $M_d \leftarrow Merge(Pearson_s, NLP_s)$
 - 3: Initialize a list as C_w
 - 4: $C_w \leftarrow AHP(AHP_{df}) \triangleright$ AHP() returns the corresponding weights of the criteria
 - 5: Initialize score of positive emotion as H_s
 - 6: Initialize score of negative emotion as S_s
 - 7: Initialize score of Pearson Correlation as P_s
 - 8: Initialize weighted Sum as W_s
 - 9: **for** each row of M_d **do**
 - 10: $W_{s_i} \leftarrow row((H_s \times C_w[0]) + (S_s \times C_w[1]) + (P_s \times C_w[2]))$
 - 11: **End for**
 - 12: Append W_s as a column to M_d
 - 13: $M_d \leftarrow$ Sort W_s in a descending order
 - 14: Initialize a list as V_{nl}
 - 15: $V_{nl} \leftarrow$ First five videos along with video names and links from M_d
 - 16: Sort the V_{nl} in a reverse order
 - 17: Recommend the list to users
 - 18: **Return** V_{nl}
-

Chapter 4

Result analysis

4.1 Performance Evaluation Metrics

The study that we have conducted on the SEED-V dataset for emotion classification has given significant results through extensive experiment. The results that have been obtained can be justified using the performance evaluation metrics, such as, F1-Score, Precision, Recall, Accuracy and Confusion Matrix. The score of these metrics portrays that the feature extraction techniques and the models that have been used are well grounded and effective.

Accuracy: The accuracy metric shows the percentage of emotion classes that have been correctly classified from all the test samples. The accuracy of the used models have been calculated using the Equation 4.1.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} = \frac{Correctly_Classified}{Total_Number_of_Samples} \quad (4.1)$$

Here,

T_P → True Positive

T_N → True Negative

F_P → False Positive

F_N → False Negative

Precision : The precision metric represents the percentage of emotion classes that have been classified correctly out of all the classification that was predicted to be on that particular emotion class. Using the Equation 4.2, the precision of the used models can be calculated.

$$Precision = \frac{T_P}{T_P + F_P} \quad (4.2)$$

Here,

T_P → True Positive

F_P → False Positive

Recall : The recall metric returns the percentage of emotion classes that have been classified correctly out of all the samples that are actually from that emotion class. The recall score of the used models have been calculated using the Equation 4.3.

$$Recall = \frac{T_P}{T_P + F_N} \quad (4.3)$$

Here,

T_P → True Positive

F_N → False Negative

F1-Score : F1-Score is the metric that combines the Precision and Recall score, representing an overall evaluation of the model. It returns the harmonic mean of Recall and Precision from which it can be evaluated how well the model performed for the dataset. The F1-Score of the used model have been calculated using the Equation 4.4

$$Recall = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4.4)$$

Confusion Matrix : Confusion matrix is a square matrix of (N x N) dimensions, where in multi-class classification, N is the number of classes present in the dataset. In this study, the confusion matrix has a dimension of (5 x 5) having 5 classes, namely Happy, Sad, Disgust, Neutral, Fear. In this matrix, 'X' axis represents the predicted emotion classes and 'Y' axis represents the actual or true emotion classes. From the confusion matrix, it can be seen how many of the classes are actually being predicted correctly.

4.2 Experimental Result Analysis

The Dataset is trained and tested on 324 features and 345 features respectively for Convolutional Neural Network (CNN) and Long Short-Term Memory(LSTM). The Table 4.1 shows the performance evaluation metrics results for both CNN and LSTM. It can be observed that both CNN and LSTM performed well for our dataset.

However, CNN has an upper hand in every metric, having almost 10% more score than the score of LSTM. It is clearly visible that CNN is the better model for our work, that means it can handle signal processing better than the LSTM. CNN has 93.01% of chances to accurately classify an emotion whereas LSTM can predict the correct emotion 82.23% of times. Similarly, the F1-Score of CNN is 12% more than that of LSTM. Moreover, the loss of CNN and LSTM are respectively 0.248, 0.416 where CNN has almost 40% less value than the other. The overall comparison of performance evaluation metrics can be visualized from Figure 4.1.

Table 4.1: Performance evaluation metrics result of CNN & LSTM

Model	Accuracy	Precision	Recall	F1-Score	Loss
Convolutional Neural Network (CNN)	93.01%	93.0%	92.0%	92.0%	0.248
Long Short-Term Memory(LSTM)	82.23%	81.0%	81.0%	80.04%	0.416

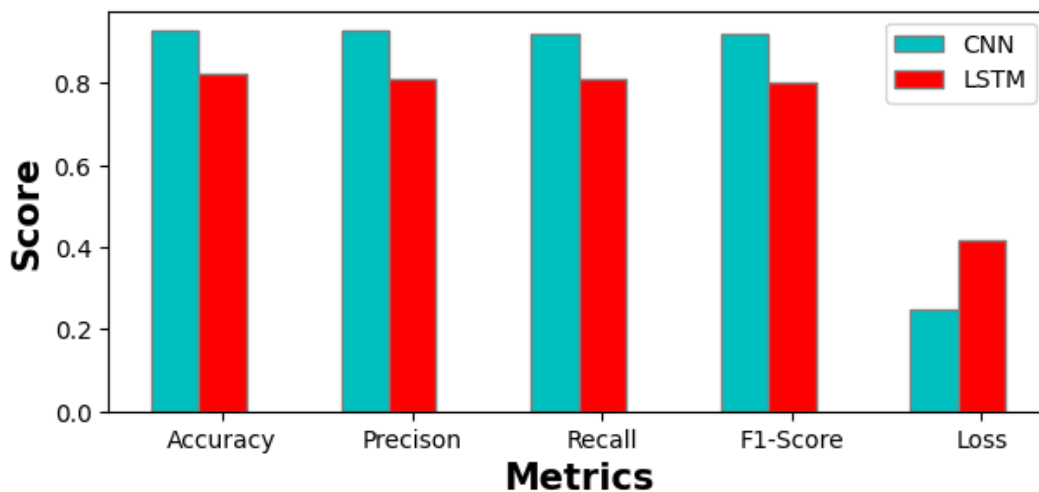


Figure 4.1: Comparisons between CNN and LSTM of the score of performance evaluation metrics

From the Table 4.2 we can observe the individual evaluation metrics score of each emotion label of CNN model. The CNN model could predict the emotion Fear, Sad and Neutral with a confident score that can be noticed by observing the F1-Score. The F1-Score of Fear and Sad emotion is 98% whereas Neutral emotion has a F1-Score of 99% which clearly states that these emotions are being classified fairly with a good margin. However, the model had a little rough time predicting the Disgust and Happy emotion as their F1-score is 79% and 86% in order.

Table 4.2: CNN: Evaluation Metrics values for each emotion label

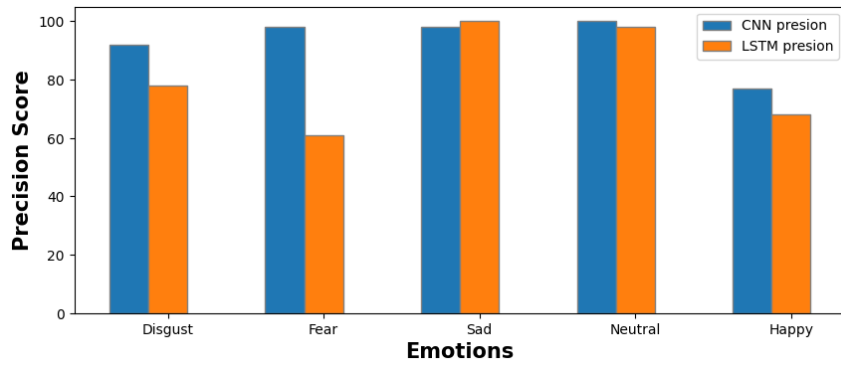
Emotion	Precision	Recall	F1-Score
Disgust	92%	70%	79%
Fear	98%	97%	98%
Sad	98%	98%	98%
Neutral	100%	97%	99%
Happy	77%	98%	86%

Similarly, the LSTM model’s individual evaluation metrics score of each emotion label can be observed from Table 4.3. The model can predict the Sad and Neutral emotion almost perfectly and has a F1-Score of 96% and 98% respectively. Even though the model can predict the Disgust emotion moderately, it shows a very decreasing result in terms of classifying Fear and Happy emotion with a F1-score of 66% and 57%.

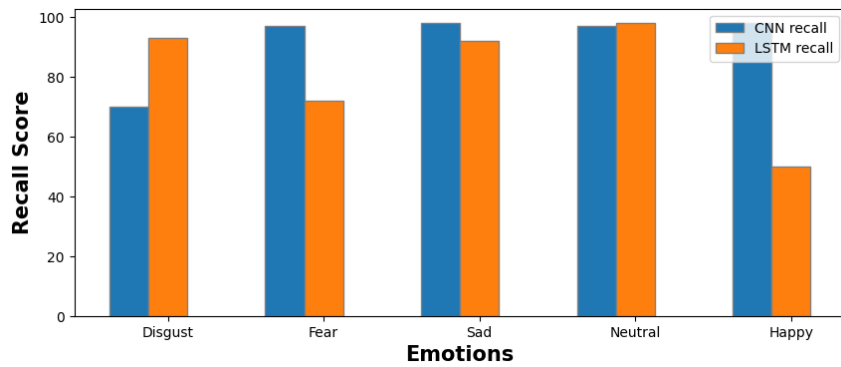
Table 4.3: LSTM: Evaluation Metrics values for each emotion label

Emotion	Precision	Recall	F1-Score
Disgust	78%	93%	85%
Fear	61%	72%	66%
Sad	100%	92%	96%
Neutral	98%	98%	98%
Happy	68%	50%	57%

A comparison of each label in terms of Precision, Recall and F1-Score can be visualized from Figure 4.2. From the Figure 4.2a we can see that for CNN model, 4 out of 5 emotions have higher precision score than of LSTM model. Only the Sad emotion has a higher precision score in LSTM model. Moreover, as for Recall scores, from the Figure 4.2b, it is noticeable that CNN model performed moderately better than the LSTM model. The LSTM model’s recall score of Disgust and Neutral emotion is higher than the score of CNN model. As for the concern of the other three emotions which are Fear, Sad and Happy, performed better under CNN model. Lastly, it is clearly visible from the Figure 4.2c that in terms of F1-Score, CNN model outperformed the LSTM model as 4 out of 5 emotions (Fear, Sad, Neutral, Happy) has a higher F1-score in CNN model. Therefore, it can be concluded that CNN model is better suited for our study than the LSTM model.



(a) Comparison of Precision Scores



(b) Comparison of Recall Scores



(c) Comparison of F1-Scores

Figure 4.2: Comparisons between CNN and LSTM of Precision, Recall and F1-Score of each label

Accuracy and loss curve of CNN and LSTM model can be observed from the Figure 4.3 and Figure 4.4 respectively. In both of the figures, for training and validation data, the accuracy curve is going upwards and the loss curve is going downwards as the number of epochs are increasing. For the CNN model in the Figure 4.3 lines for validation and training data going upwards and downwards almost smoothly. However, for LSTM model that is not the case, ups and downs are more frequent, diverging from reaching to the optimal goal.

Moreover, the validation data performed well than the training data because significant number of dropout layers were added while the data was in training phase to avoid the over fitting problem. However, during validation, all neurons are used. Hence, the model is more robust while validating and lead to a higher validation accuracy.

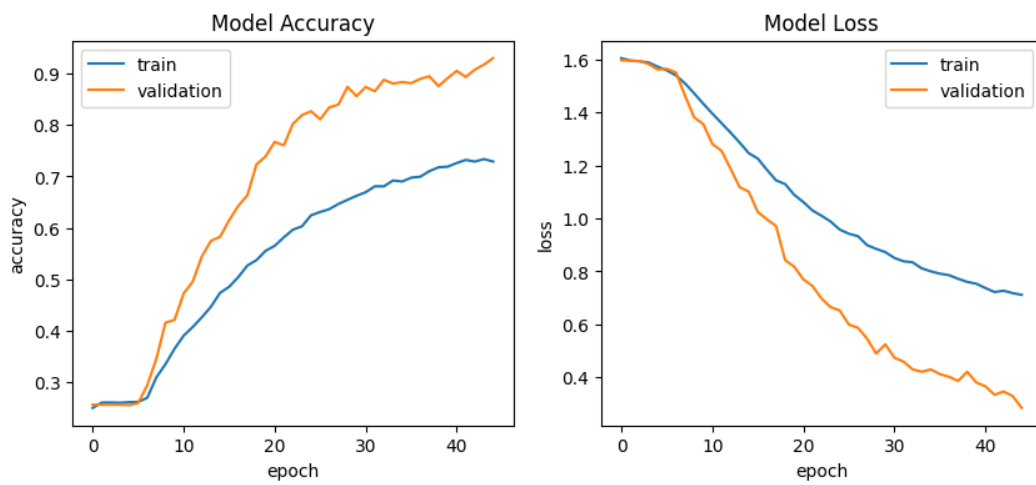


Figure 4.3: CNN Accuracy and Loss Curve

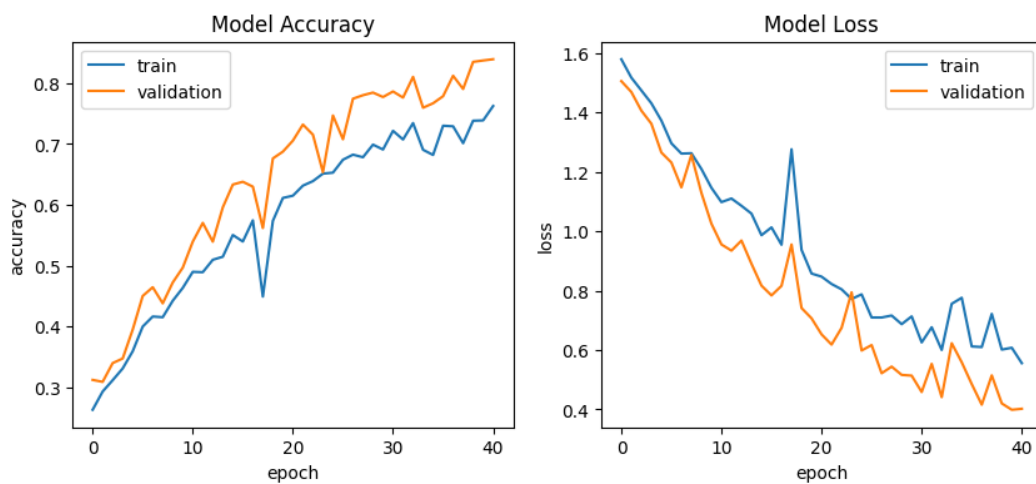


Figure 4.4: LSTM Accuracy and Loss Curve

Furthermore, Figure 4.5 is the confusion matrix of CNN and LSTM models respectively from which we can observe the numeric amount of both correctly and incorrectly predicted emotions. The values that are diagonally placed in the confusion matrix represent the correctly predicted emotions.

In addition, both the models performed well in terms of predicting the emotions. However, the CNN model has predicted more accurately than the LSTM model. For both the models, there are numerous false predictions too. However, for the CNN model, the ratio of incorrectly predicted emotions to correctly predicted emotion is very low for the Fear, Sad, Neutral, Happy Emotion, which is negligible. The case is not the same for the LSTM model as the ratio of incorrectly predicted emotion is much higher than the CNN model. The LSTM model has falsely classified some of the emotions as the Disgust emotion. Similarly, more than 50% of the testing data was classified as the Disgust and Fear emotion whereas the actual label of the emotion was Happy.

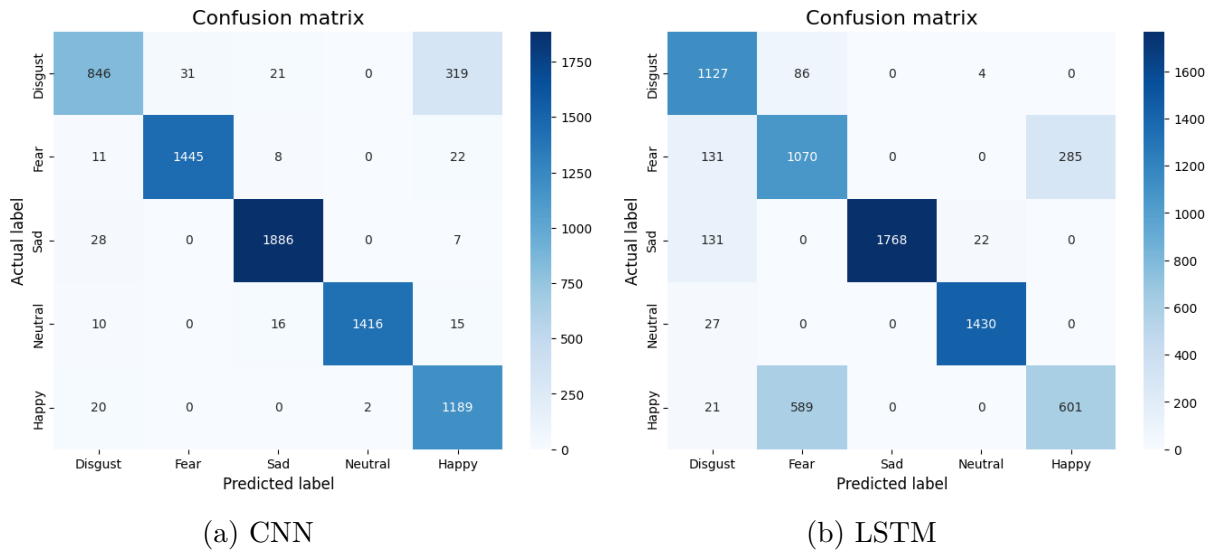


Figure 4.5: Confusion Matrix

Lastly, as for the Recommendation part, in the subsection 3.7.4 we found the Criteria weights, w which are 0.688, 0.068, 0.244 that which represents the weights of Positive emotion, Negative emotion and Pearson Correlation score each in order. In other words, 68.8% and 6.8% of total weights are allocated for the positive emotion and negative emotion and Pearson correlation score has 24.4% of total weights.

Now the weights of the corresponding criteria will be multiplied with the scores of each criteria using the Equation 3.26 and we will get the weighted sum, W_{sum} . We sort the dataframe on the basis of W_{sum} in a descending order so that the video that has the higher rating by the algorithm comes on top and rank themselves accordingly.

The first five videos will be taken into consideration in order to recommend them to the users. The higher the score of W_{sum} , the most recommended the video is.

The video that has the highest score will have the number one rank position and the rest of the ranking of the videos will be done accordingly based on the score. However, the highest scored video can not instantly lighten up a person's mood. The change of mood has to be done gradually. Therefore, after sorting the videos in a descending order based on W_{sum} , rank 1 to 5 will be selected for recommendation and the selected five videos will be recommended to users in a reverse order with the intent of lightening up the mood gradually. The scenario can be visualized from the Figure 4.6.

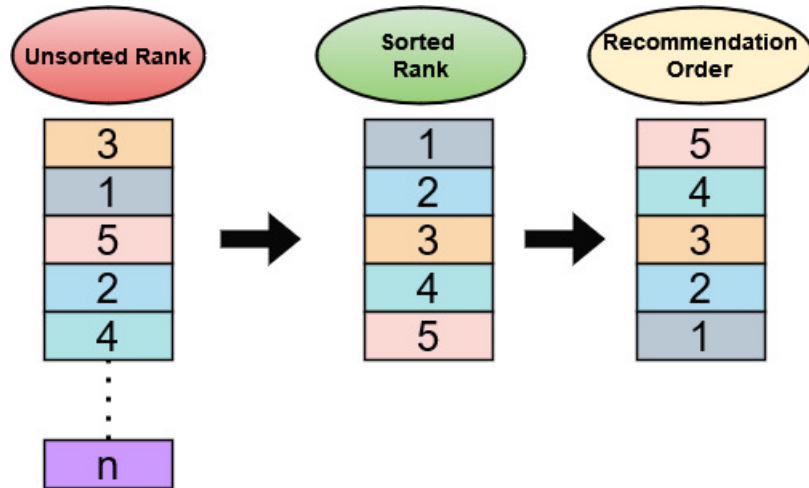


Figure 4.6: The order of recommendation

Chapter 5

Conclusion

As we continue to incorporate technology in every aspect of our lives, our system intends to incorporate this potential to address concerning issues, such as the elevation of the number of cases of depression. Despite the multitude of entertainment options available, individuals grapple with idleness or feel monotonous, leading to emotional distress. To eradicate these, in order to stabilize an individual's mood, we intend to incorporate this system, which aids in categorizing an individual's mood upon careful analysis and detection, and then work in favor of stabilizing the mood by recommending Feel Good contents.

In order to ensure proper functioning of the recommendation system, the emotions need to be detected accurately. For this, we employed EEG signals which provide higher reliability compared to other trending methods, like facial recognition, pitch in voice tones, etc. We extracted four different features namely (i) Differential Entropy (ii) Wavelet Energy (iii) Shannon Entropy and (iv) Eye movement features. After fusing these features, we explored both machine learning and deep learning models in order to classify an emotion. However, the deep learning models CNN and LSTM were better suited for our dataset. Moreover, we have also shown a detailed comparison between the models where it was clear that for our study CNN is the most suited model with an accuracy of 93.01%.

After successful categorization of the mood of an individual by identifying the model which generates the highest score of accuracy, a recommendation system was built by combining the techniques of text classification and Pearson Correlation. Score of emotion parameters such as Joy, Love, Sad, Fear etc. were extracted by classifying the subtitles of the Youtube videos and combined with the score of Pearson Correlation. Later, using the Analytic Hierarchy Process (AHP), the Youtube videos were ranked and recommended to the users accordingly. In summary, this research aspires to identify emotional instability by employing EEG signals and provide tailored content recommendations to stabilize moods and lift spirits gradually.

5.1 Challenges

Firstly, it was very challenging to understand the millions of the signal data with numeric value as there were approximately 156 millions of data samples in the whole dataset. Therefore understanding these data and finding the best approach to extract the features was time consuming. Moreover, after experimenting with various Machine Learning models like SVM, KNN, Random Forest, Decision Tree etc. we found the correct combination of the right model. Experimenting with these models required a lot of time as the process was resource intensive and required a strong setup. This issue was solved by using the Brac University Thesis Lab. Secondly, for the recommendation part, the dataset consisting of Youtube Videos with user ratings and subtitles was not available. Thus, personalized dataset had to be made with user ratings and had to extract subtitles from the youtube videos manually.

5.2 Limitations

In this study, for the emotion classification we used the State-of-the-art deep learning models. Hence, the concept of a hybrid model to detect emotion has not been addressed. Furthermore, the theory of early feature fusion type has not been discussed properly in this study. Moreover, the Pearson correlation score that has been obtained is based on a small amount of data and only has the rating of 10 users. Therefore, some of the Youtube Videos may have biased correlation. Another limitation of this study is that the extraction of emotion from the subtitles of the videos was done by using a pretrained model from the Hugging Face library.

5.3 Future Work

In future, an attempt will be made to increase the accuracy score by tuning the CNN model. We will also try different deep learning models and a thorough comparison on the evaluation metrics will be done. We will also make an attempt to extract different features and combine them with the existing features. Another comparison based study will be conducted on the newly extracted features. Moreover, we intend to work with our recommendation system by combining new techniques with the existing ones. In addition, we want to build our own text-classification model to train our dataset of subtitles of Youtube videos. Additionally, collecting a big chunk of data is also a priority so that the recommending system can get more robust.

Bibliography

- [1] P. R. Kleinginna and A. M. Kleinginna, “A categorized list of emotion definitions, with suggestions for a consensual definition,” *Motivation and emotion*, vol. 5, no. 4, pp. 345–379, 1981.
- [2] P. Ekman, R. W. Levenson, and W. V. Friesen, “Autonomic nervous system activity distinguishes among emotions,” *science*, vol. 221, no. 4616, pp. 1208–1210, 1983.
- [3] M. M. Bradley and P. J. Lang, “International affective digitized sounds (iads): Stimuli, instruction manual and affective ratings (tech. rep. no. b-2),” *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*, 1999.
- [4] A. Choppin, “Eeg-based human interface for disabled individuals: Emotion expression with neural networks,” *Unpublished master’s thesis*, 2000.
- [5] T. Partala, M. Jokiniemi, and V. Surakka, “Pupillary responses to emotionally provocative stimuli,” in *Proceedings of the 2000 symposium on Eye tracking research & applications*, 2000, pp. 123–129.
- [6] R. W. Picard, “Toward computers that recognize and respond to user emotion,” *IBM systems journal*, vol. 39, no. 3.4, pp. 705–719, 2000.
- [7] J. A. Coan, J. J. Allen, and E. Harmon-Jones, “Voluntary facial expression and hemispheric asymmetry over the frontal cortex,” *Psychophysiology*, vol. 38, no. 6, pp. 912–925, 2001.
- [8] K. Ishino and M. Hagiwara, “A feeling estimation system using a simple electroencephalograph,” in *SMC’03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, IEEE, vol. 5, 2003, pp. 4204–4209.
- [9] C. Niemic, “Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion.,” 2004.
- [10] K. Takahashi *et al.*, “Remarks on emotion recognition from bio-potential signals,” in *2nd International conference on Autonomous Robots and Agents*, Citeseer, vol. 3, 2004, pp. 1148–1153.
- [11] B. Zhou, S. C. Hui, and K. Chang, “An intelligent recommender system using sequential web access patterns,” in *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, IEEE, vol. 1, 2004, pp. 393–398.
- [12] P. J. Lang, M. M. Bradley, B. N. Cuthbert, *et al.*, *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention Gainesville, FL, 2005.

- [13] D. O. Bos *et al.*, “Eeg-based emotion recognition,” *The influence of visual and auditory stimuli*, vol. 56, no. 3, pp. 1–17, 2006.
- [14] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, “Emotion assessment: Arousal evaluation using eeg’s and peripheral physiological signals,” in *International workshop on multimedia content representation, classification and security*, Springer, 2006, pp. 530–537.
- [15] C.-F. Chen, “Applying the analytical hierarchy process (ahp) approach to convention site selection,” *Journal of travel research*, vol. 45, no. 2, pp. 167–174, 2006.
- [16] T. Pun, T. I. Alecu, G. Chanel, J. Kronegg, and S. Voloshynovskiy, “Brain-computer interaction research at the computer vision and multimedia laboratory, university of geneva,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 210–213, 2006.
- [17] O. Rosso, M. Martin, A. Figliola, K. Keller, and A. Plastino, “Eeg analysis using wavelet-based information tools,” *Journal of neuroscience methods*, vol. 153, no. 2, pp. 163–182, 2006.
- [18] G. Chanel, K. Ansari-Asl, and T. Pun, “Valence-arousal evaluation using physiological signals in an emotion recall paradigm,” in *2007 IEEE International Conference on Systems, Man and Cybernetics*, IEEE, 2007, pp. 2662–2667.
- [19] A. Heraz, R. Razaki, and C. Frasson, “Using machine learning to predict learner emotional state from brainwaves,” in *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, IEEE, 2007, pp. 853–857.
- [20] R. Horlings, D. Datcu, and L. J. Rothkrantz, “Emotion recognition using brain activity,” in *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing*, 2008, pp. II–1.
- [21] J. Kim and E. André, “Emotion recognition based on physiological changes in music listening,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
- [22] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” *International Journal of Human-Computer Studies*, vol. 67, no. 8, pp. 607–627, 2009.
- [23] 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, pp. 865–870, 2009.
- [24] H. Liu, J. Hu, and M. Rauterberg, “Music playlist recommendation based on user heartbeat and music preference,” in *2009 International Conference on Computer Technology and Development*, IEEE, vol. 1, 2009, pp. 545–549.
- [25] Q. Zhang and M. Lee, “Analysis of positive and negative emotions in natural scene using brain activity and gist,” *Neurocomputing*, vol. 72, no. 4-6, pp. 1302–1306, 2009.
- [26] Y.-P. Lin, C.-H. Wang, T.-P. Jung, *et al.*, “Eeg-based emotion recognition in music listening,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

- [27] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to pattern recognition: a matlab approach*. Academic Press, 2010.
- [28] X.-W. Wang, D. Nie, and B.-L. Lu, “Eeg-based emotion recognition using frequency domain features and support vector machines,” in *Neural Information Processing: 18th International Conference, ICONIP 2011, Shanghai, China, November 13-17, 2011, Proceedings, Part I 18*, Springer, 2011, pp. 734–743.
- [29] D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, “Asymmetric spatial pattern for eeg-based emotion detection,” in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2012, pp. 1–7.
- [30] K. Yoon, J. Lee, and M.-U. Kim, “Music recommendation system using emotion triggering low-level features,” *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 612–618, 2012.
- [31] P. Nagarnaik and A. Thomas, “Survey on recommendation system methods,” in *2015 2nd international conference on electronics and communication systems (ICECS)*, IEEE, 2015, pp. 1603–1608.
- [32] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, “Music recommendation system based on user’s sentiments extracted from social networks,” *IEEE Transactions on Consumer Electronics*, vol. 61, no. 3, pp. 359–367, 2015.
- [33] H. Koochi and K. Kiani, “User based collaborative filtering using fuzzy c-means,” *Measurement*, vol. 91, pp. 134–139, 2016.
- [34] C. M. Rodrigues, S. Rathi, and G. Patil, “An efficient system using item & user-based cf techniques to improve recommendation,” in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, IEEE, 2016, pp. 569–574.
- [35] Y. Seanglidet, B. S. Lee, and C. K. Yeo, “Mood prediction from facial video with music “therapy” on a smartphone,” in *2016 wireless telecommunications symposium (wts)*, IEEE, 2016, pp. 1–5.
- [36] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, “Emotion recognition based on eeg using lstm recurrent neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.
- [37] 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.
- [38] A. Al-Nafjan, M. Hosny, A. Al-Wabil, and Y. Al-Ohali, “Classification of human emotions from electroencephalogram (eeg) signal using deep neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 9, 2017.
- [39] D. Ayata, Y. Yaslan, and M. E. Kamasak, “Emotion based music recommendation system using wearable physiological sensors,” *IEEE transactions on consumer electronics*, vol. 64, no. 2, pp. 196–203, 2018.
- [40] O. Bazgir, Z. Mohammadi, and S. A. H. Habibi, “Emotion recognition with machine learning using eeg signals,” in *2018 25th national and 3rd international iranian conference on biomedical engineering (ICBME)*, IEEE, 2018, pp. 1–5.

- [41] P. Kumar and R. S. Thakur, “Recommendation system techniques and related issues: A survey,” *International Journal of Information Technology*, vol. 10, pp. 495–501, 2018.
- [42] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, “Amigos: A dataset for affect, personality and mood research on individuals and groups,” *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2018.
- [43] J. Papathanasiou, N. Ploskas, *et al.*, “Multiple criteria decision aid,” *Methods, examples and python implementations*, vol. 136, p. 131, 2018.
- [44] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, “Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos),” *IEEE Access*, vol. 7, pp. 57–67, 2018.
- [45] H. Zamanian and H. Farsi, “A new feature extraction method to improve emotion detection using eeg signals,” *ELCVIA: electronic letters on computer vision and image analysis*, vol. 17, no. 1, pp. 29–44, 2018.
- [46] Y. Zhang, S. Zhang, and X. Ji, “Eeg-based classification of emotions using empirical mode decomposition and autoregressive model,” *Multimedia Tools and Applications*, vol. 77, no. 20, pp. 26 697–26 710, 2018.
- [47] D.-W. Chen, R. Miao, W.-Q. Yang, *et al.*, “A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition,” *Sensors*, vol. 19, no. 7, p. 1631, 2019.
- [48] R. Alhalaseh and S. Alasasfeh, “Machine-learning-based emotion recognition system using eeg signals,” *Computers*, vol. 9, no. 4, p. 95, 2020.
- [49] M. R. Islam, M. A. Moni, M. M. Islam, *et al.*, “Emotion recognition from eeg signal focusing on deep learning and shallow learning techniques,” *IEEE Access*, vol. 9, pp. 94 601–94 624, 2021.
- [50] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, “Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition,” *IEEE Transactions on Cognitive and Developmental Systems*, 2021.