# Analysis of EEG Signals for Emotion Recognition using Deep Features


Author

Sitwat Mahd

00000205102

Supervisor

Dr. Muhammad Usman Akram

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD


MAY, 2021

# Analysis of EEG Signals for Emotion Recognition using Deep Features


Author

Sitwat Mahd

00000205102


A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor

Dr. Muhammad Usman Akram

Thesis Supervisor's Signature: ________________________________________


DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD


MAY, 2021

# Declaration

I certify that this research work titled "*Analysis of EEG Signals for Emotion Recognition using Deep Features"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Sitwat Mahd

00000205102

# Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Sitwat Mahd

00000205102

Signature of Supervisor

Dr.Muhammad Usman Akram

# Copyright Statement

# Acknowledgements

I bow my head in gratitude infront of Allah Almighty who is the best planner. His blessings gave me the strength, patience and ability to carryout the research and bring it to a satisfactory completion. I owe my all achievements to Him who is always there to guide us through thick and thin of life.

I would like to express my sincere gratitude to my advisor Dr. Muhammad Usman Akram for his always available guidance and untiring dedication which helped me in achieving this important milestone of my academic career. I consider it my good luck to have such an outstanding professional as my advisor and mentor for my research.

Along with my advisor, I would like to acknowledge my entire thesis committee : Dr. Sajid Gul Khawaja, Dr. Arslan Shaukat and Dr. Shahzor Ahmed for their valuable guidance and valuable recommendations.

My acknowledgement would be incomplete without thanking my family who have been always there for me.

*Dedicated to my beloved parents: **Dr. Abdul Khaliq & Afshan Khaliq**, and my family members who have always been there for me and are like a lighthouse in my life*

# Abstract

Recognition of human emotional states using physiological signals has become progressively popular area of research. The motivation behind the use of physiological signals such as Electrocardiogram (ECG), Electroencephalogram (EEG) and Galvanic Skin Response (GSR) for brain computer interfaces is due to the unforgeable nature of physiological signals which enhances the reliability of the system meant for people having motor disabilities. The use of deep learning models for emotion recognition using EEG signals has proved to be very useful despite the associated problems of EEG signals such as low SNR, high randomness and non-stationary nature.

Most of the emotion classification approaches involve complex signal processing techniques for extraction of features and hence underutilize the capabilities of neural networks to extract meaningful features from the raw data. Most of the approaches do not take into consideration the spatial correlation of EEG signals. Motivated by the remarkable performance of deep learning approaches, in this research, a deep learning method employing 2D CNN is proposed in which the input to CNN is formed based on spatial as well as temporal information of EEG signals and has proved to be very effective in classification of emotions. This research is focused on classifying the emotions based on valence-arousal model into 2 classes (high, low), 3 classes (high, neutral, low) and 4 classes (LVLA, LVHA, HVLA, and HVHA) using sliding window approach. The proposed approach has achieved a significant accuracy of 89.1% and 90.4% for 2 classes of Valence and Arousal dimensions respectively and 88.2% for 4 classes on publically available dataset AMIGOS. The classification task of four classes is also tested using pretrained network Alexnet and the accuracy in this case is

**Key Words:** *Electroencephalogram* **(***EEG), convolutional neural networks, deep learning, emotion recognition*

# Table of Contents

# List of Figures

# List of Tables

# CHAPTER 1: INTRODUCTION

Use of physiological signals for recognition of human emotion has progressively become a budding and rising trend in the domain of human computer interface. The research in this area is fueled by the impartial nature of the physiological signals produced by the human nervous system. A wide range of feelings, thoughts and resulting behaviors of human beings are linked with emotions. Therefore, these emotions affect our capability to act logically once we are involved in a process related to decision making, perception and human intelligence. As a result, researches in the field of emotion recognition by utilizing the emotional input use the brain computer interface systems as an efficient tool for applications in the fields of medical science and human interactions. For decades, Brain computer interface (BCI) has been one of the most important and attractive field for research in biomedical engineering. An important application of the brain computer interface is for the people having motor disabilities. The Brain computer interface (BCI) provides a way for such people to interact directly with outside world or to effectively utilize instruments using brain intentions alone. It offers a promising technology that enables human beings to use their brain waves for controlling the external devices. Focus of affective computing domain is to enable the machines to read and respond to human emotions and to further create a vivid communication between machines that are emotionally abundant and emotionally absent. One of the key criteria for the realization of an intelligent emotion interaction is to build a reliable emotion recognition system with high precision and robustness as well as reasonably good applicability. Since 2015, deep learning (DL) has achieved exceptional growth and has proven to be very successful in many complex projects. Biggest advantage of Convolutional Neural Network (CNN) is that it can directly extract the features from the raw data without using any feature extraction technique which has been validated in many difficult categorization projects. In recent years, various EEG emotion detection domain problems are being resolved by employing CNN based techniques.

## 1.1 Motivation

These days, modern ways of human centered and human-driven engagement with digital media have the potential to revolutionize the entertainment, learning, living and other fields of life. As human life is governed by emotions, the increasing role of human computer interface applications demands automatic recognition of emotions. For this purpose, human text, speech, facial expressions and gestures are being used for recognition of emotions. However, recently, emotion recognition using EEG is a more focused area of research. Emotions are important for our daily lives and work. Monitoring, evaluation and controlling of emotions in real time can change the lives of people for betterment. In case of interaction between human and machine, using the emotion recognition can make the processes simple and natural. As per trends, medical applications utilize the technology based on EEG. Main purpose of any medical treatment is to identify and mitigate pain of the patient. However, identification of the pain becomes more difficult if the patient is not in a state in which he can tell about his medical condition. Few examples of such situation are people in coma, infants and sedated patients. Hence it becomes essential to find out alternate means to detect and evaluate a medical condition even when the patient is unable to explain his problem. In order to handle such cases, introducing a systematic process capable of quantifying the patient's pain would result in providing a better quality of life for human beings. It will enable the medical practitioners to provide a more customized treatment as per condition of the patient. People having mental and emotional disturbance face anxiety, depression, disorders related to eating, conduct and attention deficit. As a result of these problems related to mental health, these people become socially isolated, fail in academic and professional domains, get addicted to drugs, have urge to commit suicide and suffer from many other health and social problems. Treatment of these patients becomes more problematic if they are unable to express their issues. Therefore, studying emotion state of these patients in real time will reveal a better picture about their medical condition and will help the doctors to provide better health care. Utilizing EEG signals for emotion recognition has been in focus in recent years. It is an essential part of the brain computer interface (BCI) system which efficiently enhances the communication between human and machines.

## 1.2 Problem Statement

A wide range of applications of BCI systems are found in the field of biomedicine. Patients suffering from disease such as schizophrenia, depression, autism, stroke are unable to communicate with the people around and express their internal feelings and emotions. Number of stroke patients are increasing at an accelerated rate. In order to assist the patients with communication, the primary requirement is to design an emotion recognition system which has high reliability, robustness, good accuracy and effortless applicability. Numerous emotion recognition approaches based on physical as well as physiological signals have been proposed but EEG based approaches are gaining popularity because of its non-invasive, low-cost acquisition and high temporal resolution. Despite these favorable attributes, EEG signals have low signal to noise (SNR) ratio, high randomness, may convey poor spatial resolution and have high dimensionality. These challenges make emotion recognition from EEG signals a tedious task. The aim of this research work is to discover deep learning methods for making the task of emotion recognition a reliable and robust one.

## 1.3 Aims and Objectives

Major objectives of the research are as follow:

- Review and comparison of recent developments in emotion recognition systems.
- To analyze EEG signals for possible emotion detection.
- To see usefulness of deep learning for this application.
- To use deep features along with other classifiers for robust segregation of different emotions
- To utilize developed algorithms for emotion charting.

## 1.4 Structure of Thesis

This work is structured as follows:

**Chapter 2** covers the anatomy of brain, emotion models and the importance of using physiological signals for emotion recognition. It further discusses the EEG bands and the associated brain states.

**Chapter 3** gives review of the literature and the significant work done by researchers in past few years related to emotion recognition.

**Chapter 4** consists of the proposed methodology in detail. It includes two main modules: Feature extraction techniques followed by classification models.

**Chapter 5** introduces the database used for evaluation purposes. All the experimental results are discussed in detail with all desired figures and tables.

**Chapter 6** concludes the thesis and reveals future scope of this research

# CHAPTER 2: EMOTION MODELS AND EEG

Brain-Computer interface (BCI) is one of the most amazing research domain in the field of biomedical engineering. It is a system which gets brain's signal as input and depicts the real-time data of that person's cognitive state [23]. Brain-Computer interface gives human beings a rewarding technological method to control the external gadgets by varying their brain waves. These applications are designed for reading the brain waves without introducing instruments into the body, thus making them more practical and easy to implement. Numerous BCI systems are in use world over which are based on EEG such as identification of spellings for words [24] and for driving the external machines like wheelchair [3]. Brain Computer Interface can be used to control the devices mentally and for studying our mental states. One such application is emotion recognition. Using algorithms for automated recognition of emotions can result in bridging the gaps between the human and machine relationship. With the advancement of technology, research related to BCI domain has been utilized to benefit people with and without physical and motor disabilities. The exponential increase in the research related to BCI domain has been fueled by the interest of researchers from multiple fields of biomedical engineering, computer engineering, neurology, rehabilitation engineering, mathematics, clinical psychology etc.

As the BCI research is applicable on multiple fields, it has shown progress in the development of BCI systems with varying end applications. During research on BCI, it has been learnt that activity of brain can be utilized either in active or passive control modes. While using BCI in active control mode, the device is controlled by the user through brain signals utilizing direct and conscious production of instructions which are transmitted to the external applications. Whereas, in passive BCI systems, the output is generated by the brain signals without any willful command. Examples of affective and cognitive input in passive BCIs are emotional states like stages of meditation, motivation, anger, enthusiasm and stress.

Principle of passive Brain Computer Interface has been employed in many domains like affective computing to understand the emotions of human beings inorder to enhance the communication between humans and machines. Subsequently, this information is utilized for development of applications which can respond to the variations in the human mental state, therefore, optimizing the output of the interaction.

Aim of the complete process of the communication is to enhance the interaction resulting into a natural and efficient user experience. Motivation about learning new functions of brain and progress in the development of BCI devices has encouraged several researchers to carryout studies which focus on the real time recognition of emotions for the patients. Moreover, for the other cognitive mental states like concentration and workload, similar advancements are underway. These and many other developments in BCI technologies have sparked the motivation in the scientists in various contexts. Although Brain Computer Interface covers a large range of applied areas, the scope of this research is narrowed down to emotions recognition through Electroencephalography (EEG) [23].

## 2.1 Brain

Emotions are a complicated phenomena. These emotions regulate and influence human behaviors as group of biological, social and cognitive components [5]. Emotions are generated in various parts of the brain which work together and are processed by a network of brain areas. Amygdala, prefrontal cortex, cingulated cortex, hippocampus and the basal ganglia are some the areas of the brain which are involved in the emotion stimulation [6]. Each part has its own task in controlling the emotions. Amygdala is a small part of the brain which has shape and size of an almond and manages the positive and negative information. It is involved when a person faces the emotion of fear. Similarly, prefrontal cortex is positioned in front of the brain. It is the control center which directs our actions and is involved in emotion regulation process. Different parts of the brain stay connected and communicate frequently. For example, Amygdala also known as the emotion center senses emotion of fear and informs the prefrontal cortex which is also known as the control center. Prefrontal cortex gets the input that some fearful event is happening. If required, this control center at the front of the head sends signal to other brain parts informing them to move the body away from the danger / fear. Multiple brain parts work together to process and respond to a situation.

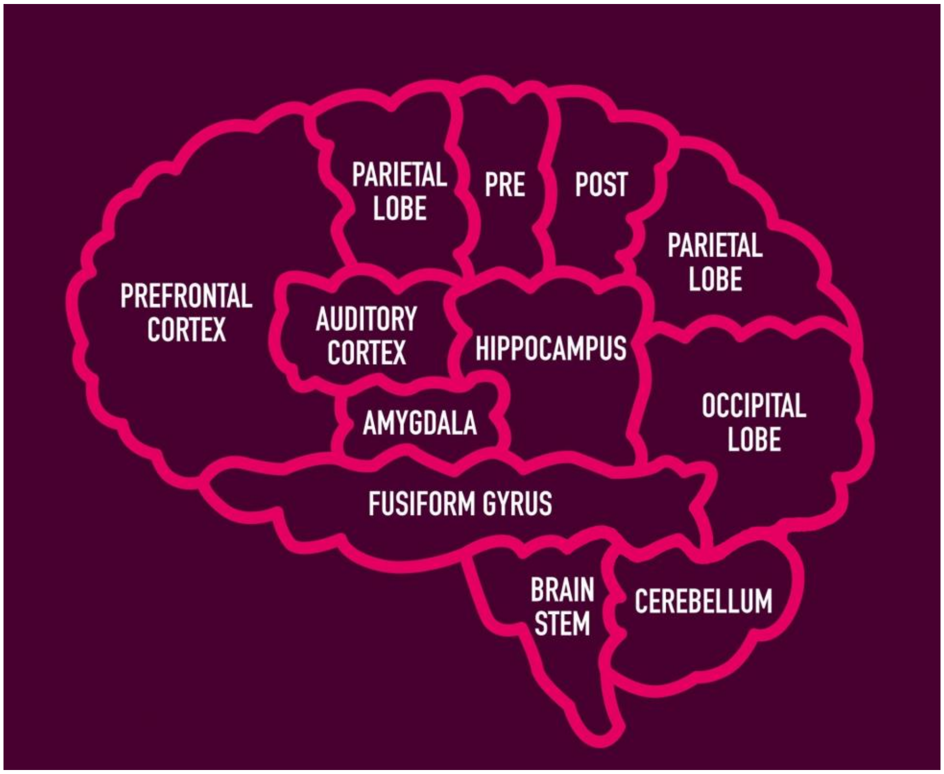



**Figure 2.1:** Brain Structure

## 2.2 Anatomy of Brain

Human brain is a complicated and important part of the body. External stimuli can evoke emotions in the brain. Emotions are the primary and natural responses of the brain to external stimuli [7]. Brain is divided into three main regions namely forebrain, mid brain and hind brain. Every region is further divided into further parts and each part is responsible for managing a particular body function.

- **Hind brain**. It consists of medulla oblongata, the pons and the cerebellum. Its main function is to coordinate vital functions of the human body like respiration, heart activity, sleep, wakefulness etc.
- The **mid brain** the narrow part of the brain that connects fore brain to the hind brain. The main function of the mid brain contributed in the control of movements of eyes and also processes auditory and visual input to the brain.
- The **fore brain** is sub divided into the diencephalon and the cerebrum.

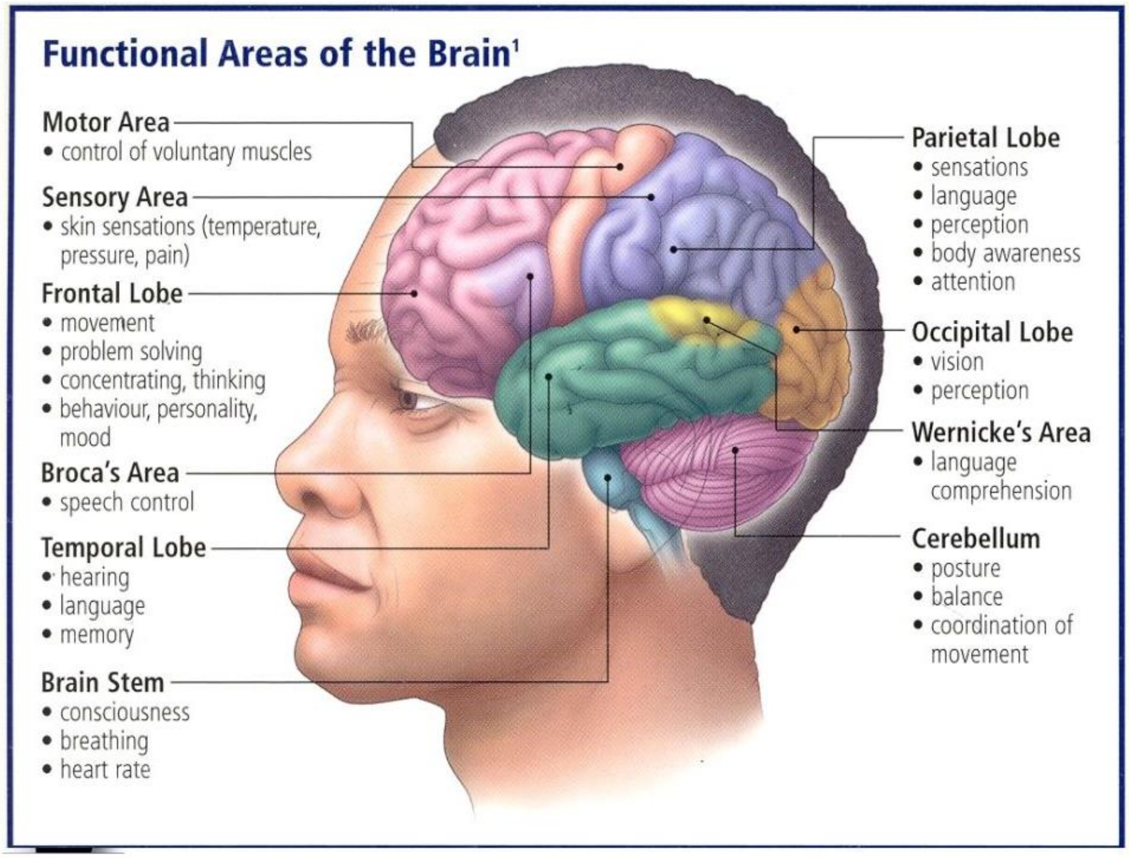



**Figure 2.2:** Anatomy of Brain

### 2.2.1 Cerebrum

It is the part of the brain which works to make human beings unique. It is the external most part of the brain. The raised portions of the brain are called gyrus and the groovy portion is called sulcus. Different capabilities of human brain like intelligence, communication, awareness, the power to think, reasoning and visualization all these take birth in the cerebral cortex. It can be grouped in four lobes.

#### 2.2.1.1 Lobes of Cerebral Cortex

Cerebral Cortex is grouped into frontal lobe, parietal lobe, occipital lobe and temporal lobe. These four parts are linked with different tasks starting from sensing to the visualization.

- **Frontal Lobe**. It is situated on the front side of the human brain and is responsible for the planning, problem solving, apprehension, interpretation, motor skills, way of thinking, intelligence and communication. Motor cortex is located at its backside. Motor cortex gets input from multiple lobes and uses the received input to perform the body actions.

- **Parietal Lobe**. It is positioned in the center section of the brain and is linked with computing the physical sensory input for example, pressure, touch and pain. With parietal lobe functions, a person is able to judge that the two objects in contact with his skin are different. A part of the brain called as somatosensory cortex is inside the parietal lobe and is necessary for interpreting the senses.

- **Temporal Lobe**. It is located on the bottom of the brain and houses hippocampus as well. This part is also closely linked with the memory creation. Primary auditory cortex is also located here which is responsible for understanding the sounds and communication we receive. Therefore, any damage to this part can result in issues with memory and communication.

- **Occipital Lobe**. It is located in the back part of the brain and is linked with understanding the information received through the eyes. This information is then transmitted to multiple visual processing regions which calculate distance, location and identification of objects. Any damage to this lobe can result in issues related to disturbed balance system, visual system, identification of things and understanding of colors, words etc.

### 2.2.2 The Cerebellum

The cerebellum or little brain is located behind the brain stem. It consists of little lobes and takes input from the ear, nerves, eyes and converts it to make the body movements more precise. It is important for the coordinated movements as well as motor skills. Its size is ten percent of the total brain but comprises of more than fifty percent of total neurons in the brain.

### 2.2.3 The Limbic System

It comprises of five main parts which are amygdala, hippocampus, hypothalamus, septal area and parts of the limbic cortex. Connections are established by these parts between the hypothalamus, cerebral cortex and thalamus. Hippocampus plays important role in memory and gaining knowledge, whereas, the limbic system controls the reactions related to emotions. Amygdala controls our responses to fear. It is attached to other systems of human body systems which are associated with fear. Alongwith managing fear responses, it also helps us to train us from the fearful conditions by remembering the details of the event in order to stay away from it in future [8]. Hippocampus is related to storage of the details in long term memory. If there is any injury to the hippocampus, new memories cannot be stored. However, the memories prior to the injury will remain there [9].

#### The Thalamus

It is located above the brain stem. It calculates and sends the information about movement to the cerebral cortex which in turn also sends back the information. This input is then transmitted to other parts.

#### The Hypothalamus

It is a cluster of nuclei which is located at the base of the brain adjacent to the pituitary gland. It is connected with other parts of the brain and manages various functions of the body like circadium rhythm, controlling the temperature of the body and food requirement. Generation of hormones through pituitary glands is also controlled by the hypothalamus.

#### The Basal Ganglia

It is a combination of nuclei which encloses the thalamus to some extent. It is necessary for controlling of the movement. Basal ganglia takes input for the intended movement from the cerebral cortex and after processing sends it to thalamus which further transmit it back to cortex. The refined movement information is then sent to the body muscles [7].

## 2.3 Emotions

Don Hockenbury and Sandra E. Hockenbury in the book “Discovering Psychology” define emotion as a complicated psychological condition which comprises of three different parts: a subjective part, a physiological reaction and a behavioral or expressive reaction [11]. Emotions are generated as multicomponent responses which comprise of coordinated variations in subjective feelings, expressing through motor skills and physiology. For understanding human behaviours, emotions are very important and they have a very critical part in the everyday nonverbal communication. Emotions influences the physiological as well as psychological state and have a very important contribution in the lives of human beings. Human health has a positive effect of positive emotions, whereas, negative emotions result in causing health issues. If these negative emotions persist for a longer time, they can result into various diseases like anxiety, depression, heart problems and suicidal thoughts in extreme cases. Owing to complex nature of shared interaction of physiology and psychology in emotions, identification of human emotions accurately and timely is still limited to our knowledge and has been a topic of research for the last many years. In order to identify the human emotions, different techniques have been recommended in the literature. Human emotions can be identified by studying their face reactions, body movements, voice signals or by monitoring the activity of their brains. Generally, human emotions are identified as a result of an action that involves few human senses. Various kinds of emotions experienced by human beings have been categorized by the psychologists. For this purpose, many theories have been proposed to differentiate the human emotions and their bifurcation. Emotions are of various kinds and they affect our ways of living and interacting with other people. These emotions appear to be ruling us at some instances in our lives. The options we choose, the decisions we make and the thoughts which we have in our minds, all are affected by our emotions at that particular time. Researchers have gone an extra mile in defining the emotions by categorizing them into different classes. The explanation and in depth knowledge of these emotions has been changing over the years. Psychologist Paul Eckman, in year 1972 proposed that 06 fundamental emotions [12] are found in all human civilizations which are fear, anger, disgust, surprise, happiness and sadness.

Robert Plutchik, in 1980s presented a different categorization of emotions which was called as the “wheel of emotions”. It exhibited that various emotions can be intermixed in a similar way as a painter mixes the primary colors to make another color [13]. Eckman, in the year 1999 elaborated his categories by adding few more emotions which are [12] embarrassment, excitement, contempt, shame, pride, satisfaction and amusement.

### 2.3.1 Key Elements of Emotions

For understanding the emotions in a better way, three key elements namely subjective, physiological and behavioral responses are taken into consideration.

### The Subjective Experience

Researcher are of the opinion that human beings experience various emotions which are not specific to one culture and going through an emotion can be highly subjective [82]. If we take one emotion of anger as an example then this emotion cannot be similar for every person. It varies from mild anger to extreme rage. Although we divide the emotions into broader categories as “angry” or “sad” etc, however, for every person the level of these emotions experienced may be different, therefore, termed as subjective.

### The Physiological Response

Emotions can also result into physiological reactions. For example, palpitation, increased heartbeat, effect of anxiety causing stomach discomfort etc. As mentioned in Cannon-Bard theory of emotion, the feelings and physiological reactions occur at the same time. These physiological reactions during an emotion are controlled by the human nervous system. This nervous system governs these physiological reactions as it has the responsibility of regulating the fight or flight responses of the body. In case of a threat, the physiological reactions of the body make the body ready either for avoiding the situation or to face it. Initial researches on the physiology of the human emotions focused on the responses, however, the researchers have shifted their focus towards the role of brain in the human emotions. Study of brain scans revealed that during emotions and specially in fear, one part of the limbic system called the amygdala has a significant part [83]. Amygdala is a small structure having shape like an almond. It is attached to the states like hunger, thirst, memory and emotion. By analyzing images of brain, research has shown that amygdala is activated once human beings see frightening pictures. The amygdala has been linked to motivational states such as hunger and thirst as well as memory and emotion. People whose amygdala gets damaged face reduction in their reaction to the fear situation [84].

### The Behavioral Response

Behavioral response is the real expressive part of the emotional response. This can include many responses like smiling, gratitude, respect, annoyance, laughter mainly depends on the cultural and social norms of the respective society alongwith the personality traits of the individuals.In our daily life, we are all the time observing other’s expressions of these behavioural responses.

The extent of our ability to correctly interpret these responses depends on the emotional intelligence of every individual.

## 2.4 Emotion Models

Emotions are categorized into two distinct models:

(a) The Categorical model

(b) The Dimensional model

In the Categorical model, a title is compulsory to pick an emotion from a collection of emotions which represents the communicated feeling in the best manner. This model tags the emotions with label. Whereas, in dimensional model, the bifurcation is done by using multidimensional scaling, on the basis of a set of quantifiable measures [85]. Each of these models makes us understand a discrete characteristic of our emotions. Both these models give us the idea about how human mind presents and interprets the emotions. Human emotions status can be evaluated by these models in real time. In dimensional model, rating scales are used for every dimension by using tools such as Self-Assessment Manikin [86] or Feeltrace [87]. Self-Assessment Manikin (SAM) has images of manikins for assessing the static degree of a dimension at a given time. On the other side, Feeltrace monitors the emotions related information with time.

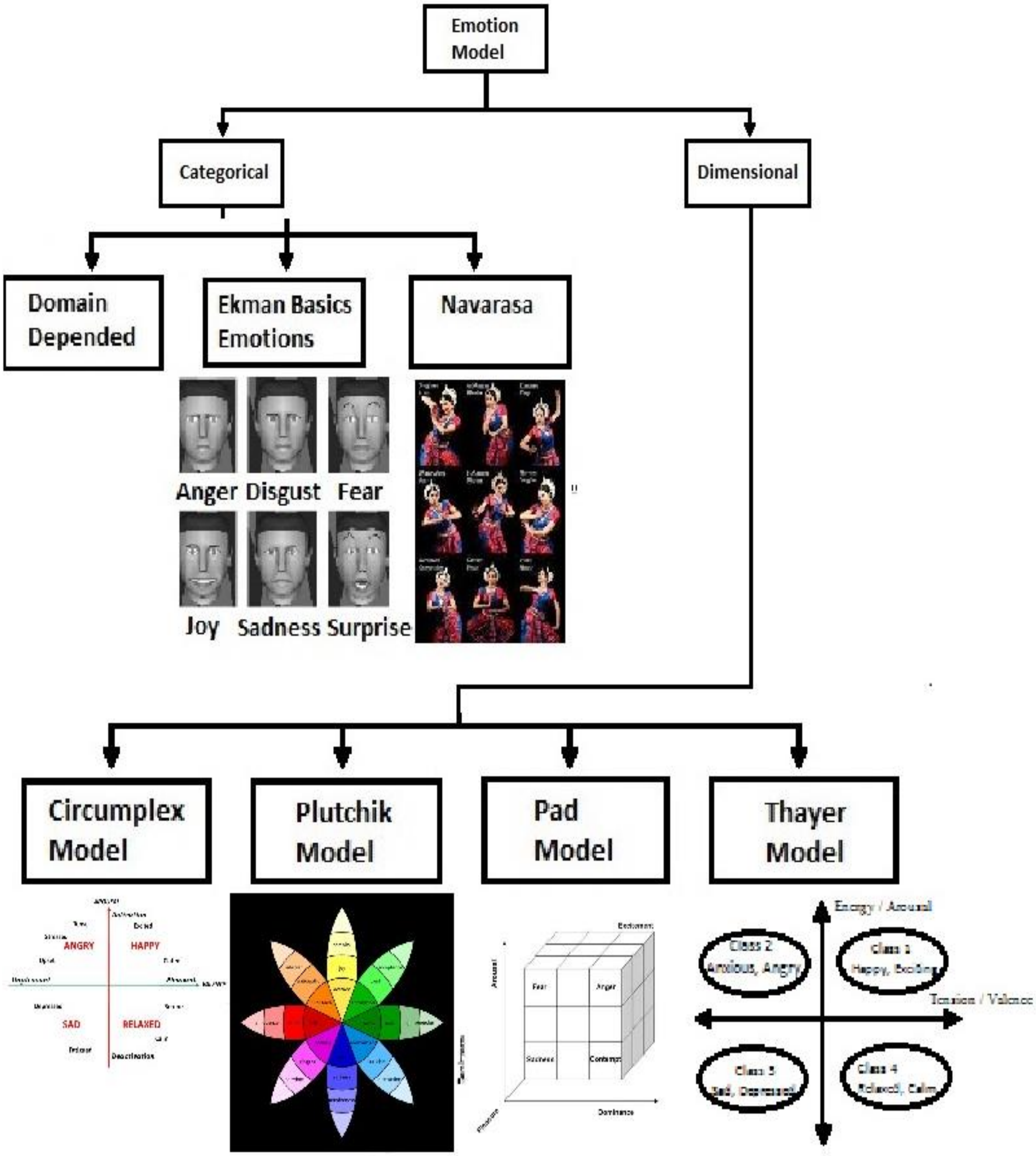



**Figure 2.3:** Emotion Models

### 2.4.1 Categorical Emotion Models

This model utilizes six basic categories of emotions. Identification of emotions is carried out by using words representing emotions or their classification tags. For example, anger, disgust, fear, joy, sadness, surprise. This model comprises of significant as well as insignificant emotions. Separate set of features are used by every emotion to signify the provoking situation. Many researches carried out in the domain of affective computing focus on the above referred 06 fundamental emotions. However, several researchers have shown that different fields might be requiring different groups of emotions

**Pros and Cons of Categorical Emotion Models**

Advantages of the categorical model are that it uses emotion tags which are easy to comprehend and it symbolizes the emotions automatically. For example, the emotions are categorized in such a way that each category has separate elements and large number of emotions. Only those emotions are included in a group, which fall in that category. Moreover, one emotional state is also shown by using multiple categories of emotion because of the variations in culture, environment, linguistics or personalities. This creates difficulty in tagging it to the real category of emotion. These observations show that despite defining the emotion categories, several emotional states are still not represented by this categorization. This can lead to inefficient recognition of the emotions. As a result, the subject is more likely to link an emotion from the available categories instead of labeling the emotion himself. Therefore, using this method can make the researchers to select a separate category. Another problem is linked to the previous one. At times it is not possible for researchers to choose a related classification as it is not there in the available labels. Therefore, a categorical model has the limitations from the identification point of view while linking the actual emotions observed by people. Still the categorical model is more common and multiple variants of this model are being used because of its flexibility and ease of implementation. Categorical models may differ only in the number of classes they are representing.

### 2.4.2 Dimensional Emotion Models

Dimensional model for identification of emotions is the 2nd method. It uses dimensional form to represent the emotional states. In dimensional model, multiple emotional states are linked with the same group of dimensions. The emotions are elaborated either in 2 (valence and arousal) or 3 (valence, arousal and dominance) dimensional space. Every emotion takes a place in this space. Valence dimension explains how positive or negative an emotion is and it varies from unpleasant feelings to pleasant feelings (happiness). Dimension of arousal indicates the extent of excitement which the emotion is showing and it varies from sleepiness or boredom to wild excitement. The dominance dimension represents the level of influence for example having control on the emotion. Various dimensional models have been proposed. In Russell's model terms pertaining to emotions are arranged in a circumplex shape that helps a subject to pick a point at any place in between two terms pertaining to specific emotions. Numerical data is acquired from the respective position of the points in the 2 dimensions i.e. valence and arousal. The dimensions of energy and stress have been utilized by Thayer. In this stress energy model, contentment is placed in low energy / low stress, exuberance in high energy / low stress, anxious / frantic in a high energy, high stress, and depression in low energy / high stress. Plutchik has proposed 8 basic emotions and the corresponding levels of these basic emotions. For example, if emotion of anger is taken, then the extreme level of anger is represented by rage and its milder level is represented by annoyance. Similarly, the rest of the emotions have been defined with their respective levels. Albert Mehrabian presented a 3-dimensional Pleasure, Arousal and Dominance model for explaining and quantifying the emotional states. The D (dominance) is utilized to ascertain if the subject is having control of the state or not. The pleasure dimension is similar to the valence in the Russel's model.

**Dimensional Emotion Model-An Overview**

Explaining the emotional state with emotion dimensions gives few advantages. Dimensional models provide the advantage that they are not linked to a specific emotional state. In these models, scores are utilized to classify the emotions in 2 or 3 dimensions. These dimensions are used to get fine emotion concepts which vary with small margin once compared with the broader categories of emotions. The dimensions provide a distinctive recognition and a large variety of emotion concepts. Specially, for measuring the discrete emotional states, a dimensional explanation is a desirable method. In contrast to the categorical model, emotional states are linked together in a dimensional space. Extent of comparison amongst the emotion categories can be calculated by using a dimensional model. The neighboring classes in the space are closely related to each other, whereas, many categories have noticeable differences between them. In nut shell, a dimensional emotion model is a helpful depiction which covers all related emotions and gives a way to ascertain the resemblance between emotional states. It proposes that we can explain every emotion in the form of combination of arousal and pleasantness. Dimensional model cannot be used for separate categories.

## 2.5 Emotion Recognition

It is the process in which human emotions are identified. Different people have different levels of correctness in identifying other's emotions. Therefore, using technology for identification of emotions is a fast growing area of research. It has been observed that the use of technology can be optimized if it utilizes multiple related inputs. Till now, maximum research has been done in the area of automatic identification of face expressions using video, audio, text and physiology through input of the wearable sensors [89]. Emotion recognition can be done using physical signals such as speech signals and facial expressions. Emotion charting with speech signals has some limitations. For example, there is an ambiguity to what extent the speech conveys the original feeling and how emotional the speech is which is being used for emotions charting. Also, privacy plays an important role and whether action by the participant is matching with his feeling or not.

Also for emotion recognition from facial expressions, privacy of the camera becomes the limitation. Also the location of the camera for monitoring of the face expressions affects the reliability of the system.

### 2.5.1 Emotion Recognition Using Physiological Signals

Various methods like ECG (electrocardiography), MRI (magnetic resonance imaging), PET (Position Emission Tomography and EEG (Electroencephalography) show varying information, about functions of the different body organs. In recent years, medical diagnostics and forecasting the patient's postural balance are being done by recording and studying the EEG signals. Similarly, charting of human emotions from physiological indications as a result of various stimuli is also very popular. Physiological responses used for charting of human emotions are as follows:

**Heart Rate Variability / Electrocardiography**

Heart Rate Variability or interval between beats is a physiological assessment and measured as time difference between the heart beats noted in milliseconds. It is also called as R-R interval as it is calculated using ECG signal as time gap between the successive R peaks. Heart Rate is defined as average number of heart beats of a human being in 60 seconds time period. Limitations of ECG that can affect the reliability of the emotion recognition model are as follows. :

i. Heart rate utilized for ECG is dependent upon the health of individual's heart and his state at that time

ii. The studies which have used ECG signals have been successful to classify only few categories such as positive / negative feelings.

iii. Heart rate was observed to sometimes increase and sometimes decrease as a result of the sadness.

iv. Once body temperature of a male person was increased from 36 to 36.6 º C, his heart rate was observed to increase by almost 40 beats per minute.

**Galvanic Skin Response (GSR)**

It is the measurement of the flow of electricity through a person's skin. It is also called skin conductance. In order to measure the GSR signals, electrodes are positioned on surface of the skin and their electrical potentials are measured. For this purpose, these electrodes are positioned on the fore and middle fingers of hand. The variations in the values vary with the state of emotion, stress and arousal of the person. Limitations of using GSR are as follows:

i. Measuring GSR is affected by external variables like temperature and humidity and this can result in erratic results.

ii. Inconsistent results can also be due to GSR fluctuations can be caused by the Internal factors like medications.

iii. There is an inherent delay of 1-3 seconds in measuring of GSR signals.

**Brain Signals**

Emotions are assumed to be linked with the activation of those areas in the brain which control our focus and attention. It is a complicated cognitive function performed in human brain and linked with numerous oscillations. Emotions are event based potentials which calculate cognitive levels and are induced as varying voltages once an event takes place. The activity in the brain can be noted through use of neuro-imaging methods like EEG (electroencephalography), PET (positron emission tomography), NIRS (near infrared spectroscopy), MEG (magnetoencephalogram) and functional fMRI (magnetic resonance imaging). Out of these methods, EEG has been much popular among researchers due to the objective assessment and portability. On the other hand, heavy equipment is used for PET and fMRI which is not easy to move. EEG is the mostly utilized method because of its better temporal presentation, non-invasive equipment and lower prices.

## 2.6 Electroencephalogram (EEG)

EEG is monitoring of electrical potential oscillations in the brain arising from the flow of ionic current between brain neurons. EEG is defined as a procedure which senses electrical activity in the brain using flat metallic electrodes attached to the scalp. The brain cells called as neurons use electrical signals for communication and are in active mode at any given time, controlling the way our thinking takes place and do interaction even during sleep time. Therefore, a thought is actually an arrangement of all the time varying electrical signals. There are approximately 100 billion neurons in the brain of an average human being. The variations in the neuron activity can be detected through an EEG diagnostics which can be helpful in identifying any disabilities of the brain [90]. The electrical signals are detected and metered for acquiring EEG signals at the locations of electrodes on the scalp. Electrodes are positioned as per the 10-20 system. This system provides an international standard to make sure that the electrodes are placed accurately each time. Therefore, by using the standardized 10-20 system, the study carried out on one participant can be compared to the other studies [24]. The 5 mostly used EEG frequency bands and their respective frequency ranges are shown in the figure. Greek signs have been used by scientists for these bands i.e. delta, theta, alpha, beta and gamma. These bands are very different and range from delta wave on slow end to the gamma wave on the hyper active end [92].

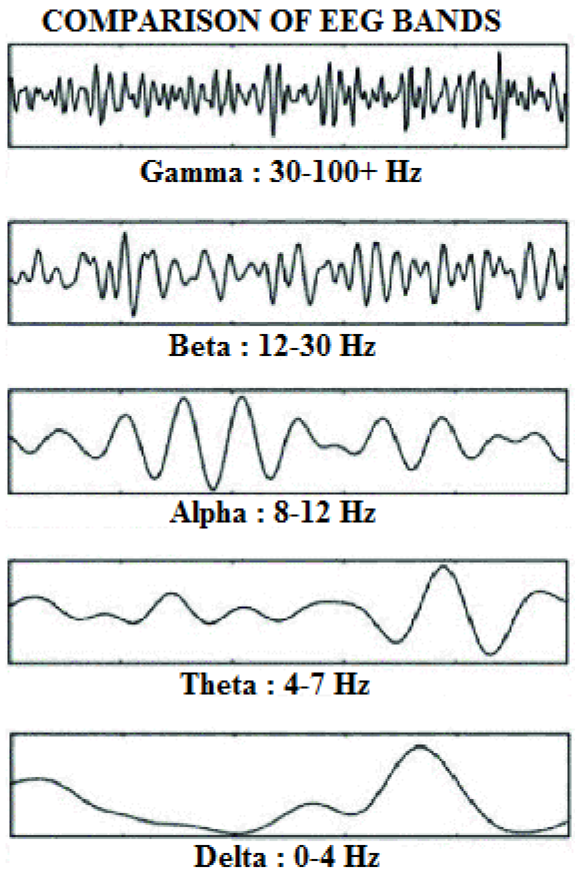



**Figure 2.4:** EEG Frequency Bands

### The Rhythm of Neurons

A neuron moves to and fro in a systematic rhythm. At times many neurons forming a neuron ensemble (neurons engaged in a specific neural calculation) move to and fro together in a coordinated manner. Once sufficient neurons carryout to and fro movement at one time with a specific frequency, this results into movement at a larger scale, sufficient enough to be monitored through EEG. As the waves become larger in amplitude, more number of oscillations are observed [92].

### Capturing Raw Data

Once the electrode detects a movement of neuron, it also records the nearby activities. Due to this reason, this data is known as raw EEG. It is a complicated wave form of not only the neuron movement but also the neuron activity in the surrounding.

### 2.6.1 10-20 Positioning System

International 10-20 system was basically devised to have a standard way of putting electrodes for normal medical EEG (electroencephalography) so that the comparison of related research can be carried out over the years and subjects can be analyzed in reference to each other. Benefit of this system over the other systems was that it was devised on separate measurement of head size and anatomical measurements which results in an acceptable relation between the position of the electrodes and the brain structures [91]. The 10-20 system is acceptable all over the world for explaining and putting the position of the head electrodes with reference to EEG examination. This system is designed on the basis of correlation between the position of electrode and the specific areas of brain. The “10” and “20” correspond to the fact that the distance between the two adjoining electrodes are either 10% or 20% of the total distance of the head from right side to left side and similarly from front to the rear side. Two anatomical references are utilized for placement of EEG electrodes. One is the nasion and the other one is inion. The perimeters of the head in the transverse and median planes are calculated from these two positions [2][17].

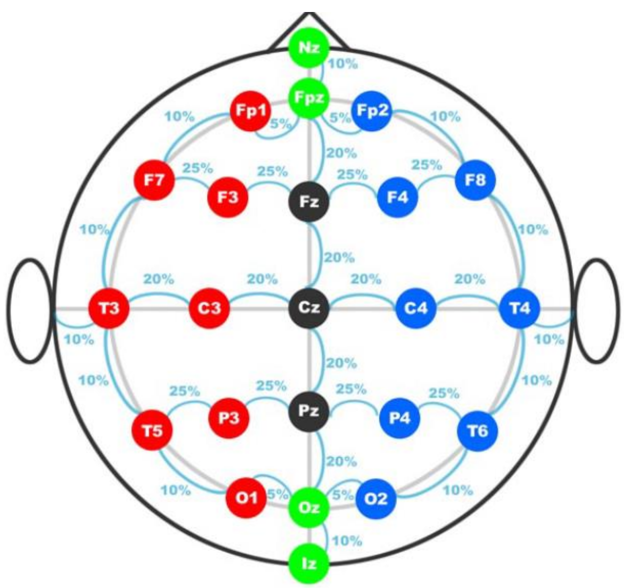



**Figure 2.5:** 10-20 Electrode Positioning System

Additional electrodes are introduced while recording a detailed EEG as per 10% division which is used for the sites mid way between the already present electrodes in the 10-20 system. New system for naming the electrode is more complex and is called Modified Combinatorial Nomenclature (MCN). This system uses 1, 3, 5 ,7 and 9 electrode locations for representing the 10%, 20%, 30%, 40% and 50% inion to nasion distance for left hemisphere. Additional letter is used for naming the intermediate electrode positions. Introducing additional letter gives the provision of naming the electrode positions in between. Following codes are used for naming these positions:

(a) AF is used for electrode position between Fp and F

(b) FC is used for electrode position between F and C

(c) FT is used for electrode position between F and C

(d) FT is used for electrode position between F and T

(e) CP is used for electrode position between C and P

(f) TP is used for electrode position between T and P

(g) PO is used for electrode position between P and O

Four electrode locations (T3, T4, T5 and T6) of 10-20 system have been renamed in Modified Combinatorial Nomenclature as T7, T8, P7 and P8 respectively [1].

Level of consciousness of a human being is linked to the EEG signals. The EEG moves to greater frequencies and lesser amplitudes once the activities increase. The EEG signal is dominated by alpha waves once the eyes are shut. Rapid Eye Movement (REM) is a part of sleep where a human being dreams. This can be observed as a distinctive EEG signal. Once a person is in deep sleep, the EEG signals have greater frequencies and less amplitude deflections and are called delta waves. However, once a person has full cerebral death, no activity can be observed from his brain [17].

### 2.6.2 EEG Bands

Every human being exhibits 5 different kinds of electrical signals or brain waves across the cortex. These can be recorded using EEG (electroencephalograph). Through EEG, the wave pattern of the brain can be monitored by the researchers. Every signal produced by the brain has a function and contributes in achieving the best possible brain functioning.

Stress management, task focusing and quality of sleep are affected by our brain's capability to be flexible and / or transition through multiple brain signal frequencies. If any of the 5 kinds of brain signals is over or under produced, problems can arise from it. Therefore, none of these signals is better or more significant than others. Every signal has a distinct function to play its role during different situations. This can be acquiring new knowledge or helping us in releasing stress after a hectic day. Gamma, beta, alpha, theta and delta are the 5 brain signals arranged in the order of higher to lower frequency [4].

Once a person is awake, the EEG shows all 5 types of brain signals at any given time. However, one specific brain signal will be dominant depending upon the state of consciousness of that person. If a person is awake but has bad Attention Deficit Hyperactivity Disorder (ADHD), he may experience slower wave (alpha and / or theta) generation as compared to beta waves. While sleeping, slower frequencies are there in combinations, however, gamma has also been observed to be linked with Rapid Eye Movement (REM).

## 2.6.2.1 Gamma Frequency Band

Signals in gamma frequency band (40-100 Hz) have been observed to be the fastest brain signals on EEG. As these are the fastest signals, they have the smallest amplitude on EEG once compared to other brain signal frequencies. In the human brain, gamma waves have a significant role in sensory binding as well as information processing. These waves have the ability to connect and process information across the complete brain. Better gamma waves result in having good problem solving skills, sympathy, self discipline and intellect. These waves also assist in perceiving reality and memorizing things. People lacking gamma waves have been observed to have learning difficulties, mental retardation and cognitive challenges. Reduced signals of gamma brain waves have also been seen in people having brain injuries. Brain of a person having healthy gamma wave activity should be functioning properly.

## 2.6.2.2 Beta Frequency Band

Beta Brain Waves are mostly observable in awake state (12-40 Hz). These are high frequency less amplitude waves. These waves can stimulate the thought process. A person's ability to focus on the daily tasks depends on the presence of the beta waves in required amount. Excessive beta waves can result in anxiety or stress. Greater beta frequency is linked with increased arousal. Taking caffeine or other stimuli can result in increased beta waves. People experience these waves during day to day tasks such as thought process, creative writing etc. Beta waves are classified into three categories:

- **Low Range** : Frequency range of Low Beta waves or 'Beta 1' is from 12 Hz to 15 Hz. Beta activity in this range is generally associated with calm and introverted focus.
- **Mid Range:** These 'Beta 2' waves are faster waves which are linked with enhanced anxiety, energy and output.
- **High Range:** These 'Beta 3' waves are the fastest waves which pertain to stress, anxiety and high arousal. After these high beta waves, gamma waves start.

#### 2.6.2.3 Alpha Frequency Band

Alpha brain waves fall in the frequency range between beta and theta waves (8-12 Hz). Purpose of these waves is to calm down a person as per requirement and enhance relaxation feelings. Once a person is stressed, alpha waves are blocked by the beta waves. These waves generally originate from occipital lobe when a person is drowsy or sleeping. Alpha waves are observed more prominent once people try to relax while keeping eyes closed. However, when eyes are opened, the alpha waves become less [10].

#### 2.6.2.4 Theta Frequency Band

Theta brain waves are related to the daydreaming, sleep and deep state of relaxation (4-8 Hz). Balanced amount of theta waves increase our capacity to be creative. However, extremely enhanced theta waves can lead to depression. If theta waves are not produced excessively while a person is awake, it is very helpful. These waves are generally more in children as compared to adults. People who practice meditation also experience increased theta activity.

#### 2.6.2.5 Delta Brain Waves

Delta waves are the slowest waves (0-4 Hz) and found mostly in infants and young children. Production of delta waves reduces with age. These are linked with deep sleep which is relaxing and healing and produce various types of hormones like human growth hormone. These waves have been known to regulate the functions like heart beat, kidney and digestive functions [4].

## 2.7 Challenges of EEG Signals:

EEG is the non invasive monitoring of the electrical signals of the brain utilizing the electrodes positioned on the scalp. Recorded electrical signals are of very low strength because the electrodes are placed at a distance from the source generating the signals, therefore, they have low signal to noise ratio. Moreover, the recorded electrical activity does not clearly exhibit the spatial information about the brain. The data of EEG signals comprises of large number of channels with several frequency bands which can provide critical information. Therefore, researchers have always been interested in the analysis of the brain activity to extract maximum information from the EEG signals [46]. Real world EEG based Brain Computer Interface systems still face some challenges. EEG signals are generally noisy. Other than the noise of the equipment, EEG signals have unique inherent noises. While the signals are being recorded, eye blinking, cardiac rhythm etc all contribute to the noise in the EEG signals. Another limitation is that the person under evaluation cannot keep his thoughts focused on the tasks during the complete process. An EEG based BCI system generally comprises of 8-128 signal channels which gives an inadequate signal resolution in comparison with a task based on an image or video. Moreover, there lies an ambiguity in the captured EEG signal and the actual electrical activity inside the deep structure of the brain. Most of the techniques used for recognition of brain signals depend on manually generated features, therefore, they require a lot of time consuming preprocessing prior to classification. Few techniques utilize removing the noise from the signal or use feature selection techniques followed by the classification model. This method involving additional preprocessing step is time consuming, requires the professional expertise and is not convenient for training as well as implementation.

# CHAPTER 3: LITERATURE REVIEW

Brain computer interfaces are dependent upon emotion recognition systems which is a popular area of research. With advancement in technology, the requirement of emotion recognition systems has risen drastically. Emotion detection systems using physical signals such as speech signals, facial expressions or gestures have proved to high classification accuracy but they are practically of lesser use for patients suffering from psychological disorders. EEG based emotion recognition models are more reliable due to the unforgeable nature of brain signals. Moreover, despite the low signal-to-noise ratio and non-stationary nature, EEG signals are equipped with high time resolution, reasonable spatial resolution and have convenient, non-invasive and low cost acquisition system by portable devices. [80]

## 3.1 Emotion Assessment Methods

The increasing research on physiological signals based emotion recognition models is due to the fact that these signals represent actual inner feelings and emotions and hence cannot be controlled subjectively. [50] The assessment of emotions can either be subjective or objective. For subjective measures self-report tools, questionnaires or pictorial tools [23] are used to denote individual's emotions. However, this method cannot offer real-time emotional experience and feelings. As far as objective measures are concerned, these do not require participant's assistance. Objective measurement methods comprise of instruments that can acquire heart signals, brain signals, skin responses and blood pressure variations.[23] Recently, hybrid methods that combine both subjective and objective assessment approaches are employed that enhance the reliability of emotional states and the recognition models utilizing them. Physiological signals such as EEG are best suited for emotion recognition models as they provide true insight of the internal brain activity of the central and autonomic nervous systems. [73]

## 3.2 Feature Extraction Techniques:

A standard EEG emotion recognition system usually consists of two steps, one is the extraction of discriminatory features which is followed by classification of emotions. In general, EEG features can be classified into four categories namely time domain features, frequency domain features, time-frequency domain features and multi-electrode features. Time domain features include, statistical features, Fractal dimension, Hjorth features, higher order crossings and non-stationarity index. [15] Time domain features primarily extract temporal details of EEG signals.

As compared to time domain features, the frequency domain features extract the information of EEG signals from frequency perspective. Most popular frequency domain feature mining method is the decomposition of EEG signals into constituent frequency bands which are termed as Delta(1-4Hz), Theta(4-7Hz), Alpha(8-12Hz), Beta(12-30Hz) and Gamma(>30Hz) bands and then extraction of features from the respective bands. Power spectral density, Differential Asymmetry, Rational Asymmetry and Differential Caudality [15] are the features being usually extracted and used for further classification. Power spectral density is most commonly used.

The time-frequency domain features consist of discrete wavelet transform, short time Fourier transform and Hilbert-Huang spectrum. [16].Time-frequency domain are most commonly extracted by using short time Fourier transform but it becomes problematic to apply in real time systems because of its greater computational time. Hjorth parameters [37] is used as a substitute for STFT as it involves simpler computation. Multi-electrode features comprise of magnitude squared coherence estimate. [16]

### 3.2.1 Time domain Features:

In previous researches, different EEG features have been suggested for emotion recognition. Time-domain analysis is carried out to obtain time series characteristics that identify essential information relevant to emotions [43]. Many recognition studies used statistical features of EEG signals.

[18] Employed Hjorth features which are statistical features computed from activity, mobility and complexity. Kroupi [19] Employed non-stationary index and Petrantonakis [20, 21] employed higher order crossings for emotion charting. Tripathi [22] used DEAP dataset, extracted time domain features and then did a comparison using DNN and CNN. First of all, the dimensionality of the massive data was reduced by dividing 8064 readings from each channel in 10 small batches. Then nine statistical features such as mean, median, maximum, minimum, standard deviation, variance, range, skewness and kurtosis values [22] were extracted for each of the batches to get 90 values for each of the channels. Convolutional neural network was employed which achieved accuracy of 81.41% and 73.36% for two classes and 66.79% and 57.58% for three classes of valence and arousal respectively.

### 3.2.2 Frequency domain features:

Many studies used extraction of frequency domain features and used different classification models for EEG emotion classification. Most widely used frequency domain feature is the power spectral density. Alnafjan et al [23] used power spectral density features to decompose EEG signals into constituent frequency bands named as theta, alpha, beta and gamma bands using Python signal processing toolbox. Power spectral density features and pre-frontal asymmetry features were extracted. The features being extracted were used to train Deep Neural Network. Two classes per dimension were detected and accuracy of 82% was achieved for two classes.

Jirayucharoensak et al. [24] also used power spectral density features for emotion classification from EEG signals. Power spectral density of EEG signals after being down sampled to 128Hz was calculated using fast Fourier transform with 128 samples as window size, in each of the five frequency bands named as theta, lower alpha, upper alpha, beta and gamma bands. To measure the asymmetry in brain signals generated as a result of stimuli, spectral power difference between 14 pairs of electrodes spread over the brain hemispheres is computed within each frequency band. In order to minimize dimensionality, PCA and covariate shift adaptation was used and reduced features were sent as input to deep learning network (DLN). The classifier achieved accuracy of 53.42% and 52.05% for three classes of valence and arousal respectively. Pan [25] also used frequency domain features for emotion recognition from EEG signals. In this analysis different power spectral features in the frequency domain and features determined by groupings of electrodes were extracted from different frequency bands of the preprocessed EEG signals. These features were used in five different classifiers including Naïve Bayes and support Vector machines amongst which logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm with Differential Entropy (DE) features achieved the highest accuracy of 77.17% and 77.03% for 2classes of valence and arousal respectively.

### 3.2.3 Time-frequency domain features:

In literature many studies used time-frequency domain features for Emotion Recognition. This includes methods based on Fourier Transform. Many studies used short time Fourier transform for classification of emotions because STFT is able to disclose significant high resolution signal characteristics in combined time –frequency domain. Valenzi et al [26] used high quality data from 9 intelligent participants to classify four emotional states( and employed short time Fourier

Transform as a feature extraction technique on data from 32 electrodes whose results were further used to compute average spectral power of the constituent frequency bands being named as Delta, Alpha Lower Beta, Upper Beta and Gamma bands. Further Linear Discriminant Analysis was employed as dimensionality reduction technique. Classification accuracy was calculated for seven algorithms (both supervised and unsupervised) and highest average accuracy of 97.2% was reported for Support Vector Machine.

**Feature Extraction using Wavelet Transform:**

Discrete wavelet transform (DWT) is another technique being found in recent studies. DWT breaks down the signal in various approximations and detail levels conforming to different frequency ranges and preserves time resolution of the signal [27].

DWT acts as a substitute to power spectral density for determining the eminence of various frequencies present in EEG signal. The DWT consists of cascaded high pass and low pass filtering stages where the low pass filter gives approximation coefficients that are forwarded to the next stage. The stage where coefficients contain the wanted frequency bands is dependent upon actual signal's sampling frequency [16]. The wanted frequency bands in case of EEG signals are the alpha beta, gamma and theta bands. The most critical concern in this process is the selection of suitable mother wavelet.

Nasehi et al. [29] used Gabor functions and Discrete wavelet transform to extract features from four EEG channels. DWT and Morlet mother wavelet was applied to decompose the EEG signals in each epoch into frequency bands and then spectrum energy was computed for each of the bands using DFT. The extracted features were then sent to probablistic Neural Network to classify six basic emotions and achieved the accuracy of 59.99%. Murugappan [30] also used DWT over five frequency bands to extract linear (power, standard deviation, and variance) and non-linear (entropy) statistical features using the wavelet functions db4, db8, sym8 and coif5 [30]. KNN classifier was used to classify these extracted features. The classification accuracy of 82.87% and 78.57% was achieved for 62 channels and 24 channels correspondingly for classification of five emotions.

Mohammadi [31] worked on DEAP dataset to extract spectral features. The 60 second signals were divided into 2 and 4 seconds wide windows. DWT was applied to disintegrate each window into five frequency bands using db4 mother wavelet function and then entropy and energy was calculated for all frequency bands as features. KNN (K=3) was used for classification of EEG signals from 10 channels and reported the accuracy of 86.75 % and 84.05 % for valence and arousal respectively. Bazgir et al. [32] also employed DWT to split the EEG signals. The EEG signals were divided into 4 and 2 seconds windows and then were split up into frequency bands (gamma, beta, alpha and theta). From each frequency band, spectral feature (energy and entropy) were computed and PCA was used as a transform to create the features mutually uncorrelated. EEG signals from 10 channels were used and the extracted features were classified using support vector machine (SVM) with eight-fold cross validation. The achieved accuracy was 91.1% and 91.3% for valence and arousal respectively.

| S/No | Title/Author | Year | Dataset | Modality | Features | Classifier | Accuracy | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Arousal | Valence | Dominance |
| 1 | Pan et al. [25] | 2020 | DEAP | EEG | Spectral features in frequency domain | logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm | 77.03% (2 classes) | 77.17% (2 classes) | … |
| 2 | Bazgir et al. [32] | 2018 | DEAP | EEG | DWT( entropy and energy extracted from each band) | SVM | 91.30% | 91.10% | … |
| 3 | Tripathi et al. [22] | 2017 | DEAP | EEG | Time domain features | Convolutional Neural Network (CNN) | 73.36% (2 classes) 57.58%(3 classes) | 81.41% (2 classes) 66.79%(3 classes) | … |
| 4 | Nafjan et al. [23] | 2017 | DEAP | EEG | PSD Frontal asymmetry | Deep Neural Network (DNN) | 82%(2 classes) | | |
| 5 | Mohammadi et al. [31] | 2016 | DEAP | EEG | spectral features using DWT(entropy and energy from each band) | KNN | 84.05% | 86.75% | … |
| 6 | Jirayucharoen et al. [24] | 2014 | DEAP | EEG | power spectral features | DLN with a stacked auto-encoder (SAE) | 52.05% (3 classes) | 53.42% (3 classes) | … |
| 7 | Valenzi et al. [26] | 2014 | 9 participants | EEG | spectral power using STFT | Support Vector Machine | 97.2%(4 emotional states) | | … |
| 8 | Nasehi et al. [29] | 2012 | 10 aprticipants | EEG | spectral, saptial and temporal features | Probablistic Neurar Network (PNN) | 59.99%(6 Emotional states) | | … |
| 9 | Murugappan [30] | 2011 | 20 participants | EEG | linear(power, standard deviation, and variance) and non-linear(entropy) statistical features | KNN | 82.87% (62 chs, 5emotions) 78.57%(24chs 5emotions) | | … |

**Table 3.1:** Literature Review of Feature Extraction Methods

## 3.3 Classification

The method of grouping a certain set of data (structured or unstructured) is known as classification. This grouping is also sometimes called as predicting the label, class or category of data points. The process of classification can be regarded as estimation of plotting function from inputs to distinct outputs variables [33]. The task of emotion recognition can be referred to as a classification problem since the aim of the process is to forecast the true class of the emotion [16].

### 3.3.1 Machine Learning Models

Machine learning is dependent upon the patterns in data to acquire the insight of a particular problem with any explicit guidelines. Machine learning can be categorized into supervised learning and unsupervised learning models. In supervised learning, the dataset is divided into two parts, one part called the training set is used to train the model and the second part called the test set is used to test the trained model during the learning phase. During supervised learning the model learns from the labelled training examples for classification and regression tasks in order to test the new data whereas in unsupervised learning the training data is not labelled and the model learns hidden patterns or clusters the data into groups. For emotion classification using EEG signals many different types of models have been proposed and employed including neural networks, linear classifiers, nearest neighbor classifiers, non-linear Bayesian classifiers and combinations of classifiers [34]. Linear classifiers including Linear Discriminant Analysis (LDA), Regularized LDA, and Support Vector Machine (SVM) use linear decision boundaries. [34] Bayesian classifier such as a Bayesian quadratic classifier and Hidden Markov Model (HMM) use Bayes rules for classification. K Nearest Neighbor classifier (kNN) does classification based on the distance of predefined number of nearest neighbors. Outputs from different combinations of classifiers are also used for classification. [35]

### 3.3.2 Deep Learning Algorithms

Deep learning algorithms are ones in which the features as well as the model are learnt from the data itself. Neural networks are inspired by the structure and working of human brain. Neural networks can learn from data just like a human brain can learn and identify different patterns in specific information. Neural networks have a series of units that are trained to learn features from the input data. Types of neural networks commonly used include Convolutional Neural Networks (CNN), Generative Adversarial Network (GAN), Long Short-Term Memory Networks (LSTM), Recurrent Neural Networks (RNN), Artificial Neural Networks (ANN) and Deep Neural networks (DNN). Deep learning methods are extensively used as compared to other machine learning models in BCI applications as these methods can adapt and learn quickly from the varying brain signals to classify them accurately . [35]

Despite the capabilities of neural networks to deal with high dimensional data, they have inherent problem of long time-consuming computation. The hardware related issue has been resolved greatly with the development of low-cost and powerful graphic processing units (GPU's) permitting the use of neural networks. This innovative development has increased the interest of researches in the use of neural networks in different applications. The automatic optimization of parameters in neural networks makes them very appealing without involving much domain specific knowledge. Lately, deep learning has attracted many researchers in the field of emotions classification from EEG signals despite the associated challenges with EEG signals such as low SNR, high randomness and high dimensionality. [36]

Numerous machine learning techniques and algorithms have been proposed to tackle with high dimension vast volume EEG data used for emotion recognition. In the recent years deep leaning has attained remarkable progress and has affirmed to be very effective and proficient in complex problems. Deep learning algorithms have been employed widely in the areas of image recognition, face and speech recognition but the design of a suitable deep learning model for EEG based emotion classification is still a challenging task. This is because EEG signals are associated with many problems. EEG signals have low signal to noise ratio, high randomness, non-stationary nature, and convey poor spatial resolution related to the electrical activity inside the brain due to limited number of channels. Not only the noises of sensory equipment like power line interference and faulty electrode connections , some physiological actions such as blinking of eyes , muscle movement and heart beat also contribute to the collection of EEG signals having low signal to noise ratio [15]. Also the associations between intentions inside the brain and the corresponding EEG signals are vague to some extent. Moreover, EEG signals have substantial variations between two different subjects or fluctuate within a single subject over time.

For emotion classification from EEG signals many algorithms have been presented in literature amongst which use of deep neural networks have proven to be most effective. The most popular DNN approach is the Convolutional neural network which has been used increasingly widely to solve numerous classification problems [15].

The ability of CNN to extract features directly from the raw data without the extraction of hand crafted features makes it the most powerful method employed for various classification problems. Lately, convolutional neural networks and its modified versions have been employed in the area of emotion recognition problems from EEG signals. [37] Mei et al. [38] employed CNN for emotion classification using EEG signals. The matrices were created which contain functional relationship between EEG channels. EEG signals for each trial were segmented into 4 second long segments and each segment was decomposed into four frequency bands named as theta, alpha, beta and gamma bands and Pearson Correlation Coefficient (PCC) [38] was calculated between every pair of signals of each channel, therefore 4x32x32 matrices were computed for each trial and sent to CNN as input. The achieved accuracy was 85%(2 classes), 78%(3classes) and 75%(4 classes(LVLA, HVLA,LVHA, HVHA).

Li et al. [39] proposed a hybrid model that combines Convolutional neural network and Recurrent neural network. The input to CNN was given in the form of 2D frames. These frames were created first by taking continuous wavelet transform of each channel's 1D signal by using db-4 wavelet, then scalogram was formed from the output of CWT. The elements of each channel's scalogram were added up to get a vector. These vectors from all the channels were stacked together to form a 2D frame corresponding to one window. The window was then moved ahead to get all the 2D frames for single trial and were used as input to CNN which extracted the cross-channel correlation features. Which were fed to long short-term memory (LSTM) unit that was used as RNN to extract contextual information. The decision layer outputs the results for all trials and the accuracy of proposed C-RNN was reported to be 74.12% and 72.06% for arousal and valence respectively.

Zhang et al. [40] proposed two neural network models that used spatio-temporal features extracted from EEG signals. The 1D EEG signals were converted to 2D frames according to the location of electrodes in the acquisition system. The incoming 2D frames were divided by employing the sliding window approach and sent as input to both neural network models. The cascaded Convolutional Recurrent network model employed 2D-CNN to extract spatial feature sequences which were sent to cascade RNN to extract temporal features. The parallel model employed 2D CNN to extract spatial features and RNN in parallel extracted temporal features from 1D EEG vectors, both of which were concatenated and sent to softmax layer for final intention recognition. RNN used in both models employed two layers of long short-term memory (LSTM) units. The cascade and parallel models achieved accuracy of 98.31% and 98.28% respectively for multiclass intentions prediction. Song et al. [15] proposed Dynamical Graph convolutional neural network (DGCNN) that included an adjacency matrix which showed functional relationship between the nodes of graph i.e the EEG channels. The adjacency matrix is learnt by the network dynamically. The EEG signals are disintegrated into five frequency bands named as delta theta , alpha, beta and gamma, then the frequency domain features of power spectral density feature (PSD), the differential asymmetry feature (DASM), the rational asymmetry feature (RASM), and the differential caudality feature (DCAU) were determined. The DGCNN model consisted of graph convolutional process through learnt adjacency matrix, convolution, relu activation and fully connected layer. The achieved accuracy was reported to be 84.54%, 86.23% and 85.02% for 2 classes of arousal, valence and dominance respectively on DREAMER dataset.

Salama et al. [41] proposed 3D CNN for emotion classification. To avoid overfitting data augmentation was first applied which involved adding of some noise to the EEG signals to create e noisy signals which were then fed to the CNN model. The signals from each channel were divided into segments and were combined together to form 2D matrices. Then the third dimension was time which was added by appending fixed number of consecutive frames together which makes a chunk. The label to each chunk was assigned according to majority rule based on the number of frames in each chunk. The accuracy of proposed model was dependent on the chunk size and varied with varying chunk sizes. The proposed model which was able to extract spatio-temporal features from 3D input achieved the accuracy of 87.44% for valence and 88.49% for arousal for two classes each. Yang et al. [42] also proposed a hybrid neural network that fused together Convolutional neural network and Recurrent neural network was used for classification on DEAP dataset. 2D EEG structures were obtained from single dimensional EEG signals according to the location map of electrodes. Sequence of 1D vectors were converted to sequence of 2D frames and then incoming 2D frames were divided into groups using sliding window method and fed into CNN to extract spatial features. 1D vectors are also fed to RNN to extract temporal features. The two stacked RNN layers were created by using long short-term memory unit. The temporal and spatial features were then concatenated to be fed to softmax layer for classification of emotions. The achieved accuracy was 91.03% for valence (2classes) and 90.8% for arousal (2classes). Chao et al. [43] proposed a capsule network for emotion classification in which multiband feature matrices were used as input. To construct these matrices the power spectral densities (PSD) were calculated for each of the four frequency bands named as theta, alpha, beta and gamma. The sub-matrices of PSD feature values were made for each of the frequency bands according to electrode positions of 10-20 system. Further these sub matrices were concatenated to form final feature matrix. Capsnet was employed that is a multilayer network composed of capsules which perform different operations layer wise and can translate relative associations between individual parts and entire object. Classification accuracy of 0.6828, 0.6673, and 0.6725 for arousal, valence and dominance (2 classes each) respectively was achieved using capsnet. Zeng et al. [44] also employed deep learning and proposed sinc net-R classifier to classify emotions in 3 classes (positive, neutral, positive) on SEED dataset. The input data was comprised of temporal as well as spatial features that showed relationship between channels. EEG signals from 62 channels were down sampled to 200Hz and 0.5 sec sliding window was used so that 100 samples made up one epoch. The raw data matrix of dimensions 62x100 was reshaped to get column vector which was sent as input. The proposed sinc net-R was a fusion of CNN and DNN, the CNN part comprised of 3 convolutional layers and DNN part consisted of 3 fully connected layers. In CNN part, the first layer involved convolution of input data with parameterized sinc functions and the second and third layer performed standard convolution task to extract features which were sent as input to fully connected layers of DNN. The achieved accuracy of proposed network was 94.5% for 3 classes.

Chen et al. [45] employed deep CNN for emotion classification. Amplitude of each time point and its normalization were taken as raw and norm time domain feature. For frequency domain features, 64 power spectral densities were computed by applying FFT on frequency bands and the overall signal from all 32 channels. The time and frequency domain features were joined together to form combined features. Three type of Deep CNNs were used for binary classification in valence and arousal dimension and achieved accuracy of around 99% for all the three models. Cheah et al. [46] used normalized EEG signals data from 32 channels as input to

CNN without extracting any features. The signal from each channel was divided into 5 discrete segments (10 seconds long) and then 1 second overlapping (90% sliding window overlap) sub divisions were extracted from within each segment in order to create additional training examples. The CNN was trained with 1 second overlapping divisions and achieved an accuracy of 97.59% for valence (3 classes) and 98.4% for arousal (3 classes). 5 temporal convolution layer and 3 spatial convolution layers were included in the CNN model followed by fully connected multilayer- perceptron network [46]. Ozdemir et al. [47] used 2D images of EEG signals as input to CNN. EEG signals were divided into fifteen seconds frames. Each frame of the EEG signals was split up into three frequency bands alpha, beta and gamma then average power spectrum of each band was computed using FFT, which were used as feature. Azimuthal Equidistant Projection was used to get two dimensional electrode locations. The obtained values for each band and 2D electrode locations were combined to get three topographic plots for each band by using Clough Tocher scheme. The three colored 2D images were used as input to CNN to achieve the accuracy of 95.96%, 96.09% and 95.90% for valence, arousal and dominance respectively (2 classes each). Cao et al. [48] employed CNN for emotion recognition using the data from four electrodes (F3, F4, P3, P4) from DEAP dataset. Data normalization (z-score normalization) was performed followed by Principal Component Analysis (PCA) to reduce the dimensionality of input data. Two-dimensional images of size 40*40 were sent as input to CNN to achieve the accuracy of 84.3±4.0% for valence and 81.2±3.0% for arousal (two classes each).

Many studies used spectrograms produced by EEG signals as input to the convolutional neural network [36]. Donmez et al. [49] classified EEG signals using spectrogram. The spectrograms were made using short time Fourier transform. The EEG signals were collected from 10 participants using single channel device. STFT was applied to extract 2D time frequency information. The values of STFT coefficients were squared to get the spectrogram. The input spectrogram images were transformed to 224x224x3 size to be sent as input to Google net pretrained neural network and the achieved accuracy was reported as 84.69% for 3 emotions named as fear, sadness, and fun. Zhao et al. [50] proposed a 3D convolutional network model for the classification of four quadrants (LALV, LAHV, HALV, HAHV) in the valence-arousal dimensional space. The data from electrodes was represented in 2D topological arrangement according to the position of electrodes in 10-20 positioning system to cater for the vanished topological location information of the electrodes. EEG signals were converted to 9x9x128 sized matrices, where 9x9 matrix showed 2D topological electrode arrangement and 128 represented the window size, were sent as input to the 3D CNN to achieve the classification accuracy of 93.53% and 95.86% in DEAP and AMIGOS datasets respectively.

| S/ No | Title/A uthor | Year | Dataset | Modal ity | Features | Classifier | Accuracy | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **Arousal** | **Valence** | **Dominan ce** |
| 1 | Cho et al. [28] | 2020 | DEAP | EEG | Spatio-temporal features (EEG stream) | 3D CNN | 98.74% (2 classes) | 99.28% (2 classes) | … |
| | | | | | | | 99.32% (4 classes) | | |
| 2 | Zhao et al. [50] | 2019 | DEAP/A migos | EEG | Spatio-temporal features | 3D CNN | 93.53% DEAP/ 95.86%AMIGOS (LVLA,LVHA,HVLA,HVHA) | | |
| **3** | Ozdem ir et al. [81] | 2019 | DEAP | EEG | 2D three colored images | Convolutional Neural Network (CNN) | 96.09%(2 classes) | 95.96% | 95.90% |
| **4** | Donme z et al. [49] | 2019 | 10 participan ts (Neurosk y) | EEG | spectrogram | Convolutional Neural Network (CNN) | 84.69%(for 3 emotions) | | |
| **5** | Cao et al. [48] | 2019 | DEAP | EEG | F3,F4,P3,P4/Norm alization /PCA | Convolutional Neural Network (CNN) | 81.2%(2 classes) | 84.3%(2 classes) | … |
| **6** | Cheah et al. [46] | 2019 | DEAP | EEG | Normalized data/1sec sliding window | Convolutional Neural Network (CNN) | 98.48%(3 classes) | 97.59%(3 classes) | … |
| **7** | Chen et al. [45] | 2019 | DEAP | EEG | time domain, frequency domain features and their combinations | Convolutional Neural Network (CNN) | 99%(2 classes) | | |
| 8 | Zeng et al. [80] | 2019 | SEED | EEG | temporal and spatial features | SincNet-R | 94.5% (3 classes) | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 9 | Chao et al. [43] | 2019 | DEAP | EEG | poewr spectrlal density features | Capsule Network(CapsNet) | 68.28% (2 classes) | 66.73%(2 classes) | 67.25%(2 classes) |
| 10 | Zhong et al. [79] | 2019 | SEED, SEED-IV | EEG | Differential Entropy features | Regularized Graph Neural Network (RGNN) | SEED 94.24%( negative, neutral, positive) | SEED-IV 79.37%(n eutral, sad, fear, happy ) | subject dependent |
| 11 | Yang et al. [42] | 2018 | DEAP | EEG | Spatial features extracted by CNN, temporal features extracted by RNN | Parallel Convolutional Recurrent Neural Network | 91.03% (2 classes) | 90.8% (2 classes) | … |
| 12 | Yang et al. [78] | 2018 | 5 participan ts | EEG | entropy measure extracted by Recurrence Quantification Analysis(RQA) | channel-frequency convolutional neural network (CFCNN), | 90.51% (3 emotional states) | | |
| 13 | Yang et al. [14] | 2018 | DEAP | EEG | Differential Entropy features | Convolutional Neural Network (CNN) | 90.24% (2classes) | 89.45% (2classes) | … |
| 14 | Granad os et al. [77] | 2018 | Amigos | ECG,G SR | time domain, frequency domain | Deep Convolutional Neutral Network (DCNN) | 76%(2 classes) | 75%(2 classes) | … |
| 15 | Song et al. [15] | 2018 | DREAM ER | EEG | frequency domain features(PSD, DE, DASM, RASM, DCAU) | Dynamical Graph Convolutional Neural Networks(DGCNN) | 84.54%(2 classes) | 86.23%(2 classes) | 85.02%(2 classes) |
| 16 | Salama et al. [41] | 2018 | DEAP | EEG | Spatiotemporal features | 3D CNN | 88.49%(2 classes) | 87.44%(2 classes) | … |
| 17 | Mei et al. [38] | 2017 | DEAP | EEG | Pearson correlation coefficient matrices | Convolutional Neural Network (CNN) | 85% (2 classes) | 78%(3 classes) | 75%(4 classes- LVLA,L VHA,HV LA,HVH A) |
| 18 | Goshva rpour et al. [76] | 2017 | Private | ECG | MP Coefficients extracted by wavlets(Coiflet, and Daubechies) and discrete cosine transform | Probabilistic neural network (PNN) | 100%(3 classes) | 100%(3 classes) | 100%(for 5 emotional states) |
| 19 | Yin et al. [75] | 2017 | DEAP | EEG | time domain, power features | multiple-fusion layer ensemble stacked autoencoder(MESAE) | 77.2%(2 classes) | 76.17%(2 classes) | … |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 20 | Zhang et al. [40] | 2017 | PhysioNet EEG Dataset | EEG | Spatial and temporal features | cascade/Parallel Convolutional Recurrent Neural Network | 98.31%/98.28% (multiclass intentions prediction) | | … |
| 21 | Li et al. [39] | 2016 | DEAP | EEG | Spatial features extracted by CNN, temporal features extracted by RNN | Convolutional Recurrent Neural Network (C-RNN) | 74.12%(2 classes) | 72.06%(2 classes) | … |
| 22 | Wang et al. [74] | 2013 | DEAP | EEG | raw EEG data | Deep Belief Network (DBN) | 60.9%(2 classes) | 51.2%(2 classes) | … |

**Table 3.2:** Literature Review of Classification Techniques

# CHAPTER 4: METHODOLOGY

This research work proposes a classification technique that employs two-dimensional images of EEG signals as input to convolutional neural network. AMIGOS Dataset has been used for emotion recognition from EEG signals. The dataset contains self-assessment of emotions by 40 participants who were being shown 20 videos (16 short and 4 long) as stimulus during the recording of physiological signals. These videos were of variable duration ranging from 51 seconds to 150 seconds. The physiological signals were recorded using low cost sensors. The recorded physiological signals were self-annotated by participants using Self-assessment Manikin (SAM) on a scale of 1-9. These self-assessed values of emotions for all the trials are presented in the dataset as valence, arousal, dominance, liking, familiarity and six basic emotions including Neutral, Happiness, Sadness, Surprise, Fear, Anger, and Disgust [60]. Some studies have used the basic emotions as classification labels while others have used valence, arousal and dominance values for two and three class classification techniques.

The most widely used approach for classifying emotions is training the classifier for each label one at a time. This type of approach is known as single-label classification [41]. Usually in emotion recognition, the emotions are classified as two (high and low) or three classes (high neutral and low) for valence, arousal and dominance dimensions. The other approach could be to perform classification of training data where each training example belongs to more than one class at the same time. This type of approach is known as multi-label classification [41]. For multi-label classification, the training examples are classified as four groups of valence and arousal.

Most of the classification techniques in the reviewed literature presented remarkable results based on binary classification of both valence (low/high) and arousal (low/high) [15, 31, 32, 39, 41-43]. However this research work proposes a CNN architecture for classification of EEG signals into four classes/labels. These labels are not present in the dataset so they are required to be derived from the given data. For this purpose the self-assessed values of arousal and valence are used for preparing the new labels. The two-dimensional valence arousal plane is divided into four quadrants named as Low Valence Low Arousal (LVLA), Low Valence High

Arousal (LVHA), High Valence Low Arousal (HVLA) and High Valence High Arousal (HVHA) [54]. Each trial is given a new label based on the values of valence and arousal dimensions and placed in one of the four quadrants. As the valence and arousal dimensions for each trial are assigned values using SAM on a scale ranging from 1 to 9, value of 5 is chosen as the cut-off to mark the valence and arousal dimensions as high and low. If the value of valence and arousal dimension for any trial is greater than 5 then it is taken as high valence (HV) and high arousal (HA) and if the value of valence and arousal dimension is lesser than 5 then it is marked as low valence (LV) and low arousal (LA) respectively. So for a particular trial if the values of valence and arousal are termed as LV and LA then for that particular trial the assigned label would be LVLA (first quadrant in Valence-Arousal dimensional plane). Similarly if the values of valence and arousal dimensions are LV and HA, HV and LA, HV and HA then the assigned labels would be LVHA, HVLA and HVHA respectively for that particular trial.

The two-dimensional valence arousal plane is shown in the figure below which shows that all the basic emotions lie in this Valence-Arousal dimensional plane and depend on levels/value of valence and arousal dimensions.

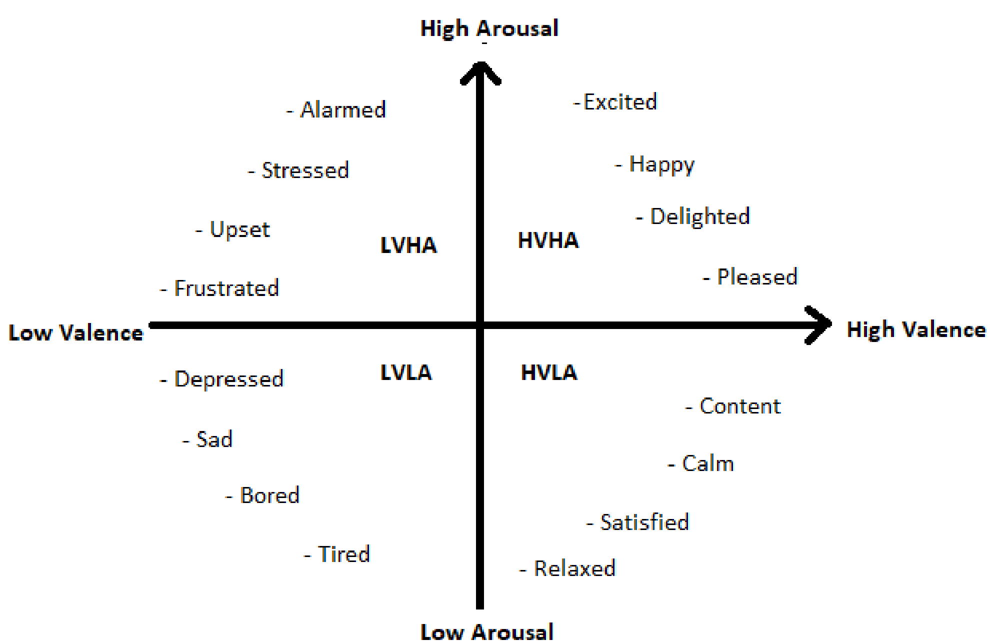



**Figure 4.1:** Valence Arousal Model for Representation of Emotions

## 4.1 Proposed Methodology

The overall methodology comprises of three main stages. The first stage comprises of specialized preprocessing of one dimensional EEG signals from all 14 channels in which the signals from all channels have been segmented in segments of one second length. The second stage involves the construction of two dimensional topological images from one dimensional preprocessed EEG signals. This conversion is required because the spatial information of EEG signals from the positions of electrodes in 10-20 system has been lost in the dataset and needs to be recovered for classification process and for improving the classification accuracy. In the third stage these two dimensional topological images are sent as input to the convolutional neural network for the classification of emotions in four classes. The four classes of labels have been derived from the values of valence and arousal dimensions.

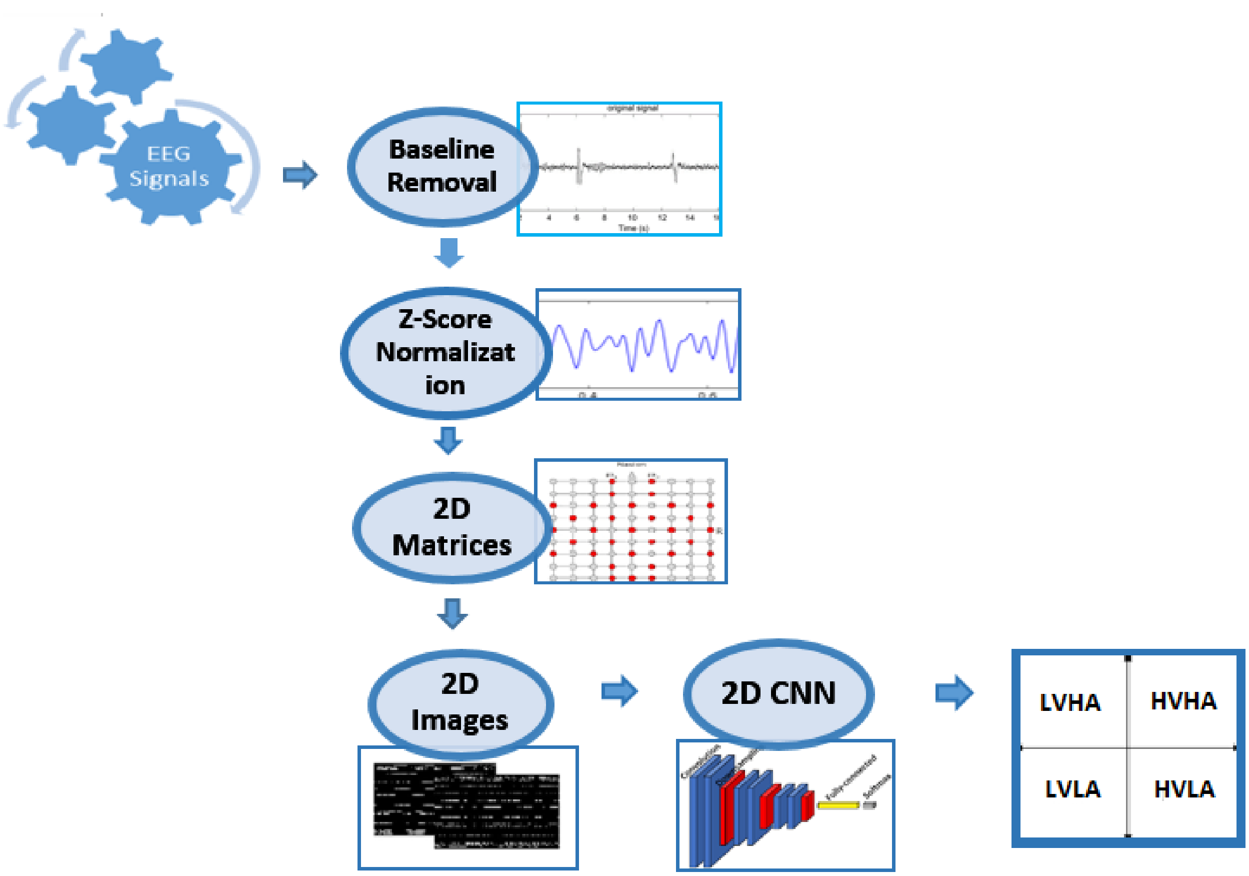



**Figure 4.2:** Block Diagram of Methodology

AMIGOS preprocessed data has been used in this research work for classification of emotions. The EEG signals recorded while participants were being shown short videos have been used in this research work. The preprocessed version of data has been used for further processing. This basic preprocessing involved averaging the recorded EEG signals to common reference and band pass filtering. The common average being calculated by computing the average signal of the EEG signals from all 14 channels and then subtracting this average signal from each of the 14 EEG channels. After averaging band pass filter was being applied with pass band frequency ranging from 4 to 45 Hz. These preprocessed EEG signals cannot be sent directly to the classification model and therefore require some specific preprocessing as well before the signals can be sent as input to the neural network. The second stage comprises of conversion of 14 channels of 1D preprocessed EEG signals into 2D images so that the benefits of 2D CNN can be exploited. All the three stages of the proposed methodology have been explained in detail under the following headings.

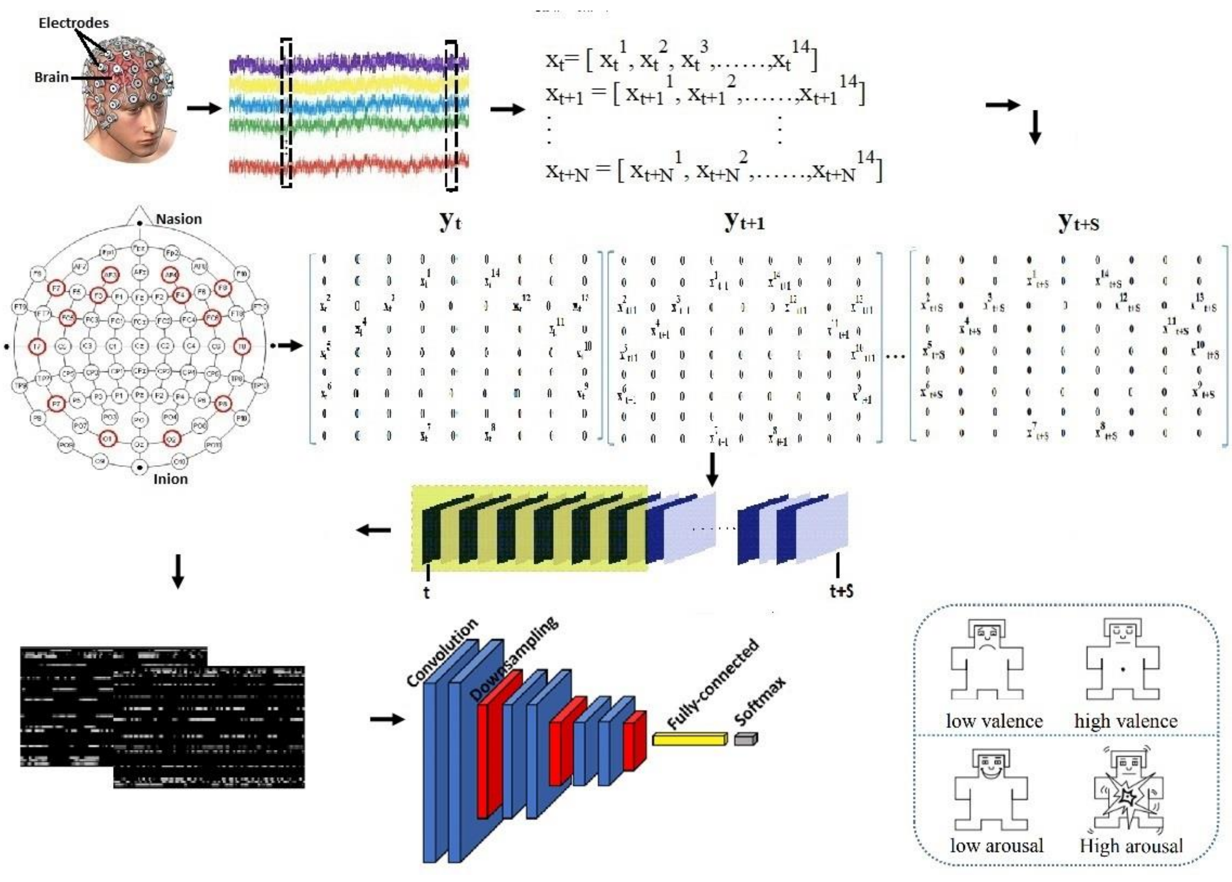



**Figure 4.3:** Proposed Methodology

### 4.1.1 Pre-Processing Stage

Preprocessing of data is an essential step for improving the quality of data and prepares the data for further processing. The main objective of preprocessing is that it suitably converts and scales the complete data and eliminates outliers that assists in the training process. This preprocessing stage is an essential step before any machine learning model can be trained. For enhancing the classification accuracy many preprocessing techniques have been proposed in literature. Most of the proposed methodologies in literature that employ CNN such as [39, 43, 47, 53, 54] use complicated preprocessing steps and extract hand crafted features which result in underutilization of neural network's capabilities. In many studies the EEG signals are directly used as input without considering the influence of baseline signals.

In order to improve classification accuracy, a preprocessing method which was proposed by yang [42] and other studies such as [55-59] has been employed. Yang [42] stated that the classification accuracy can be improved by nearly 32% by employing this preprocessing method on EEG signals. This method is computationally simple and inexpensive and finally converts the 1D EEG signals into 2D frames. This includes the removal of baseline signals from the actual data. The initial part of EEG signals is the baseline signal which is recorded when the stimulus is not being shown to the subject. It is the duration of EEG signal when there is no neural activity related to the stimulus. The removal of this baseline signal is essential otherwise it affects the recognition accuracy. The length of this baseline signal is different in different databases. For AMIGOS dataset, the duration of this baseline signal is 5 seconds. The preprocessing method involves subtracting the mean of baseline signal from the actual signals of all the channels. For this purpose first of all the baseline signals from all channels are separated and divided into N segments of 1s second each. Afterwards the mean of these 1 second long segments is calculated for each individual channel. The mean of these baseline signal segments for each channel is calculated as shown in equation below.

$$M\_BLsig = \frac{1}{N}\left(\sum_{1}^{N} BLsig\right) \quad (1)$$

The actual signal of each channel with length of 60 seconds each is also divided into S, 1 second long segments to get S number of C X L matrices where C represents the number of channels and L represents the length of each segment. The calculated mean of baseline signal for each channel is subtracted from each of the S segments of each signal for each individual channel as shown in equation below where PP_sig represents the preprocessed signal, Act_sig is the actual EEG data from which the mean of baseline signal segments is removed. This preprocessing step removes the effect of neutral emotional activity from the actual EEG signals used for recognizing emotions and hence improve the classification accuracy.

$$PP\_sig_i = Act\_sig_i - M\_BLsig \quad (2)$$

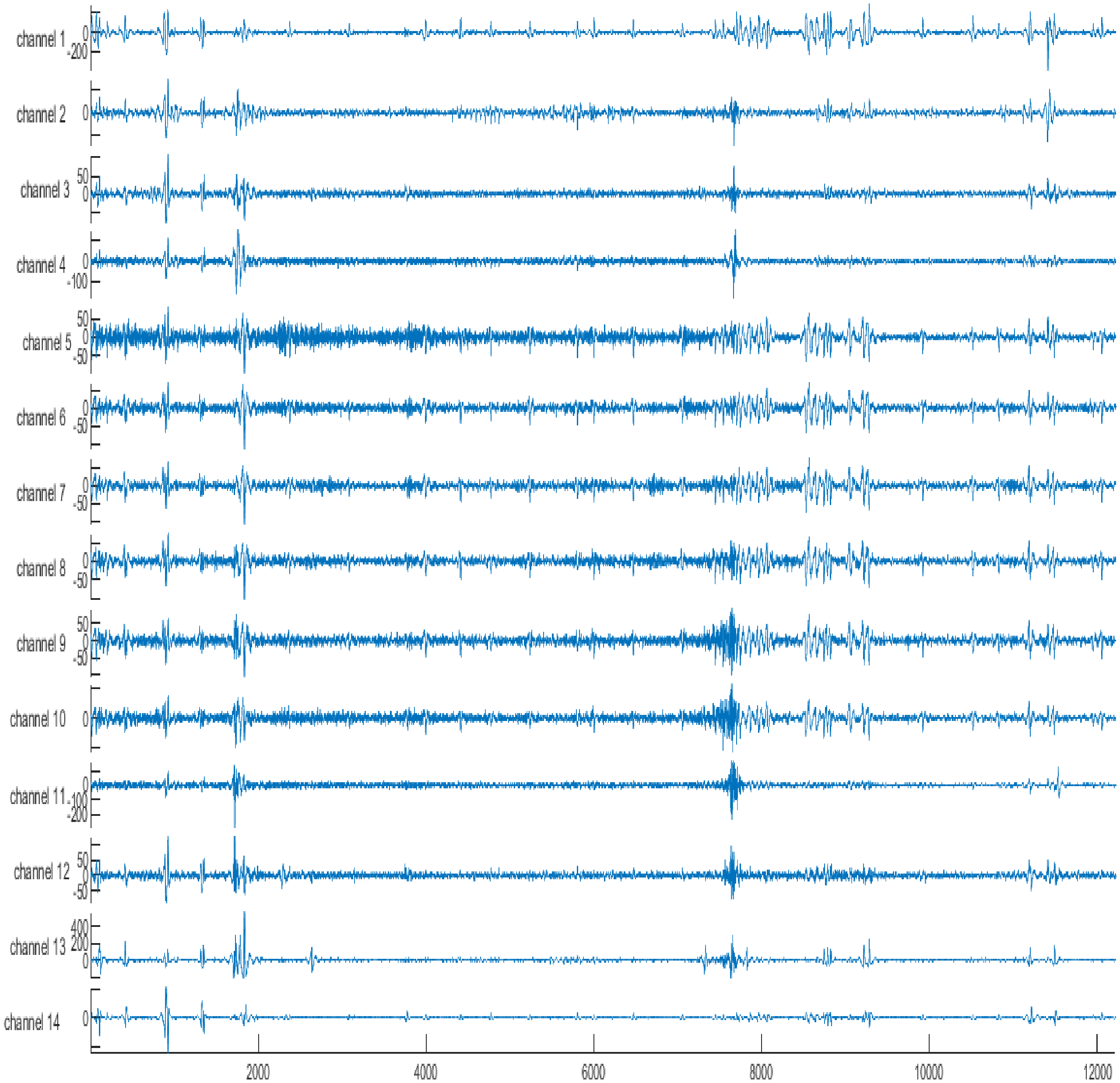



**Figure 4.4 (a):** Original EEG signals of all channels

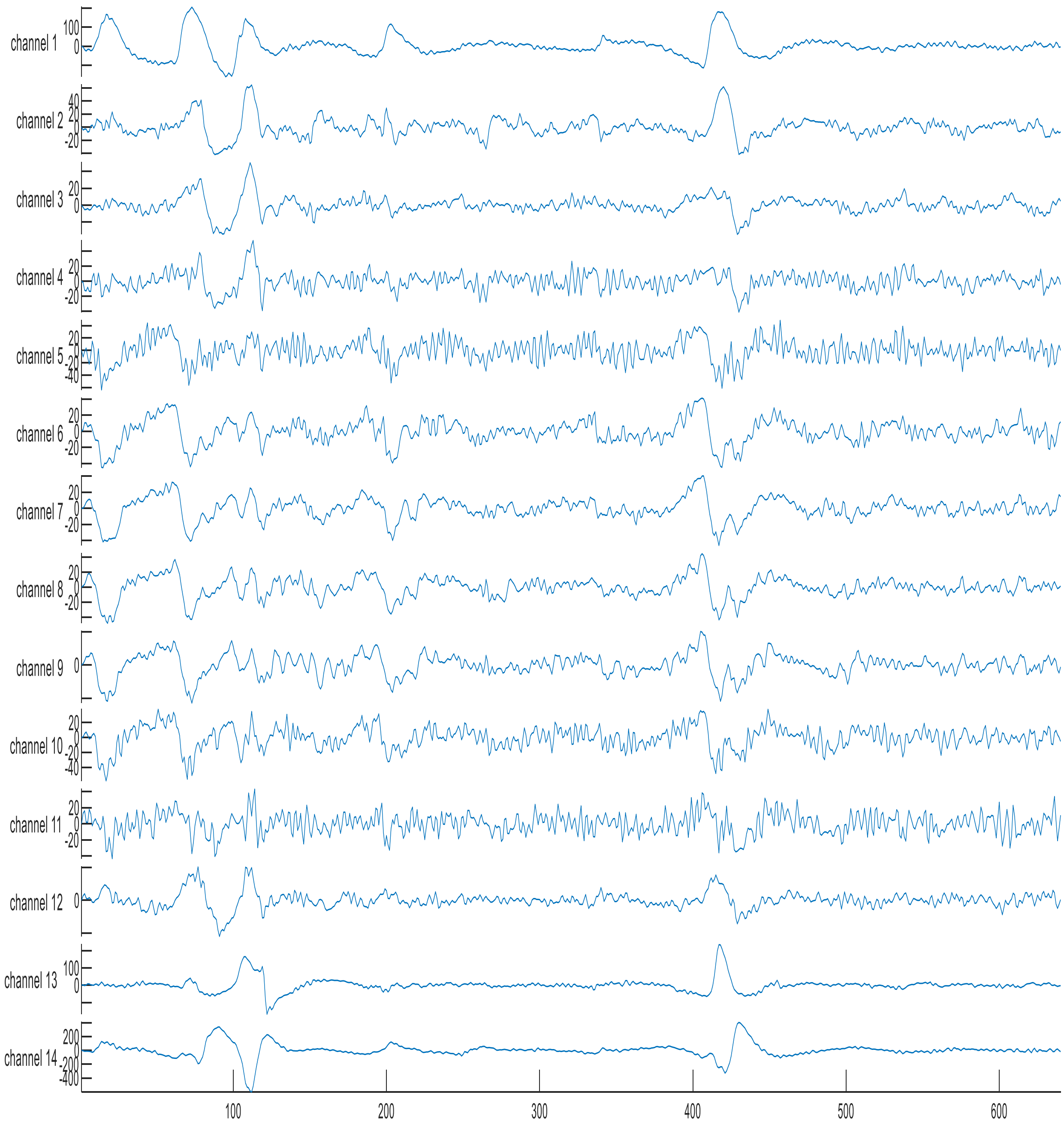



**Figure 4.4 (b):** EEG Baseline signals of all channels4.4

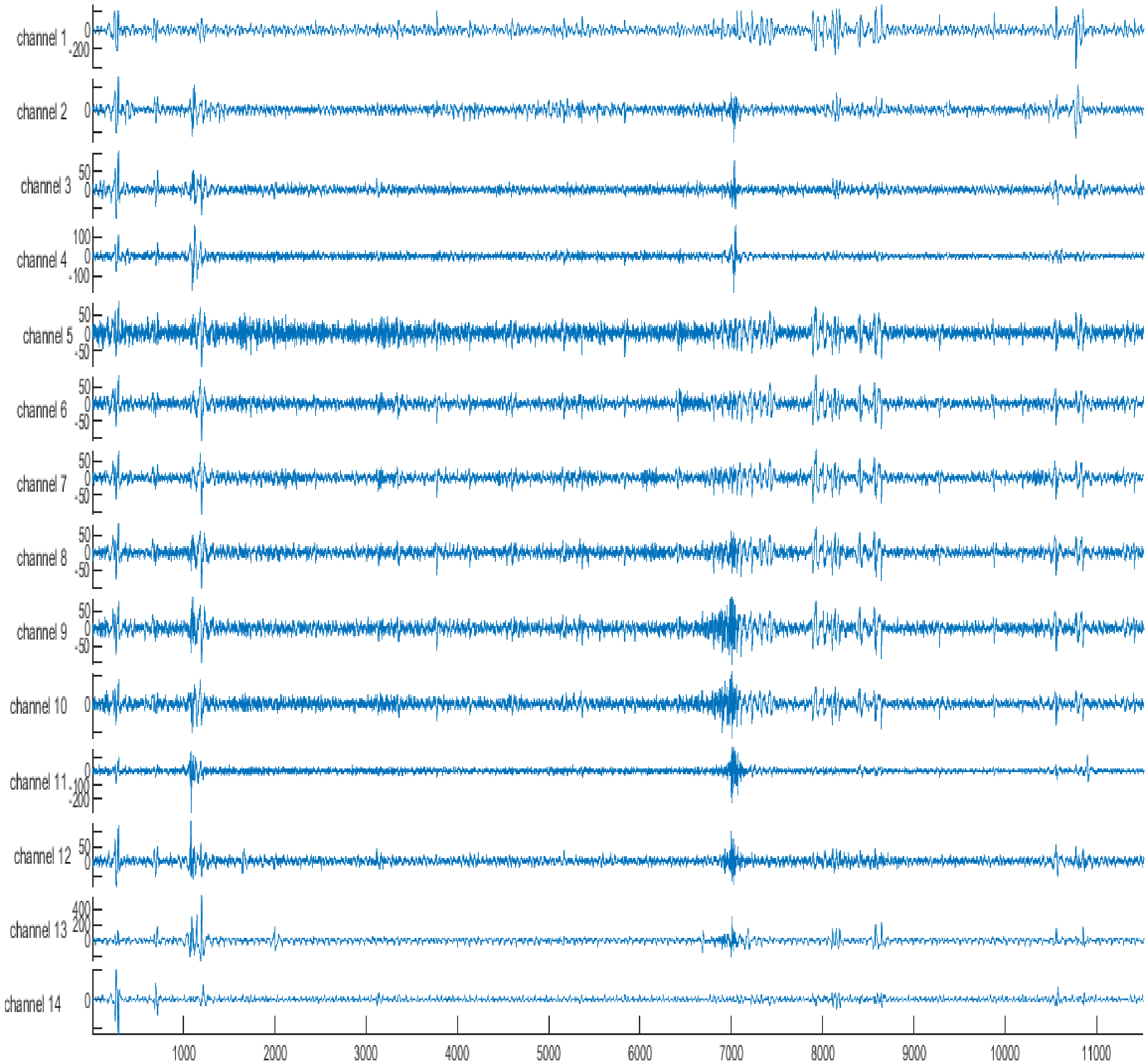



**Figure 4.4 (c):** Baseline removed EEG signals of all channels

### 4.1.2 Z- Score Normalization

In data mining, normalization of data is a very important step. Data normalization involves recognizing and processing the data within a database in a particular manner so that the database can be effectively utilized for analysis. Data normalization helps in removing the inconsistencies in data which make the analysis of data problematic. These inconsistencies and irregularities in data can adversely affect the data analysis and produce undesirable results. After the normalization of raw data the order of magnitude of data points becomes same which is appropriate for relative evaluation [48]. In the proposed methodology the EEG signals of all trials are normalized using z-score normalization to convert the EEG data to more usable form and remove the inconsistencies from the data. The mean and standard deviation of the original data are normalized by employing this technique of normalization. The EEG data after being Z-score normalized comes in accordance with the standard normal distribution which has zero value of mean and standard deviation of one [48]. The following equation is followed in performing Z-score normalization of the EEG data of all the trials.

$$V = \frac{x - mean}{\sigma} \tag{3}$$

In the above equation, x represents a non-zero value at any particular location in the matrix, mean represents the mean value of all non-zero values in the matrix and $\sigma$ represents the standard deviation of all values in the matrix. This z-score normalization is performed segment by segment for all channels while converting 1D EEG signals into 2D topological images.

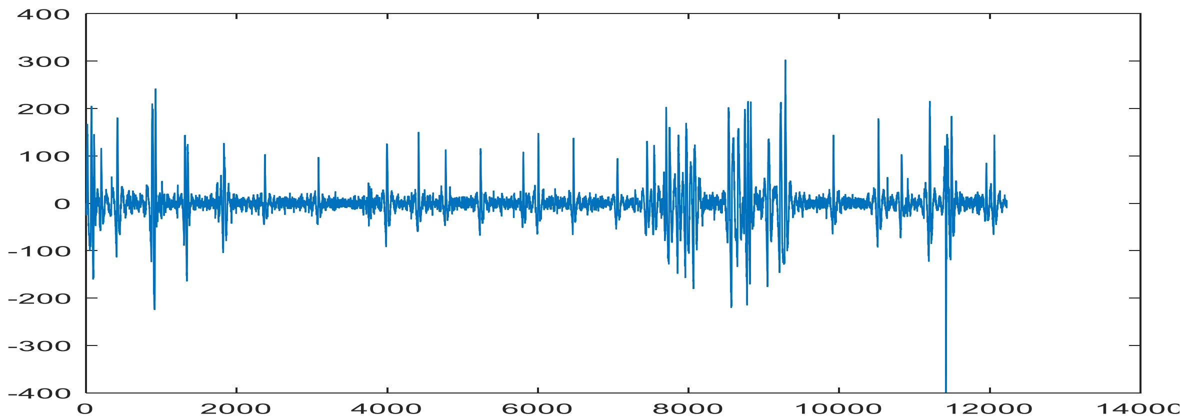



**Figure 4.5 (a):** EEG signal of one channel before pre-processing

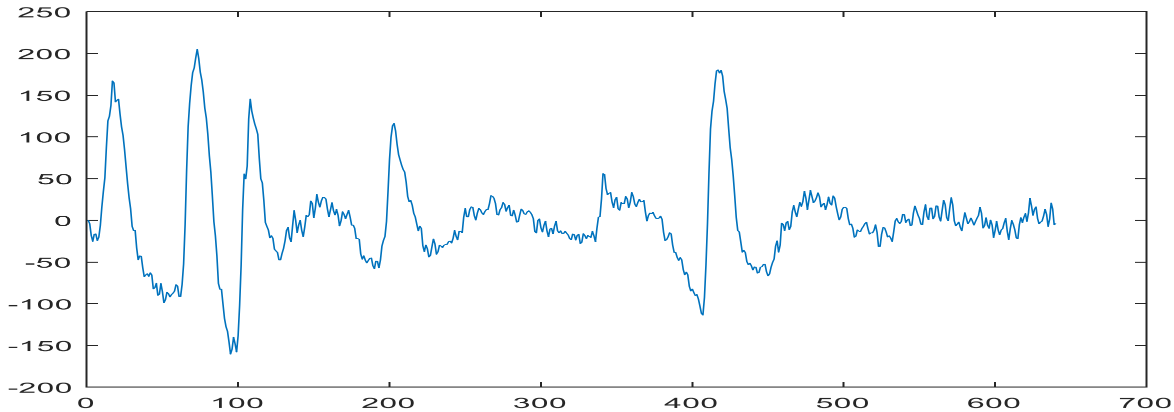



**Figure 4.5 (b):** 5s long baseline signal of one channel

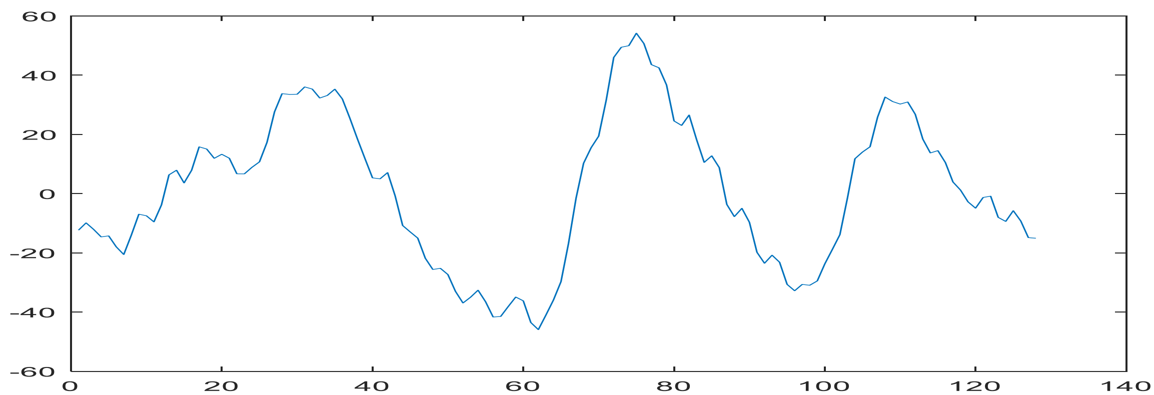



**Figure 4.5 (c):** Mean of five segments of baseline signal of one channel

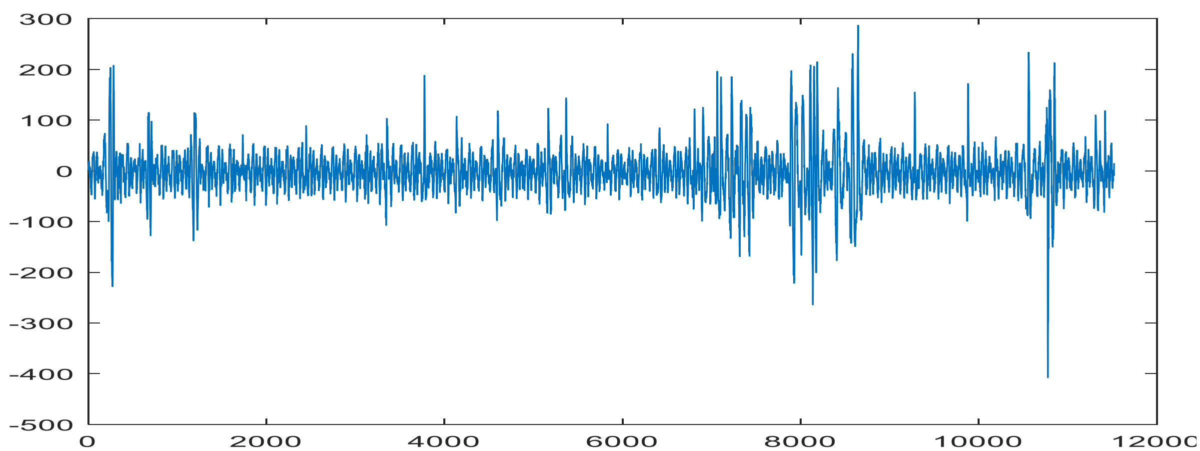



**Figure 4.5 (d):** Pre-processed baseline removed signal of one channel

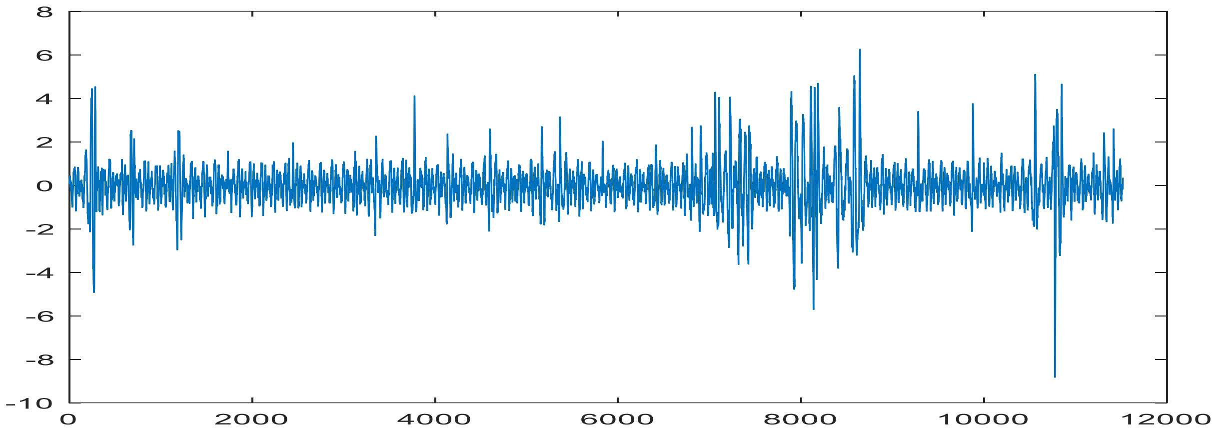



**Figure 4.5 (e):** Z-score Normalized Signal of One Channel

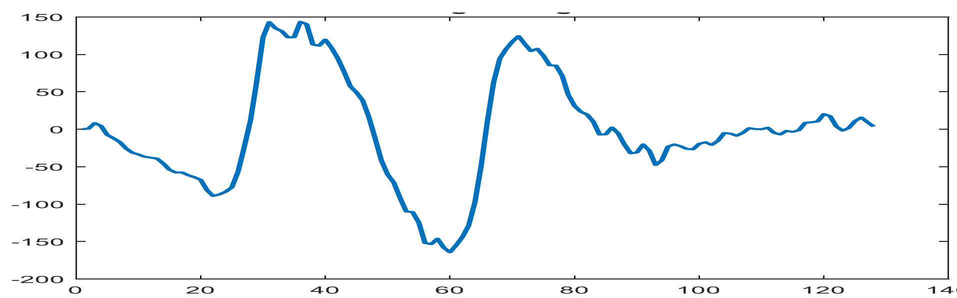



**Figure 4.6 (a):** EEG signal segment before pre-processing

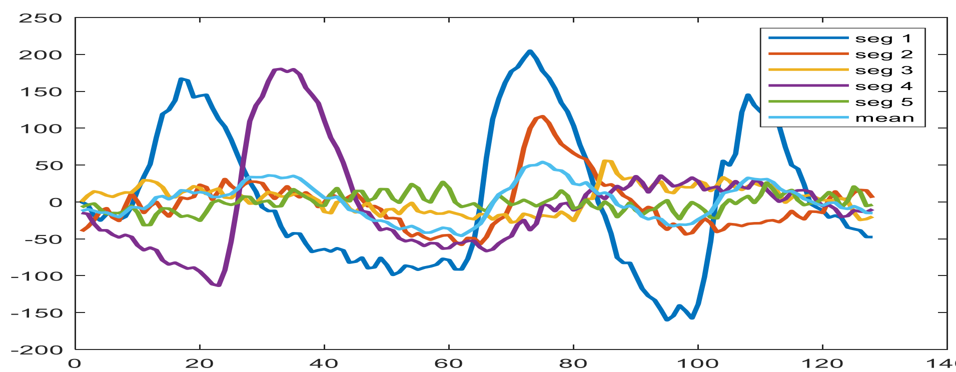



**Figure 4.6 (b):** Long Five Segments of EEG Baseline Signal and Their Mean

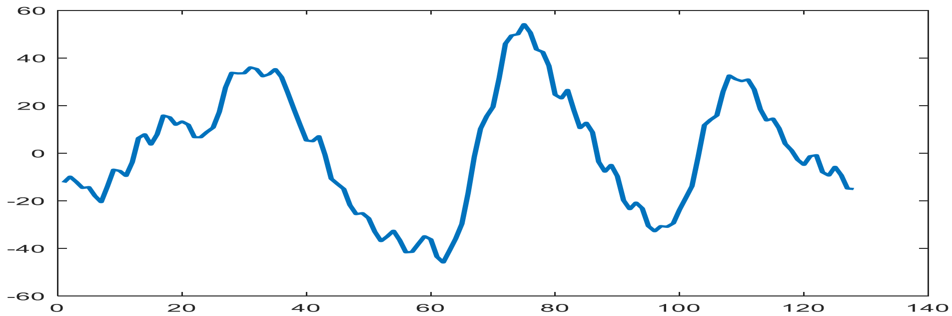



**Figure 4.6 (c):** Mean of Five Segments of EEG Baseline Signal

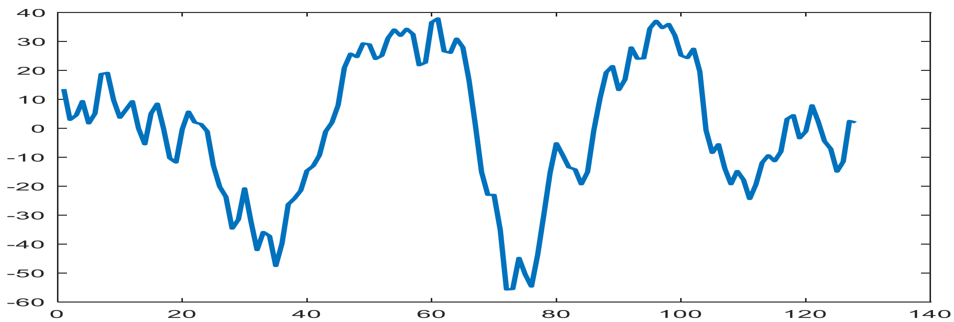



**Figure 4.6 (d):** Pre-Processed Baseline Removed Signal Segment

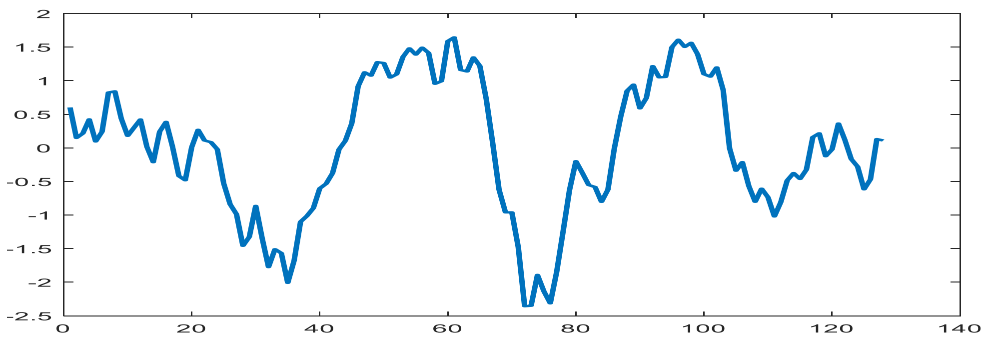



**Figure 4.6 (e):** Z-Score Normalized Signal Segment

### 4.1.3 Conversion of 1D Data to 2D Matrices

Machine learning approaches employing neural networks have proved to be very successful for emotion classification from EEG signals. Many researchers are attracted towards use of deep learning approaches for EEG based emotion recognition.

The EEG signal acquisition system employs wearable headsets that have several electrodes for acquiring the EEG signals from brain. The electrodes for capturing the brain waves have fixed positions and are placed according to internationally standardized positioning system called 10-

20 positioning system. In this positioning system the electrodes are placed at fixed locations of underlying regions of cerebral cortex. The numbers 10 and 20 of the positioning system specify that the definite distance between any two electrodes located adjacent to each other is either 10% or 20 % of the total front and back or left and right distance of the skull. The EEG signal data being captured by all the electrodes simultaneously at any particular moment of time can be shown in the form of 1D vector as

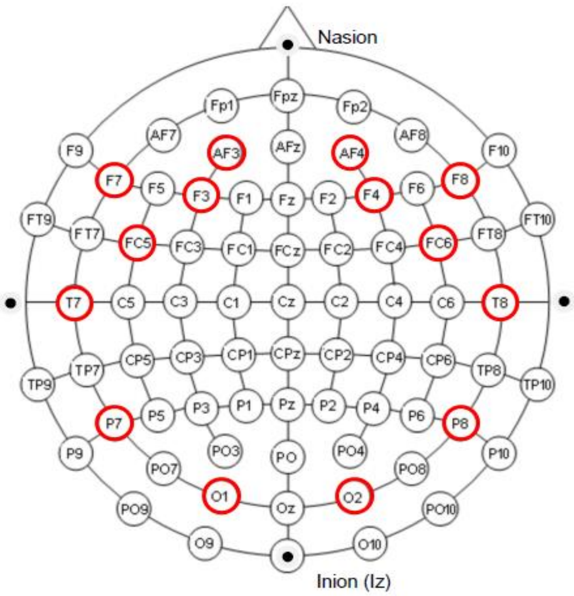
$$x_t = [\, x_t^1, x_t^2, x_t^3, \ldots\ldots, x_t^n \,] \tag{4}$$

where $x_t^i$ represents the data sample of ith channel recorded at time instant t when there are in total n number of electrodes in the EEG acquisition system. The number of electrodes used for recording EEG signals is 14 in case of AMIGOS dataset. The number of these 1D data vectors is (S + 1) which are generated by the EEG acquisition system during the time period of [t, t + S] [42]. Each of these 1D data vectors are composed of n elements as there are n number of electrodes in the EEG acquisition system. If the data of time interval of length L is to be represented then a 2D frame is formed by combining the sequential 1D vectors in that time interval L. The resultant 2D frame has L number of rows which represents no. of samples in the given time interval and C number of columns which represents total number of channels. In this 2D frame any channel has only two adjacent neighbors whereas in the 10-20 positioning system every electrode has a particular location and is surrounded by more than two neighboring electrodes which are positioned to record brain waves of particular underlying area of brain. Different areas of brain are responsible for different types of brain activities [40]. In this way the spatial information of adjacent electrodes on scalp is lost due to amalgamation of electrodes' data in 2D frames. This loss of spatial information adversely affects the performance of any BCI system. The recognition rates can significantly drop to low levels because of this loss of spatial information of electrodes.

Many researches have emphasized the significance of spatial information of EEG electrodes and are based upon converting the 1D EEG data to 2D frames which preserve the spatial information of electrodes. Wang [61] highlighted that the 2D matrices of format (channels x time samples) lost their topological position of electrodes and needed to be converted to some format that can preserve the relative positions of electrodes of the EEG acquisition system. For this reason the proposed methodology by Wang converted the 2D matrices to 3D tensors which preserved the topological positions of electrodes. Afterwards 3D CNN was used for extracting spatial as well as temporal features and for classification of emotions in two classes of valence (72.1%) and arousal (73.3%). Similarly Yang [42] proposed the construction of 2D frames from 1D vectors of EEG data in order to retain spatial information between the multiple channels neighboring each other. The stream of 2D frames were divided into segments using sliding window and sent to CNN for extracting spatial features. These spatial features were concatenated with temporal features extracted by RNN for final classification of two classes of valence (91.03%) and arousal (90.8%).

Zhang [40] also proposed similar conversion of 1D EEG signals to 2D meshes for classification of intentions by a hybrid model of CNN and RNN. Yuxuan [50] also employed conversion of 1D signals to 2D topological arrangement to cater for the vanished topological location information of the electrodes. EEG data was converted to 3D matrices of size 9x9x128 where 9x9 showed 2D topological arrangement and 128 denoted the window length. This 3D input was sent to 3D CNN architecture and achieved recognition accuracy of 93.53% and 95.86% for DEAP and AMIGOS datasets respectively.

The voltage oscillations generated in the brain as a result of stimulus are captured from the scalp by the electrodes of the acquisition system. Keeping in view the significance of spatial information of electrodes the chain like 1D EEG signals are converted to 2D matrix to preserve the spatial information of EEG electrodes according to their location in the 10-20 positioning system. 2D matrices from 1D signals of individual channels is formed based on the location map of electrodes in 10-20 positioning system. The 1D data vector $x_t$ at any instant of time t is converted to its corresponding 2D matrix according to figure 4.6.




Nasion

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C1 & 0 & C14 & 0 & 0 & 0 \\ C2 & 0 & C3 & 0 & 0 & 0 & C12 & 0 & C13 \\ 0 & C4 & 0 & 0 & 0 & 0 & 0 & C11 & 0 \\ C5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C7 & 0 & C8 & 0 & 0 & 0 \end{bmatrix}$$

Inion (Iz)

**Figure 4.7:** 2D Electrode Location Map

The resulting matrix is of 9x9 dimensions. The unused electrode positions are represented by zeroes in the matrix. The 14 EEG channels named as AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 [51] are represented by channel locations from 1 to 14 in 2D electrode location map. The resulting matrix $y_t$ at any particular instant of time t can be represented as

$$\mathbf{y_t} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_t^{1} & 0 & x_t^{14} & 0 & 0 & 0 \\ x_t^{2} & 0 & x_t^{3} & 0 & 0 & 0 & x_t^{12} & 0 & x_t^{13} \\ 0 & x_t^{4} & 0 & 0 & 0 & 0 & 0 & x_t^{11} & 0 \\ x_t^{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_t^{10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_t^{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_t^{9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_t^{7} & 0 & x_t^{8} & 0 & 0 & 0 \end{bmatrix} \qquad (5)$$

### 4.1.4 Creation of 2D Images

The total length of 1D EEG signals from each channel is divided into S number of segments of 1s each. The number of samples in each segment is 128. Therefore, the number of 2D matrices formed according to the positions of electrodes in 2D electrode location map is also 128 for each segment. These 128 2D matrices are processed together to form 2D images. Thus, in this way the 1D EEG vectors $(x_t, x_{t+1}, x_{t+2}, \ldots, x_{t+S})$ for the time interval [t, t+S] are converted to 2D matrices $(y_t, y_{t+1}, y_{t+2}, \ldots, y_{t+S})$. These 2D matrices are concatenated together to form 3D segment Wj which has the dimensions of 9x9x128 where 9x9 shows the dimensions of 2D electrode location map and 128 represents the length of one segment. The 3D chunk can be represented as

$$Wj = (y_t, y_{t+1}, y_{t+2}, \ldots, y_{t+S}) \tag{6}$$

This 3D chunk is then converted to 2D image in ".png" format which has dimensions of 81x128. For this conversion, EEG channels AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 in the 2D topological matrix are mapped to the corresponding indices of 4, 13, 19, 21, 29, 31, 37, 39, 47, 49, 55, 57, 67, and 76 of 1D vector respectively. The unused indices of 1D vector of length 81 are filled with zeroes. The resulting matrix after conversion has dimensions of 81x128 which is used as feature matrix. This feature matrix is considered as one example of a particular trial of a certain person. In this way all the1D EEG signals are converted into 2D topological images which are given as input to 2D CNN architecture for classification. The total number of images with dimensions of 81x128 generated from data of 33 persons (data of 7 persons containing invalid data is omitted) each having 16 trials is 45,474 which are used as the final feature matrices.

## 4.2 Classification

Convolutional neural networks are the most leading classification technique in the area of machine learning. Deep learning has become a dominating topic in artificial intelligence because of the prevailing and influential advantages of neural networks. Neural networks have gained popularity due to its remarkable ability to learn features without any human intervention.

Convolutional neural networks can provide powerful and promising results based on their ability of automated hierarchical feature extraction [61]. In addition to providing with remarkable accuracies, neural networks are also computationally efficient. The research on neural network has spiked due to the growing availability of EEG datasets and also due to the limitation of hardware being removed by influential and economical graphic processing units (GPU's) [62]. Deep learning based approaches have been employed increasingly for the classification of EEG signals which are commonly rated as having high dimensionality of data and low signal to noise ratios.

A neural network is combination of processes and algorithms which work step by step to identify the hidden relations of the incoming data. The neural networks have remarkable capability to adjust with varying input and produce finest results without remodeling the output requirements. For design of neural networks scientists were inspired by the biological neural networks. The working of a neural network resembles that of a human brain and consists of an arrangement of neurons which work systematically. A neuron in the neural network works like a mathematical function that categorizes the data according to the particular architecture. Just like the layered network of neurons in human brain, a neural network is composed of multiple layers of nodes that have specific interconnections. These interconnections have specific weights assigned to them. Initially all of these weights are set to arbitrary values and are repeatedly altered and fine-tuned during the training of neural network until similar output is obtained for the data having similar labels.

A CNN is composed of an input layer, an output layer and several hidden layers including convolutional layers, activation layers, pooling layers and fully connected layers. Typically every CNN architecture has repetition of convolutional layers and pooling layers and are responsible for learning features. These layers are followed by fully connected layers and the last fully connected layer performs the classification. These layers are explained in detail. The forward pass of all the data from the input layer to the output layer is termed as forward propagation. [63]

### 4.2.1 Convolution Layer

The most basic and important building block of CNN is the convolutional layer which is used for extracting the features. This layer performs operations which can be both linear and non-linear in nature. Convolution is a linear operation and convolutional layer consists of kernels of filters each of which is composed of small array of numbers. The input is also in the form array of numbers called a tensor [63]. Each of the filters is convolved with the input tensor by sliding the filter over the complete input tensor. While sliding the filter element wise product is calculated between sectors of input and the filter and these products are summed to get the final value for the corresponding location in the input. The convolution results are obtained in the corresponding locations in the output tensor which is known as feature map. This process of convolution is repeated with a number of kernels to obtain multiple feature maps which are representative of various features of the input.

The typical sizes of these kernels used in convolution operation are 3x3, 5x5 and 7x7 and the depth of resulting feature maps is defined by the number of these kernels used. The size of feature maps is always less compare to the input tensor because the center of each kernel never gets to be directly above the outermost number of input. Usually this problem is resolved by a technique called zero padding in which rows and columns of zeros are added all around the input tensor so that the center of kernel can overlap the outermost element of input tensor. The advantage of zero padding is that it does not allow the dimensions of successive feature maps to reduce and hence more convolution layers can be added in the particular architecture.

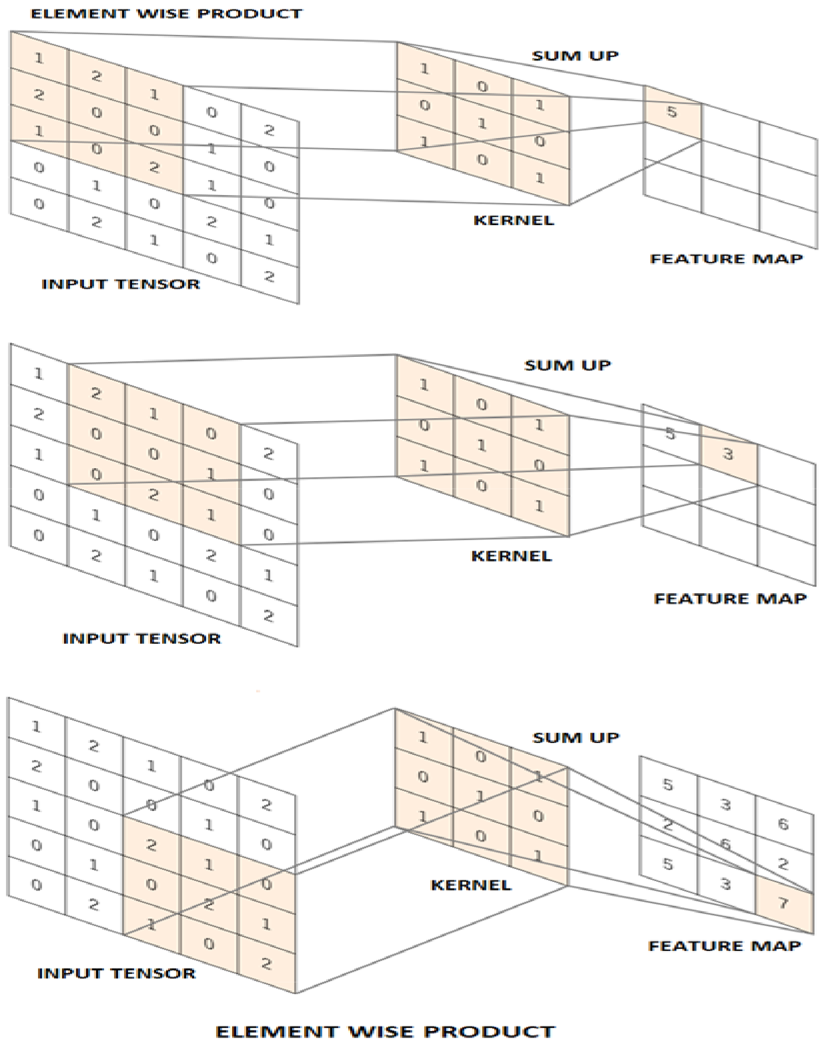



**Figure 4.8:** Convolution Operation with 3x3 kernel

Another variable which controls the size of feature maps is known as stride whose value is typically 1 but if the down sampled version of feature map is desired then greater values can be taken as stride specifies the distance between two consecutive locations of kernel while sliding over the input tensor. Max pooling can also be employed for the said purpose. An important part of training a CNN a model involves identifying the kernels that are most suitable for specific task and particular database for training. During the training process the kernels are learned automatically whereas the size and number of kernels along with the stride and padding needs to be specified before the start of the training process.

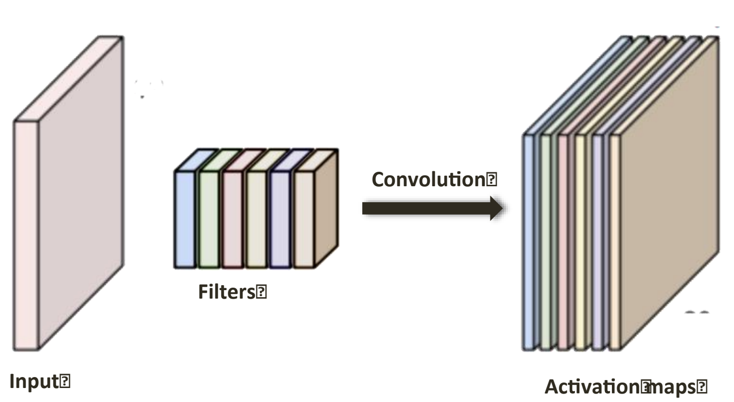



**Figure 4.9:** Activation Maps of Convolution Layer

### 4.2.2 Batch Normalization

The step of training of a deep neural network is a difficult and challenging task. These networks have large number of layers in them and the weights of the layers are initialized at the start. When the weights are updated, the distribution of layer's input deep inside the network may change after the pass of every mini-batch [66]. Due to this the algorithm keeps on running after a changing target endlessly. This alteration in the distribution of inputs to deep layers is termed as internal covariate shift [66]. When the parameters of a layer prior to the current layer change, the distribution of inputs of current layer is changed consequently and now the current layer has to realign according to the new distribution continuously. These small changes in distribution of inputs in shallow layers propagate through the network and produce large changes in deeper layers. Batch normalization can resolve this problem whereby the layer's inputs are standardized for each mini-batch. In this way the training process smooths out and the learning procedure is stabilized which decreases the number of training epochs drastically. Standardization means to rescale the data so that it has zero mean and standard deviation equal to one. Batch normalization speeds up the training process.

### 4.2.3 ReLU (Rectified Linear Units) layers

Conventionally, next to convolutional layer a non-linear activation layer is applied to introduce some non-linearity in the network that goes through linear operation of convolution that comprises of element-wise multiplications and summations in the convolutional layer. This non-linear layer assists in minimizing the overfitting in the system. This ReLU layer applies the function f(x) = max(0,x) [63] to all the input values. Non-linear functions such as sigmoid and hyperbolic tangent were used in the past because these characterize behavior of biological neurons mathematically but nowadays most commonly used function is the rectified linear unit. Since this ReLU function is efficient computationally, it accelerates the training process without deteriorating the accuracy of the network. The issue of vanishing gradient whereby the training of lower layers slows down due to the exponentially decreasing gradient through the layers, is also lessened by employing ReLU activation function. As shown in figure 4.10 the ReLU unit just converts all negative values of input volume to zero and enriches the network with non-linear properties without upsetting the receptive fields of convolution layer.

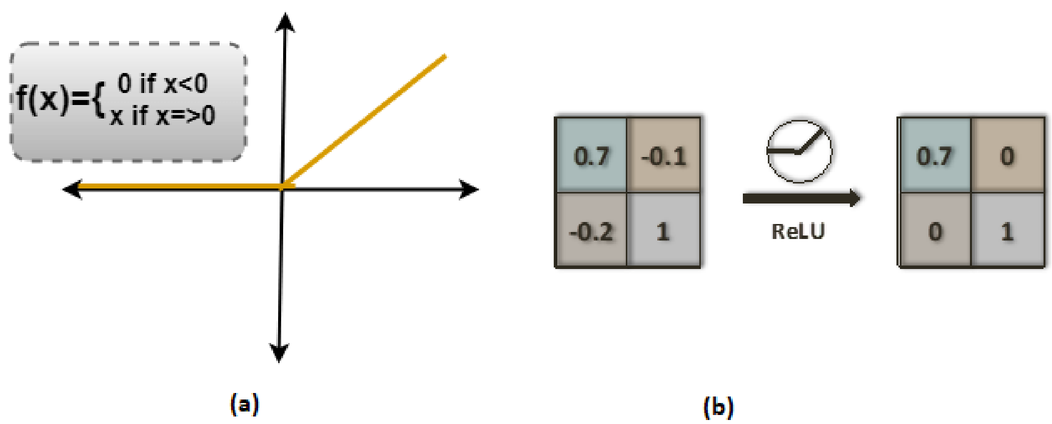



**Figure 4.10:** (a) ReLU actvation function. (b) ReLU operation in CNN

### 4.2.4 Pooling Layer

Pooling layer usually comes after the ReLU layer. Pooling layer performs the operation of down sampling as it reduces the dimensions of the feature maps. The reduction in dimensionality provides the advantage of decreasing the number of learnable parameters in successive layers. Different types of pooling layers including average pooling and L2-norm pooling can be used but most commonly employed pooling layer is the max pooling layer. For max pooling the filter is slid over the input which is divided into sectors and maximum value of all the values in the specific sector is retained only thereby reducing the dimensions of length and width of input volume. The depth of input volume is not altered by max pooling. The pooling layer does not have any parameters that are to be learnt but the hyper parameters such as size of filter, stride and padding need to be specified. Due to the reasoning that the location of a specific feature relative to the other features is of more importance as compared to the exact location of that feature, the max pooling layer is beneficial in two ways. Firstly it reduces the computation cost as the number of weights and parameters is reduced due to reduced dimensions and secondly, overfitting in the network is controlled. Overfitting refers to the idea that the model gives reduced accuracy for test set as compared to the accuracy of training set. In the proposed 2D CNN architecture the classification results are checked both with max pooling layer as well as with it. Figure 4.11 shows how max pooling is done.

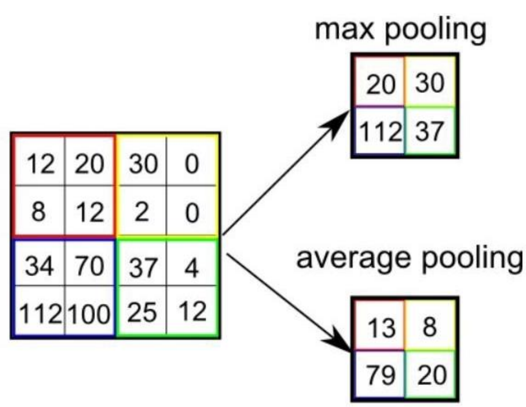



**Figure 4.11:** Comparison of Average and Max Pooling

### 4.2.5 Fully Connected Layer

Fully connected layers are applied towards the end of the network. The set of convolutional layer, ReLU layer and pooling layer is applied repeatedly in the network after which the fully connected layers are applied. The feature maps obtained from last convolution or pooling layer are one- dimensional arrays which are then sent as input to one or more than one fully connected layers. The neurons of every fully connected layer is connected as input to every neuron in the next fully connected layer. Every connection between two consecutive fully connected layers has a particular learnable weight. . Every neuron in the fully connected layer receives the input feature vector from the previous layer in the form of (x0, x1, x2,…, xn) as well as a vector of weights (w0, w1, w2,…, wn) which is learnt during the training process and the bias which is added to the weighted sum of input vector elements. The final feature maps being extracted from convolutional layers and pooling layers are sent as input through fully connected layers to the final output layer. The number of nodes in the final output layer is equal to the number of classes/labels in the training data.

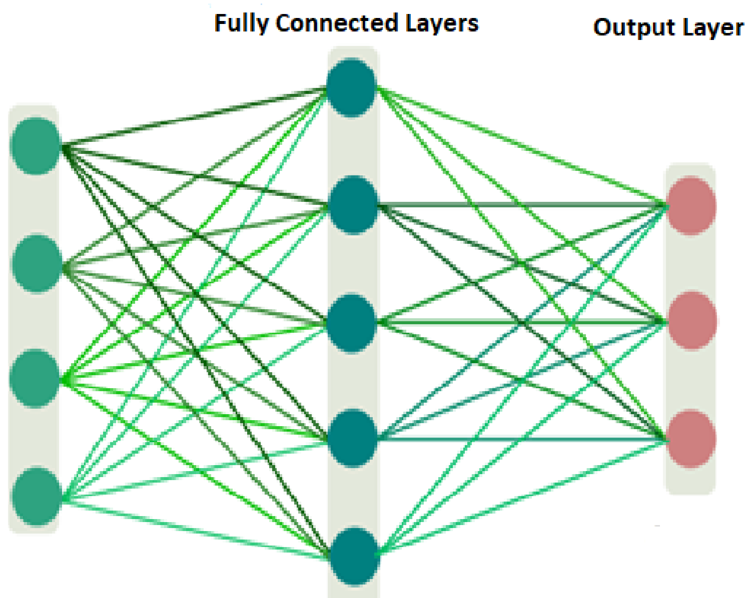



**Figure 4.12:** Fully Connected Neural Network Layers

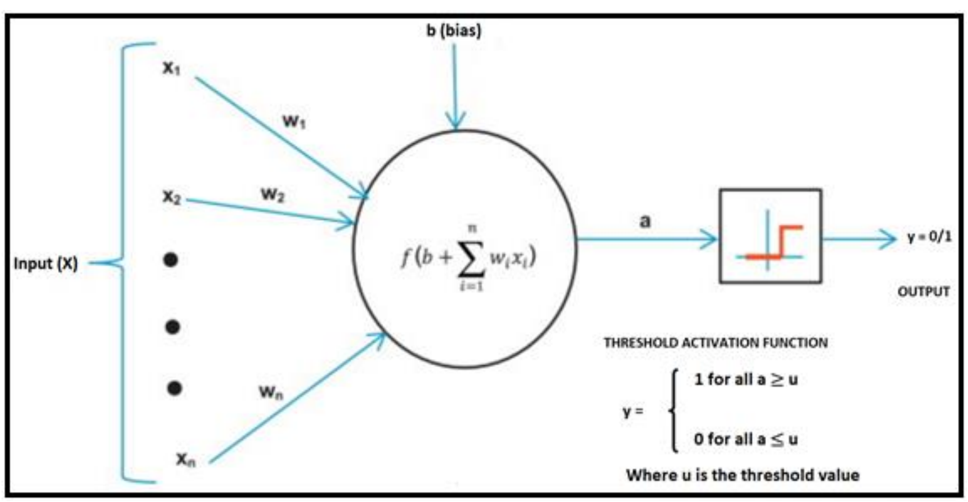



**Figure 4.13:** Representation of a Neuron in Fully Connected Layer

## 4.2.6 Softmax Activation Function Layer

The activation function applied to the last fully connected layer before the output layer needs to be carefully chosen which is suitable for the particular classification task. Typically sigmoid activation function is used for binary classification and softmax function is used for multiclass classification tasks. In a neural network, Softmax layer is applied before the output layer and it has similar number of nodes as that of the output layer. Softmax function converts the output of last fully connected layer into probabilities so that all the output values from the output layer add up to one [64]. Softmax function can be defined as a function that takes a vector of values as input and gives at output a vector of values that sum up to one. The values in the input vector can be negative, zero, positive or greater than 1 which are converted to values between 0 and 1 by the softmax function. This conversion is done by taking the exponent of each input value and normalizing each value by the sum of the exponents of all the input values. The output of the softmax layer is a vector which depicts the probability distribution of all the possible outcomes [64].

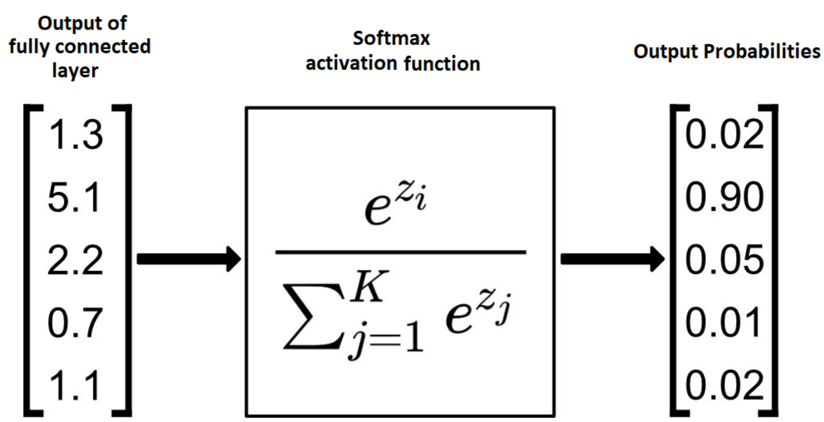



**Figure 4.14:** Softmax Activation Function

## 4.2.7 Training of a Neural Network

Training of a neural network is a process of finding the suitable weights for the neural connections in fully connected layers and finding kernels for the convolution layers by using a feedback mechanism [63]. For training purpose a training dataset is required which contains the training examples along with their labels. The weights and kernels are updated based on error between the predicted output and the actual labels. The algorithm that is usually used for training of a network is called back propagation where loss function and gradient descent algorithm play an important role. The outputs are predicted through forward propagation of the training dataset. During back propagation, the output value of the last layer which is calculated against specific kernels and weights is used to calculate the error value [67]. The weights of each neuron in the last layer are then updated based upon this calculated error value. The updated weights of this last layer are then used to calculate error value for the weights of preceding layer which are then updated based on their error value. This process of calculating error value and updation of weights is continued in the backward direction until it reaches first layer. Another optimization algorithm often used is called gradient descent. The process of training and updation of weights is shown in figure below.

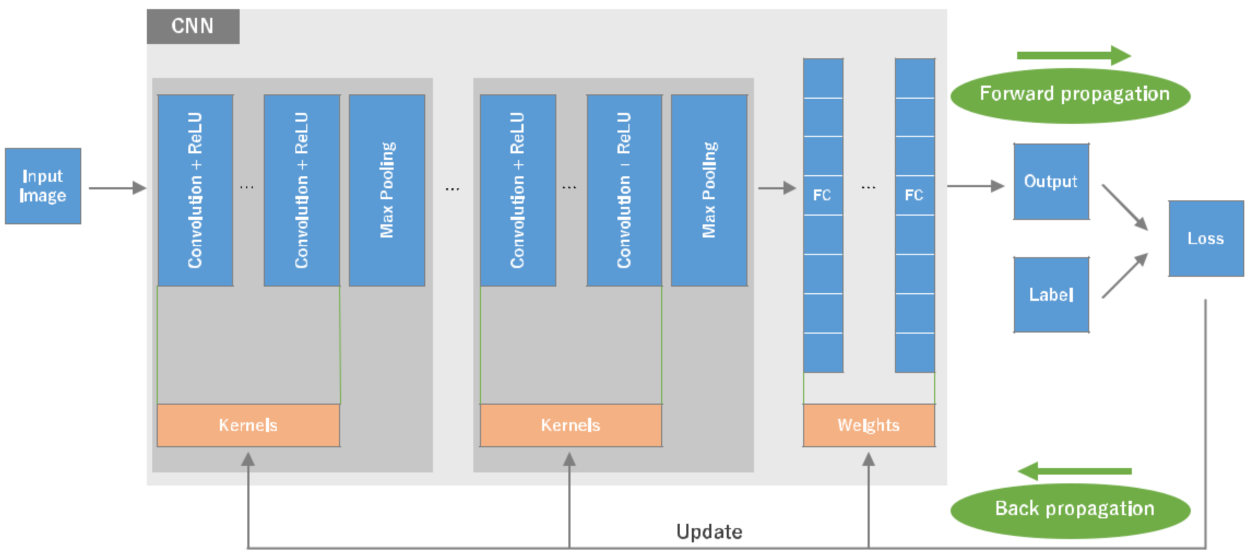



**Figure 4.15:** Training of a Neural Network

## 4.3 Architecture for 2D Topological Images

Convolutional neural networks are complex networks which are widely employed for classification of images. CNNs are feed forward networks in nature and used for recognition tasks as they provide high recognition accuracies. Scientists were inspired by visual perception of humans in identifying objects and hence proposed CNN which has a hierarchical layered architecture. The more the number of layers, the more the number of parameters to be learnt in the neural network and more will be computationally loaded training process of the network. However the CNN can reduce the time required for learning these parameters. CNNs have the ability to lessen the number of parameters without losing the quality of extracted features [65]. Images have greater number of pixels and therefore have high dimensionality which makes images suitable for classification by CNNs because of CNNs' remarkable ability to deal with high dimensional input data. Taking into consideration that CNNs are best suited for classifying images, one dimensional data is represented in the form of images and sent as input to CNN for exploiting CNN ability to deal with images. In the proposed methodology, 1D EEG signals are converted to 2D topological images and sent as input to the CNN for classification exploiting CNN's ability to deal with images. The detailed CNN architecture that is employed for classification of 2D EEG images is given in the figure below.

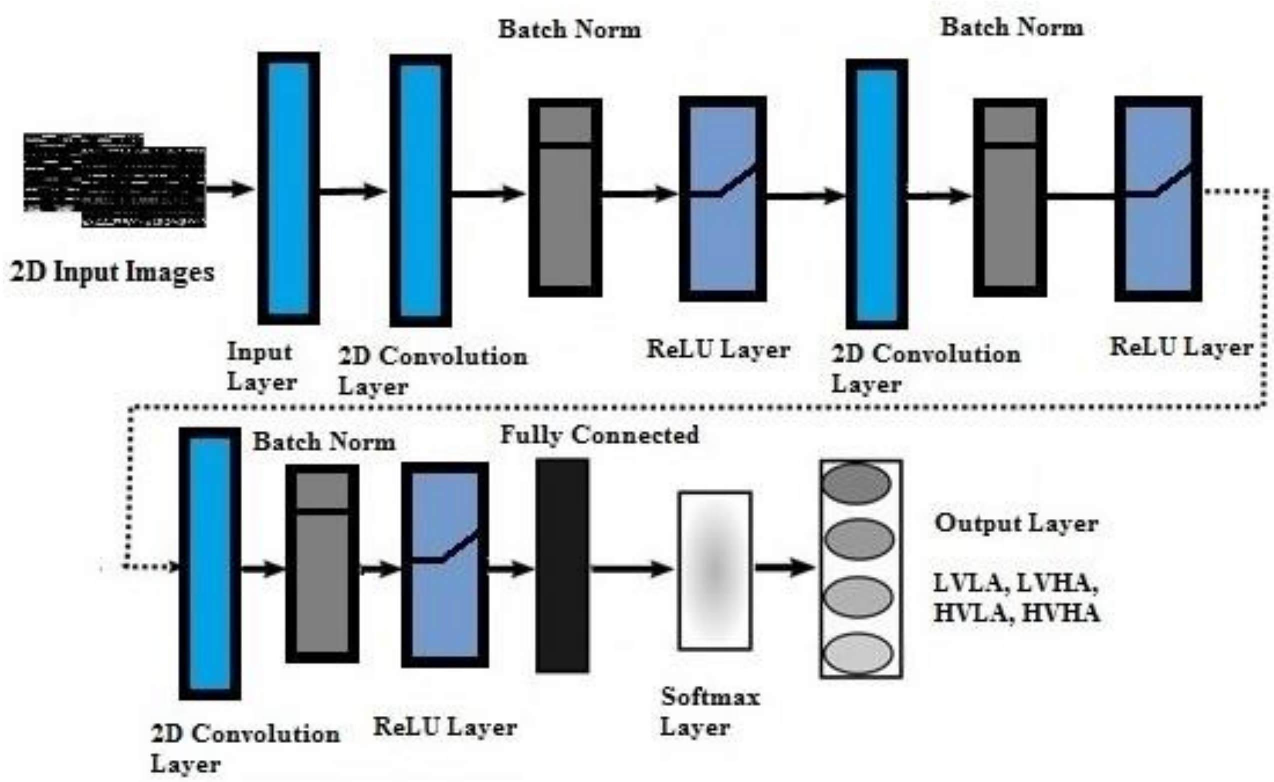



**Figure 4.16:** Proposed CNN Architecture

The input images that are sent to input layer have the size of 81x128. After the input layer comes the convolutional layer. This convolutional layer has 8 kernels each having size of 3x3. Each of the kernels is convolved with the incoming image to produce activation maps. After the convolution layer comes the batch normalization layer, then is ReLU layer which is followed by max pooling layer. The set of convolution layer, batch normalization layer, ReLU layer and max pooling layer is repeated three times in the proposed network architecture. The second and third convolution layers employ 16 and 32 kernels respectively with size of 3x3 each and same zero padding at the borders. All the three max pooling layers have size of 2x2 and are applied with stride of 2. The third max pooling layer is followed by fully connected layer. The output of fully connected layer is sent to SoftMax layer which is the last layer before the output layer. As suggested by Basha [68], the datasets which are wider in nature, that is, the datasets having more number of classes and lesser examples in each class need lesser number of fully connected layers in contrast to datasets which are deeper in nature that is, the datasets having lesser number of classes and more examples in each class, for the 2D network architecture. Since the dataset used for the current research work is not very deep therefore only one fully connected layer is employed in the proposed CNN architecture. The details of parameters for each layer are described in the table below. The images that are sent as input to the CNN are created by taking 1s long segments each containing 128 samples of all 14 channels of EEG signals. The total number of parameters for each of the convolutional layers and fully connected layers are mentioned in the table 4.1 below.

| S/ No. | Layers | Activations | Weights | Biases | Total Parameters |
|---|---|---|---|---|---|
| 1. | 2D Image Input | 81x128x1 | - | - | 0 |
| 2. | Conv | 81x128x8 | 3x3x1x8 | 1x1x8 | 80 |
| 3. | Batch norm | 81x128x8 | 1x1x8 | 1x1x8 | 16 |
| 4. | ReLU | 81x128x8 | - | - | 0 |
| 5. | Conv | 81x128x16 | 3x3x8x16 | 1x1x16 | 1168 |
| 6. | Batch norm | 81x128x16 | 1x1x16 | 1x1x16 | 32 |
| 7. | ReLU | 81x128x16 | - | - | 0 |
| 8. | Conv | 81x128x32 | 3x3x16x32 | 1x1x32 | 4640 |
| 9. | Batch norm | 81x128x32 | 1x1x32 | 1x1x32 | 64 |
| 10. | ReLU | 81x128x32 | - | - | 0 |
| 11. | Fully connected | 1x1x4 | 4x331776 | 4x1 | 1,327,108 |
| 12. | SoftMax | 1x1x4 | - | - | 0 |
| 13. | Ouput | - | - | - | 0 |

**Table 4.1:** Details of 2D CNN Architecture

# CHAPTER 5: EXPERIMENTAL RESULTS

## 5.1 Database

The proposed methodology is being assessed by using AMIGOS dataset which is an offline dataset used for research based on physiological signals. This newly published dataset is widely used in recent literature based on emotion classification from physiological signals. For this dataset two types of experiments were conducted to collect the data, one with 16 short videos and the other one with 4 long videos. In the first experiment 40 participants watched videos individually while in second experiment participants watched videos individually as well as in groups. The short videos had variable duration between 51 second to 150 seconds [52] and long videos had duration of longer than 14 mins. Physiological signals of participants such as Electroencephalogram (EEG), Electrocardiogram (ECG), and Galvanic Skin Response (GSR) were recorded for both configurations by means of wearable sensors. Emotiv Epoc Neuroheadset [51] was used to record EEG signals at 128Hz sampling frequency whereas Shimmer sensor was used to record ECG and GSR data at 256Hz and 128Hz sampling frequency respectively.

For the first experiment participants did their self-assessment for valence, arousal, dominance, liking, familiarity and seven basic emotions (Neutral, Disgust, Happiness, Surprise, Anger, Fear and Sadness [60]). Valence dimension shows the level of negativity or positivity, arousal shows the level of calmness or excitement and dominance shows level of submissiveness or authority. Valence, arousal and dominance levels were rated by using Self-Assessment Manikin (SAM) on a 9-point scale. For valence 'negative' was rated as '1' and 'positive' as '9', for arousal 'calm' was rated as '1' and 'excited' as '9' and for dominance 'submissive' was rated as '1' and 'in command of emotions' as '9'. Familiarity was rated on a scale of 1 to 9 with 'never seen' taken as '1' and 'already seen' as '9'. Liking was also rated from 1 to 9 with 'dislike' rated as '1' and 'like' rated as '9'. External assessment of participants' levels of valence and arousal was also made for both the experiments. The videos of participants from front were recorded while they watched the stimuli videos 20 in number and cropped to show only head and neck areas. These recorded videos were then split up into 20 second clips and shown to annotators who rated them on valence and arousal continuous scales ranging from -1 to 1 for low valence/arousal and high valence/arousal respectively.

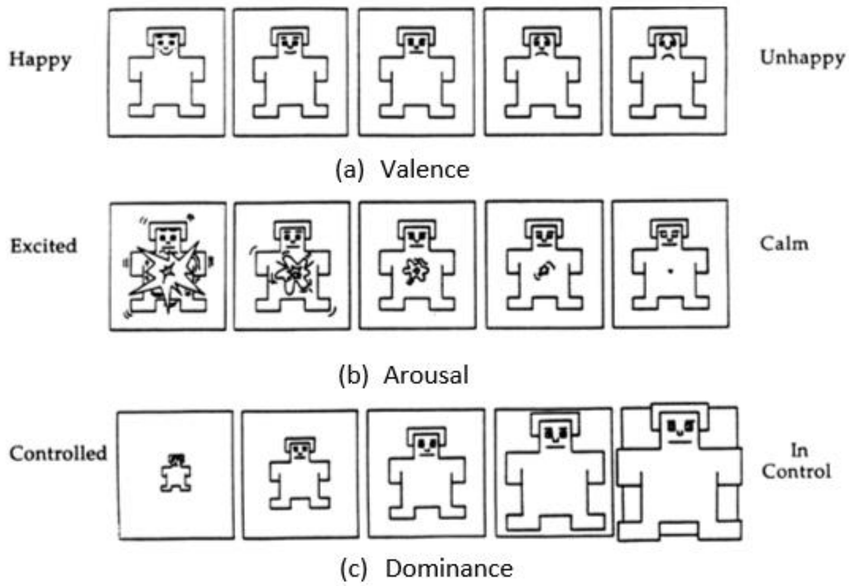



**Figure 5.1:** Self-Assessment Manikin (SAM) scale (1-9) for (a) valence (b) arousal (c) dominance

The dataset contains EEG signals from 14 channels and peripheral physiological signals from three channels which include ECG Right, ECG Left, and GSR. The dataset contains both the original data recordings as well as the pre-processed data. The length of recorded signals varies with the length of stimuli videos and for each trial the recorded signal contains baseline signal of 5s duration which was recorded when no stimulus was being shown to the participant [50].

The preprocessed data after being down sampled to 128Hz is averaged to common reference and then passed through a band pass filter with pass band of 4 to 45 hertz. The preprocessed data is present in 40 .zip files in matlab format one for every participant. Preprocessed data from the dataset has been used in this research work. Each file of preprocessed data consists of three 1x20 matrix lists named as joined_data, labels_selfassessment and labels_ext_anotation. The cell named as joined_data contains 20 matrices for each video and each matrix contains data of EEG, ECG and GSR samples where the number of samples depends upon the duration of video. Columns 1 to 14 of each matrix contain data samples corresponding to 14 EEG channels which are named as AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 [51] whereas 15$^{th}$, 16$^{th}$ and 17$^{th}$ columns contain data of EEG right, EEG left and GSR signals respectively.

The cell named as labels_selfassessment contains twenty 1x12 matrices for each of the trial videos where each of the 12 columns contain the values for arousal, valence, dominance, liking, familiarity, neutral, disgust, happiness, surprise, anger, fear, and sadness respectively. Columns for arousal, valence, dominance, liking and familiarity contain floating values ranging from 1 to 9 whereas the columns for basic 7 emotions contain values of 1 or 0 depending upon if that emotion is selected by the participant or not. The data contained in each of participants' .mat files has been summarized in the following table.

| List name | Resulting matrix shape | Matrix contents |
|---|---|---|
| joined_data{1,YY} | XX x 17 | samples (depends on the duration of the video) x channels for trial YY |
| labels_selfassessment{1,YY} | 1 x 12 | 1 x label (arousal, valence, dominance, liking, familiarity, neutral, disgust, happiness, surprise, anger, fear, and sadness) for trial YY |
| labels_ext_anotation{1,YY} | ZZ x 3 | segments (20 second clips) x channels (segment_index, valence and arousal) for trial YY |

**Table 5.1:** Details of Pre-processed Data

## 5.2 Experimental Results

Most of the emotion recognition studies employing physiological signals perform binary classification of high and low on valence and arousal dimensions. This research work not only involves the classification of emotions in two classes of each Valence and Arousal dimensions but also involves classification in three and four classes.

For classification purpose, the 2D input images are divided randomly into training set and test set. 70% of the total data is kept for training of the CNN architecture and the rest 30% is used for testing purpose. During the training process the classification accuracy for every mini-batch of the training set is calculated to monitor the progress of training process. The smoothed (less noisy) version of this training classification accuracy is also calculated by employing a smoothing algorithm. Along with this the training loss for each mini-batch is calculated during the training process for monitoring the training progress.

The subsets of training set are used to update weights by computing the gradient of the loss function and these subsets of training set are called as mini-batches. The number of times it takes for all the training examples to pass through the algorithm is called iteration. An iteration is completed when a mini-batch is passed through the algorithm and every iteration marks a step headed for minimizing the loss function in the gradient descent algorithm because the loss is calculated for every mini-batch. The training data is divided into mini-batches which pass through the algorithm one by one and when the complete training data has passed through the algorithm once then one epoch is completed. The number of epochs specified for the training process are equal to 15. Different number of epochs up to 100 were used for the training of the network but it is observed that the network starts to over fit the training data as the number of epochs are increased. The training data is shuffled for every epoch during training. When the training data is divided into mini-batches then the last mini-batch may not have same amount of data as that of the rest of the batches. This shuffling of data prevents the same data from being discarded in every epoch. During the training process of the network, the validation data is used for calculating the validation accuracy as well as the validation loss. The classification accuracy and classification loss which is calculated using the validation data is termed as validation accuracy and validation loss respectively. For this purpose the validation frequency is specified to be 30. This validation loss is termed as cross entropy loss since the emotion classification problem is a multiclass classification problem. The validation data is also shuffled before each time the network validates itself.

An optimization algorithm is used to optimize the network parameters in every iteration. The regular gradient descent algorithm is normally employed for updation of network parameters which minimizes the loss function at every iteration by finding the gradient of the loss function using the complete dataset at once. On the other hand, the stochastic gradient descent calculates gradient and performs updation of network parameters by utilizing different subset (mini-batch) of training data for every iteration. This mini-batch size is variable and has to be specified at the start of training process of the CNN. The parameter updates of stochastic gradient descent evaluated by using mini-batches of training data are a noisy approximation of parameter update that would be computed using complete training data at once. This algorithm of stochastic descent is oscillatory in nature and for reducing these oscillations a momentum term is added to the parameter update. So the stochastic gradient descent with momentum ‘sgdm’ is used as an optimizer during the training process. With ‘sgdm’ optimizer, different values of learning rate $\alpha$ equal to 0.01, 0.001 and 0.0001 is used in the training process and best results are obtained for $\alpha = 0.001$. Same training parameters are used for classification in two, three and four classes.

### 5.2.1 Experimental Results for Two Classes

The proposed CNN model is trained for classification of emotions in two classes. The Valence and Arousal dimensions have self-assessment values starting from 1 to 9. Each dimension having continuous values is divided into two classes. The Valence dimension is separated into two classes named as Low Valence (LV) with values ranging from 1 to 5 and High Valence (HV) with values ranging from 6 to 9. Similarly the arousal dimension is divided into two separate classes named as Low Arousal (LA) ranging from 1 to 5 and High Arousal (HA) ranging from 6 to 9. The classification results are shown by the confusion matrices for both Valence and Arousal dimensions. Table 5.2 shows the distribution of examples/instances in two classes for Valence and Arousal dimensions. The total number of images is 45,474. Table 5.3 shows the number of instances in each class when 50% overlapping sliding window is used and the total no. of images created is 90,420. Figure 5.2 and 5.3 show the confusion matrices of Valence and arousal dimensions respectively. Figure 5.4 and 5.5 show the confusion matrices of Valence and Arousal dimensions respectively with 50% overlap combined with interpolation.

| **Classes** | **LV** | **HV** |
|---|---|---|
| **Valence Values** | $1<x\leq5$ | $5<x<9$ |
| **No. of instances** | 22,505 | 22,969 |

**(a)**

| Classes | LA | HA |
|---|---|---|
| Arousal Values | 1<x≤5 | 5<x<9 |
| No. of instances | 22,431 | 23,043 |

**(b)**

**Table 5.2 (a):** No. of Instances in Two Classes of Valence Dimension. **(b):** No. of Instances in Two Classes of Arousal Dimension.

| Classes | LV | HV |
|---|---|---|
| Valence Values | 1<x≤5 | 5<x<9 |
| No. of instances | 44,742 | 45,678 |

**(a)**

| Classes | LA | HA |
|---|---|---|
| Arousal Values | 1<x≤5 | 5<x<9 |
| No. of instances | 44,598 | 45,822 |

**(b)**

**Table 5.3 (a):** No. of Instances in Two Classes of Valence Dimension with 50% Overlap. **(b):** No. of Instances in Two Classes of Arousal Dimension with 50% Overlap

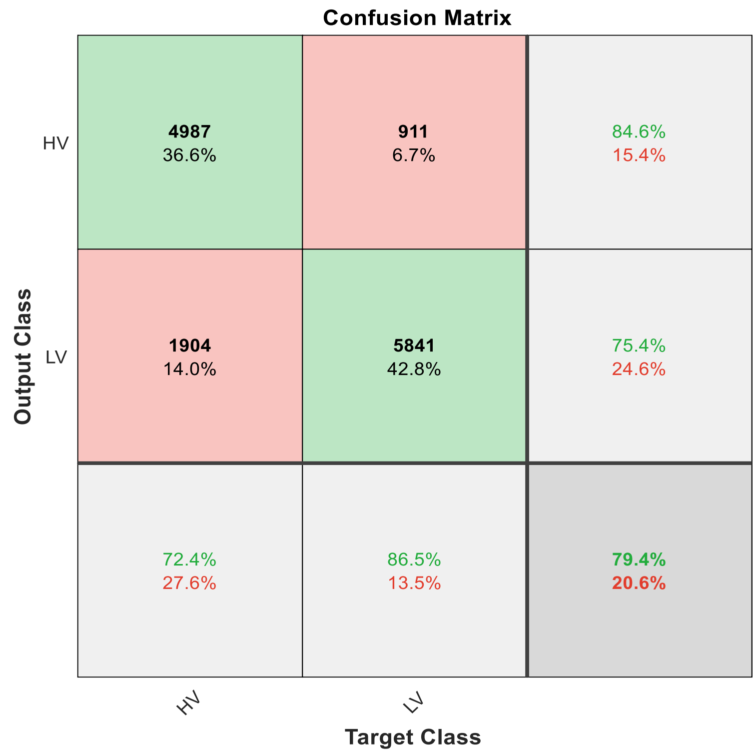



**Figure 5.2:** Confusion Matrix for 2 Classes of Valence Dimension

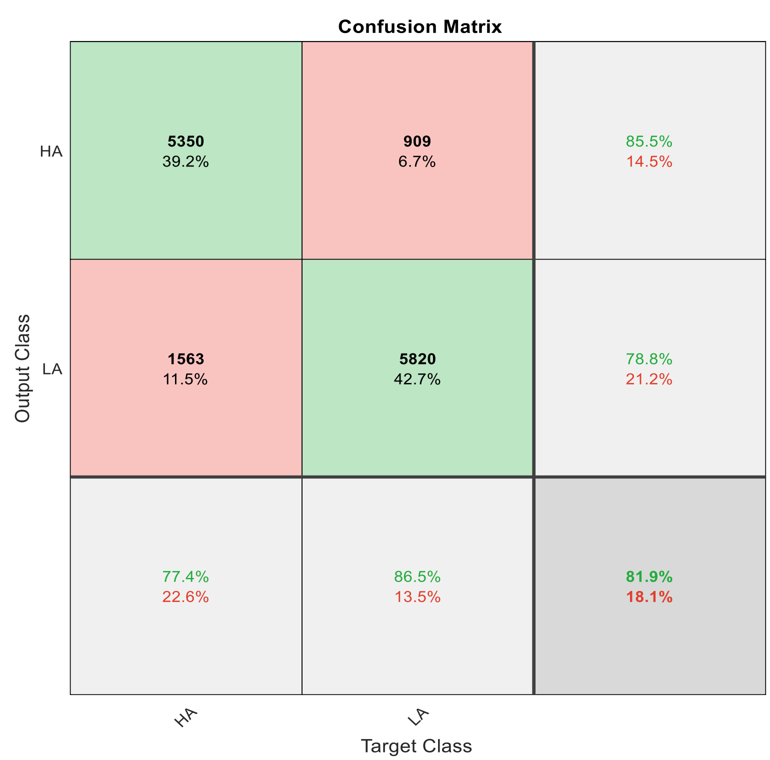



**Figure 5.3:** Confusion Matrix for 2 Classes of Arousal Dimension

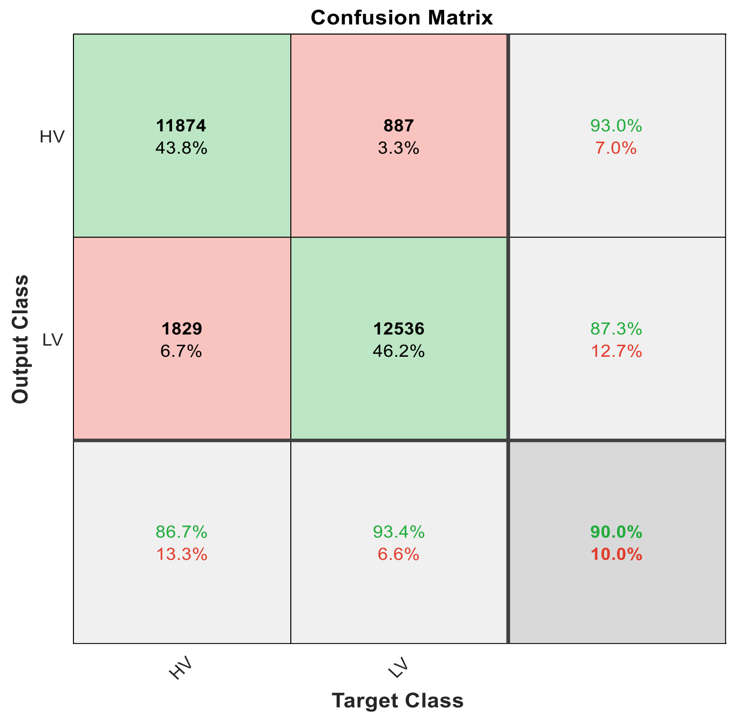



**Figure 5.4:** Confusion Matrix for 2 Classes of Valence Dimension with 50% overlap + Linear Interpolation

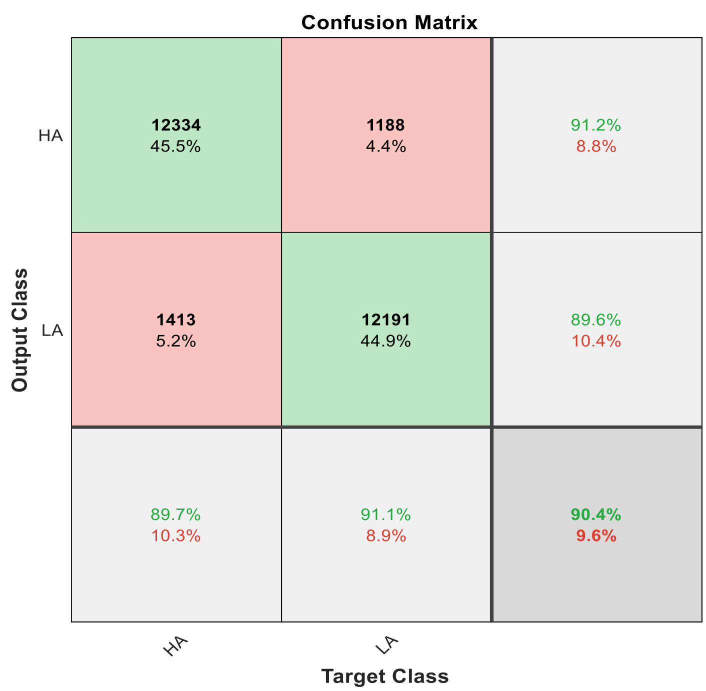



**Figure 5.5:** Confusion Matrix for 2 Classes of Arousal Dimension with 50% overlap + Linear Interpolation

### 5.2.2 Experimental Results for Three Classes

The CNN model is also trained for three classes**.** The three classes are obtained by segmenting each of Valence and Arousal dimensions in three separate ranges. The valence dimension is split up into three classes named as Low Valence with values ranging from 1 to 4, Neutral with values ranging from 4 to 6 and High Valence with values ranging from 6 to 9. Similarly for Arousal dimension, the values ranging from 1 to 4 are termed as Low Arousal the values of 4 to 6 are termed as Neutral and the values ranging from 6 to 9 are termed as High Arousal. The classification results are shown by the confusion matrices for both Valence and Arousal dimensions. Table 5.4 shows the number of instances in each of the three classes of both Valence and Arousal dimensions. Table 5.5 represents the number of instances in three classes of both Valence and Arousal with 50% overlap. Figure 5.6 and 5.7 show the confusion matrices of Valence and arousal dimensions respectively. Figure 5.8 and 5.9 show the confusion matrices of Valence and Arousal dimensions respectively with 50% overlap combined with interpolation.

| **Classes** | **LV** | **Neutral** | **HV** |
|---|---|---|---|
| **Valence Values** | $1<x\leq 4$ | $4<x<6$ | $6\leq x<9$ |
| **No. of instances** | 16,702 | 8,247 | 20,525 |

**(a)**

| **Classes** | **LA** | **Neutral** | **HA** |
|---|---|---|---|
| **Arousal Values** | $1<x\leq 4$ | $4<x<6$ | $6\leq x<9$ |
| **No. of instances** | 14,274 | 13,307 | 17,893 |

**(b)**

**Table 5.4 (a):** No. of Instances/Examples in Three Classes of Valence Dimension. **(b):** No. of Instances/Examples in Three Classes of Arousal Dimension.

| **Classes** | **LV** | **Neutral** | **HV** |
|---|---|---|---|
| **Valence Values** | 1<x≤4 | 4<x<6 | 6≤x<9 |
| **No. of instances** | 33,208 | 16,392 | 40,820 |

**(a)**

| **Classes** | **LA** | **Neutral** | **HA** |
|---|---|---|---|
| **Arousal Values** | 1<x≤4 | 4<x<6 | 6≤x<9 |
| **No. of instances** | 28,381 | 26,455 | 35,584 |

**(b)**

**Table 5.5 (a):** No. of Instances/Examples in Three Classes of Valence Dimension with 50% Overlap. **(b):** No. of Instances/Examples in Three Classes of Arousal Dimension with 50% Overlap

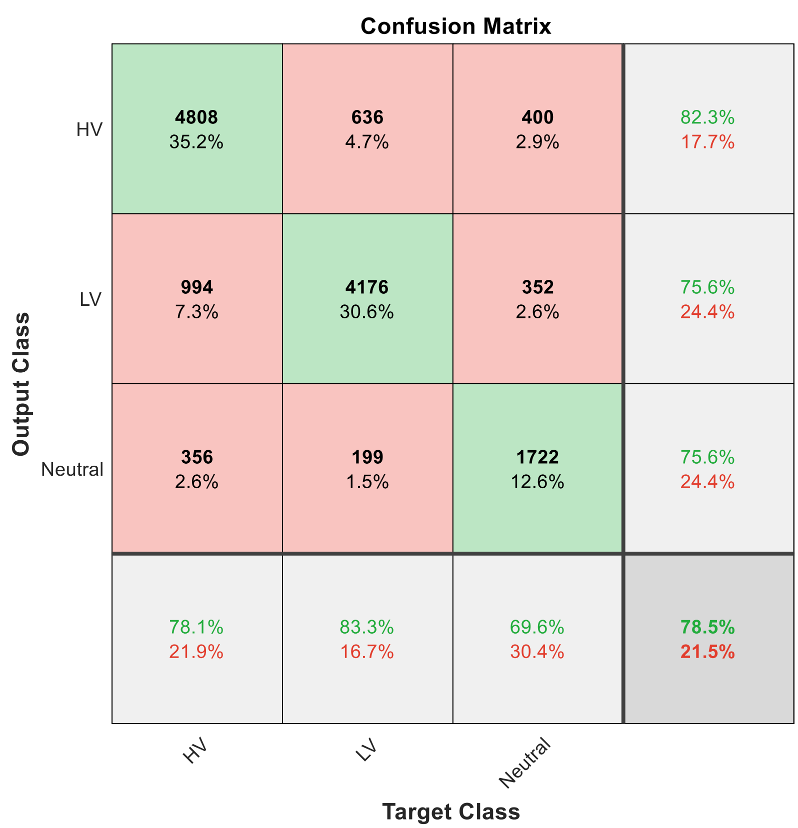



**Figure 5.6:** Confusion Matrix for 3 Classes of Valence Dimension

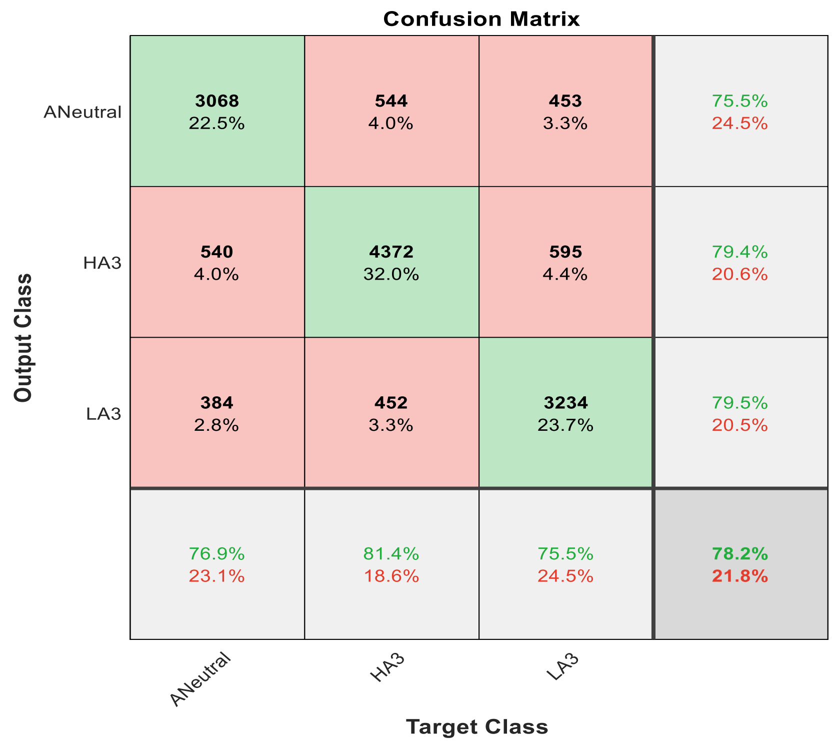



**Figure 5.7:** Confusion Matrix for 3 Classes of Arousal Dimension

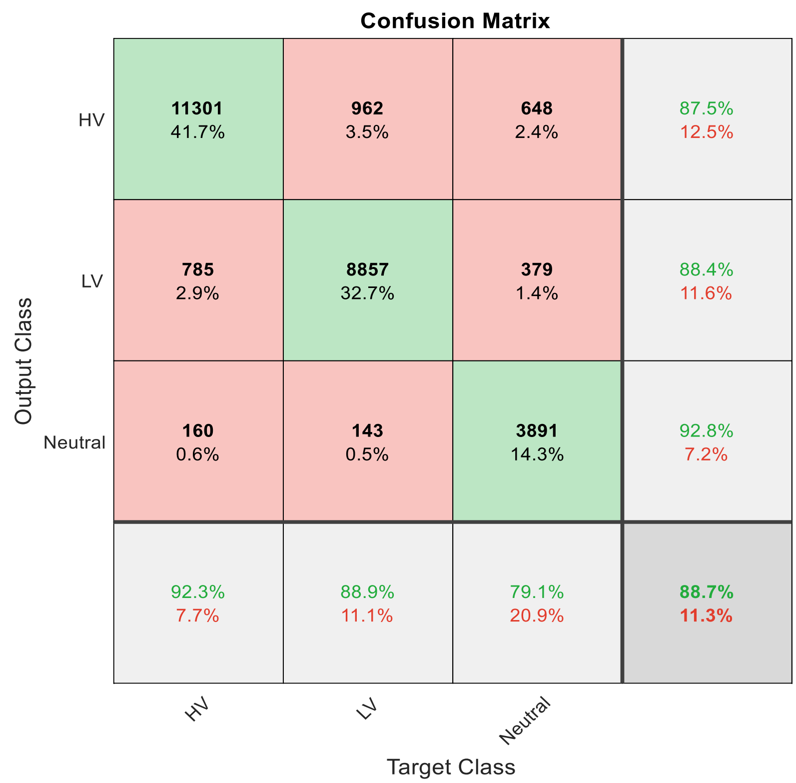



**Figure 5.8:** Confusion Matrix for 3 Classes of Valence Dimension with 50% Overlap + Interpolation

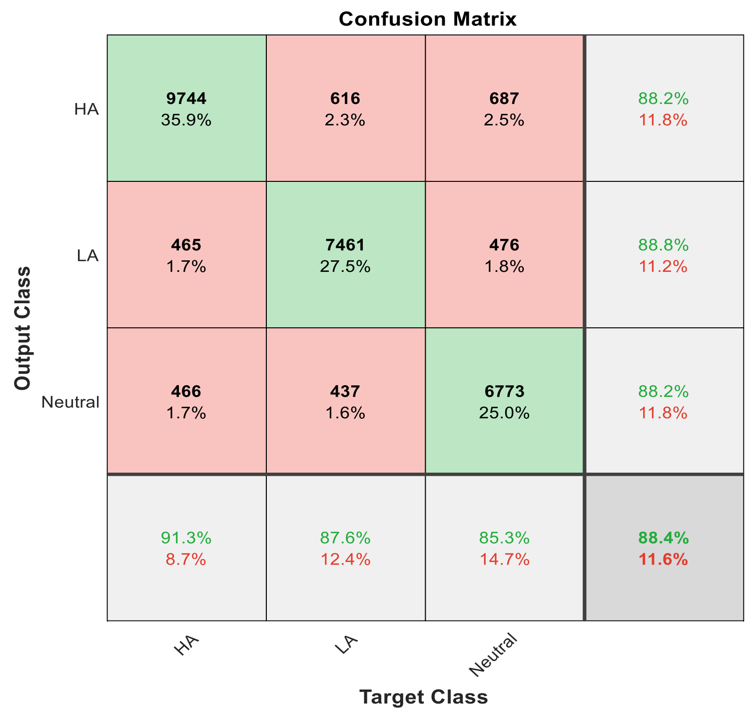



**Figure 5.9:** Confusion Matrix for 3 Classes of Arousal Dimension with 50% Overlap + Interpolation

### 5.2.3 Experimental results for Four Classes

The four classes are named as Low Valence Low Arousal (LVLA), Low Valence High Arousal (LVHA), High Valence Low Arousal (HVLA) and High Valence High Arousal (HVHA) [52]. These four class labels are obtained by dividing the complete dataset into four classes. As the dimensions of valence, arousal and dominance are self-assessed by participants on a scale of 1 to 9, value of 5 is used as threshold to divide the valence and arousal dimensions into high and low levels. The data of 33 persons out of 40 persons is used as input to the CNN architecture as the data of 7 persons with ID numbers 9, 12, 21, 22, 23, 24 and 33 has been omitted as it contains partially invalid data in the preprocessed dataset. The total number of 2D images with dimensions 81x128 from AMIGOS dataset prepared without overlap as final input to the CNN architecture is 45,474 and 90,420 with 50% overlapping sliding window. Table 5.6 shows the distribution of examples/instances in each of the four classes. Table 5.7 shows the number of instances in four classes with 50% overlap. Figure 5.

| **Classes** | **LVLA** | **LVHA** | **HVLA** | **HVHA** |
|---|---|---|---|---|
| **Arousal Values** | $1<x\leq 5$ | $5<x<9$ | $1<x\leq 5$ | $5<x<9$ |
| **Valence Values** | $1<x\leq 5$ | $1<x\leq 5$ | $5<x<9$ | $5<x<9$ |
| **No. of instances** | 9,939 | 12,656 | 12,582 | 10,297 |

**Table 5.6:** No. of Instances/Examples in Four Classes without Overlap

| **Classes** | **LVLA** | **LVHA** | **HVLA** | **HVHA** |
|---|---|---|---|---|
| **Arousal Values** | $1<x\leq 5$ | $5<x<9$ | $1<x\leq 5$ | $5<x<9$ |
| **Valence Values** | $1<x\leq 5$ | $1<x\leq 5$ | $5<x<9$ | $5<x<9$ |
| **No. of instances** | 19,750 | 25,171 | 25,027 | 20,472 |

**Table 5.7:** No. of Instances/Examples in Four Classes with 50 % Overlap

The distribution of actual labels of 16 trials of 33 person in 2D space for four classes (LVLA, LVHA, HVLA, and HVHA) is shown in figure 5.10. The black dots represent the mean value of all the examples in one class. Figure 5.11 represents the confusion matrix for four classes of emotions without overlap when 2D CNN is used. Figure 5.12 shows confusion matrix for four classes of emotions without overlap when Alexnet is used. Figure 5.13 shows the confusion matrix with 50% overlap when 2D CNN is used. Figure 5.14 shows the confusion matrix with 50% overlap combined with linear interpolation when 2D CNN is used. Figure 5.15 shows the confusion matrix with 50% overlap combined with linear interpolation when Alexnet is used. Figure 5.16 shows the confusion matrix with 50% overlap combined with nearest neighbor interpolation when 2D CNN is used.

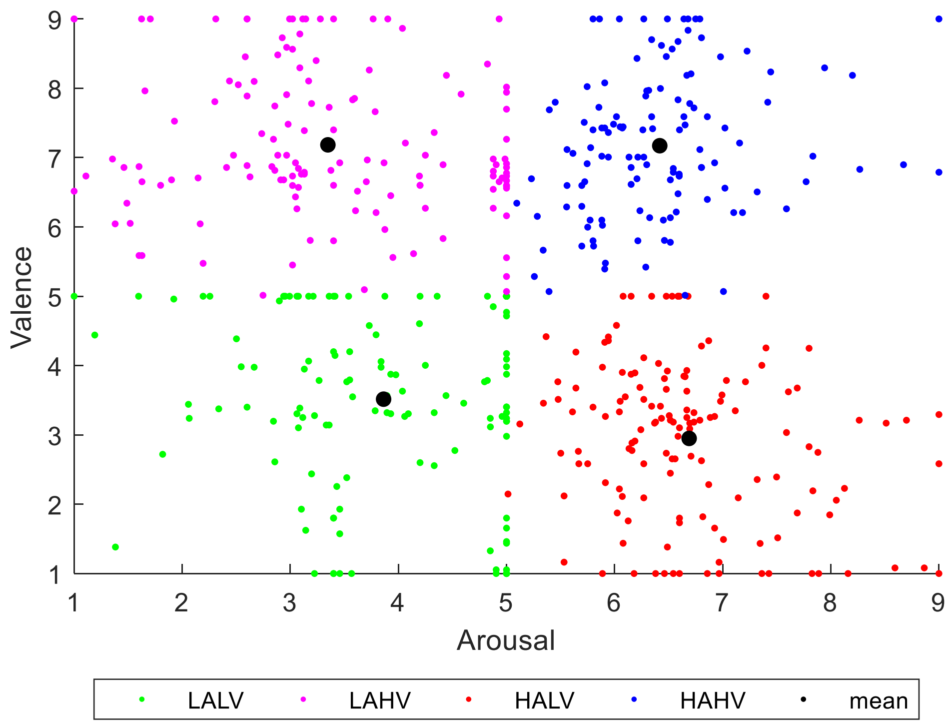



**Figure 5.10:** Distribution of Actual Labels in Four Classes of Dataset

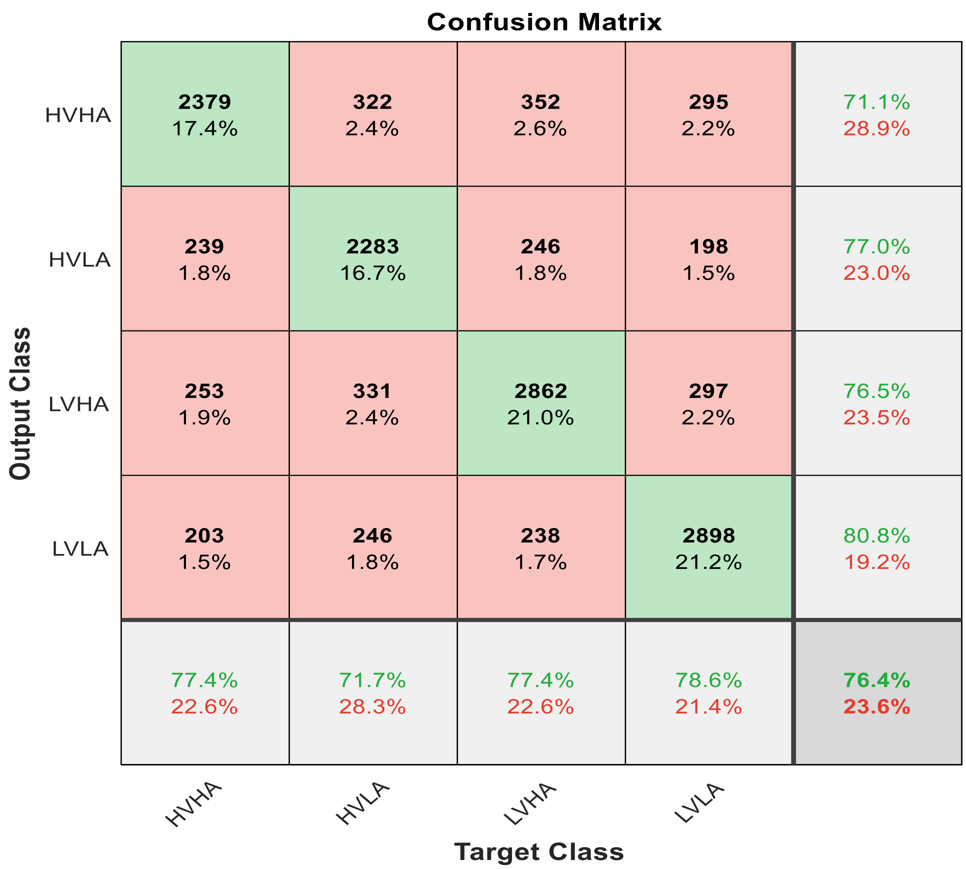



**Figure 5.11:** Confusion Matrix for Four Classes of Emotions without Overlap

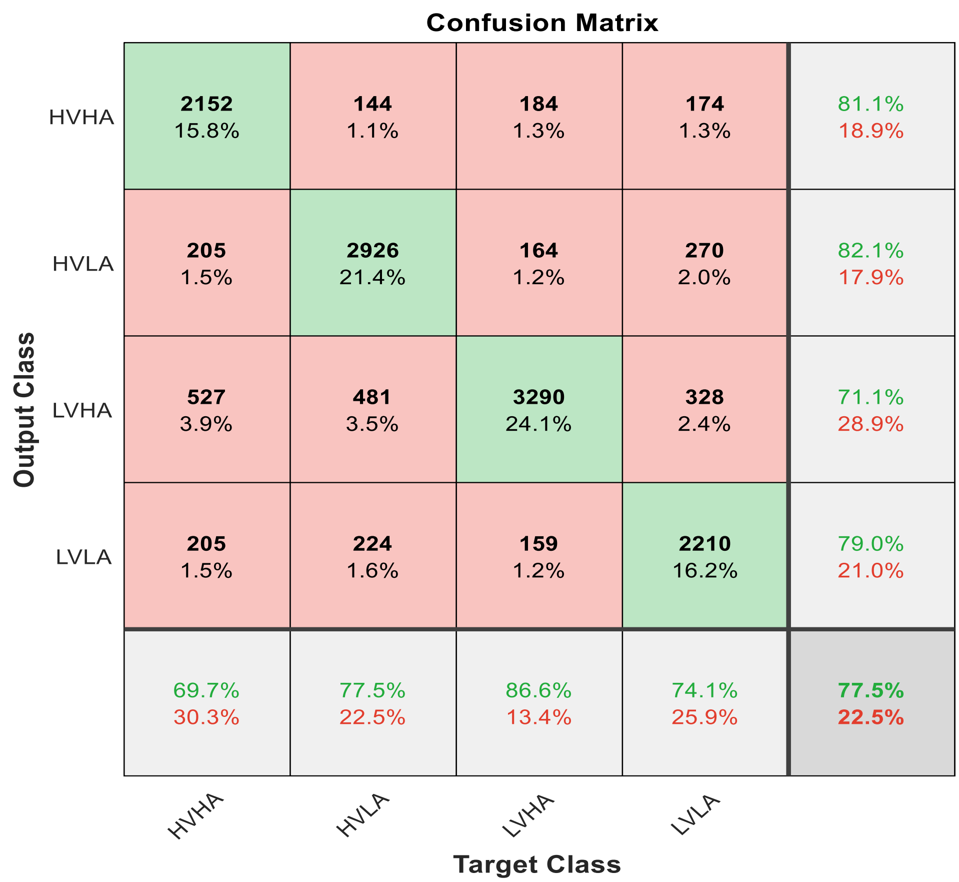



**Figure 5.12:** Confusion Matrix for Four Classes of Emotions with Alexnet without Overlap

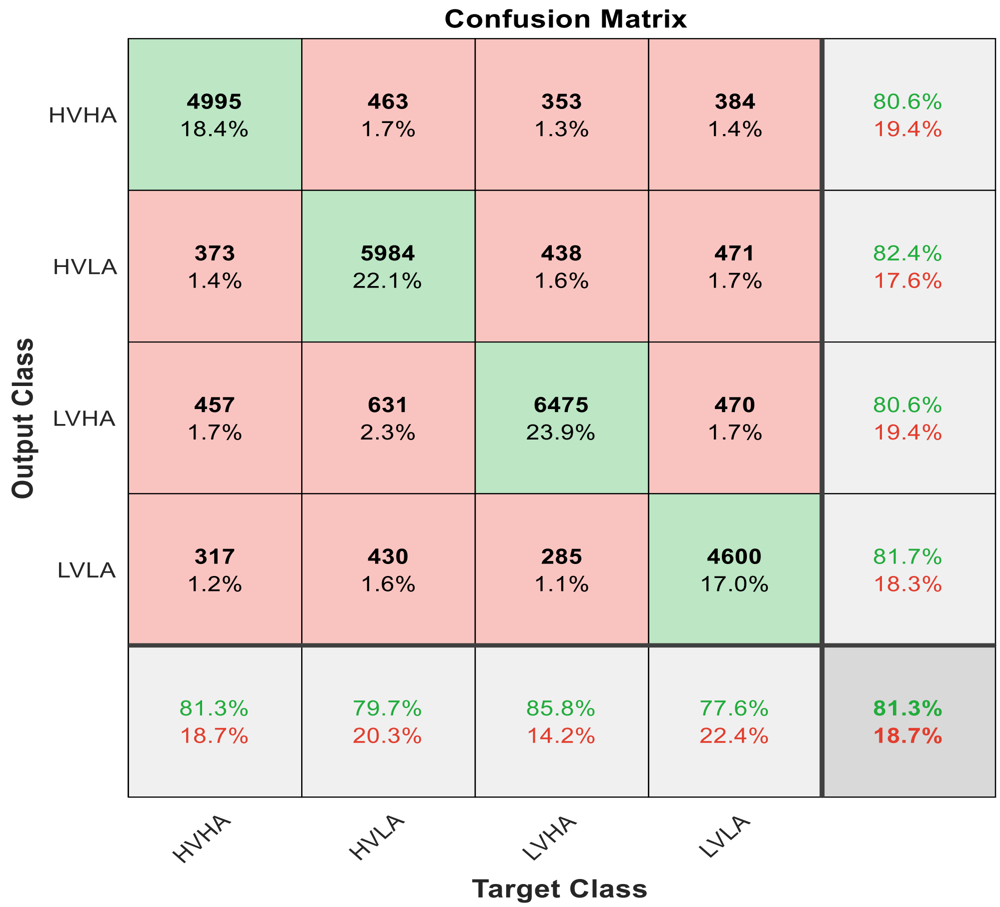



**Figure 5.13:** Confusion Matrix for Four Classes of Emotions with 50% Overlap

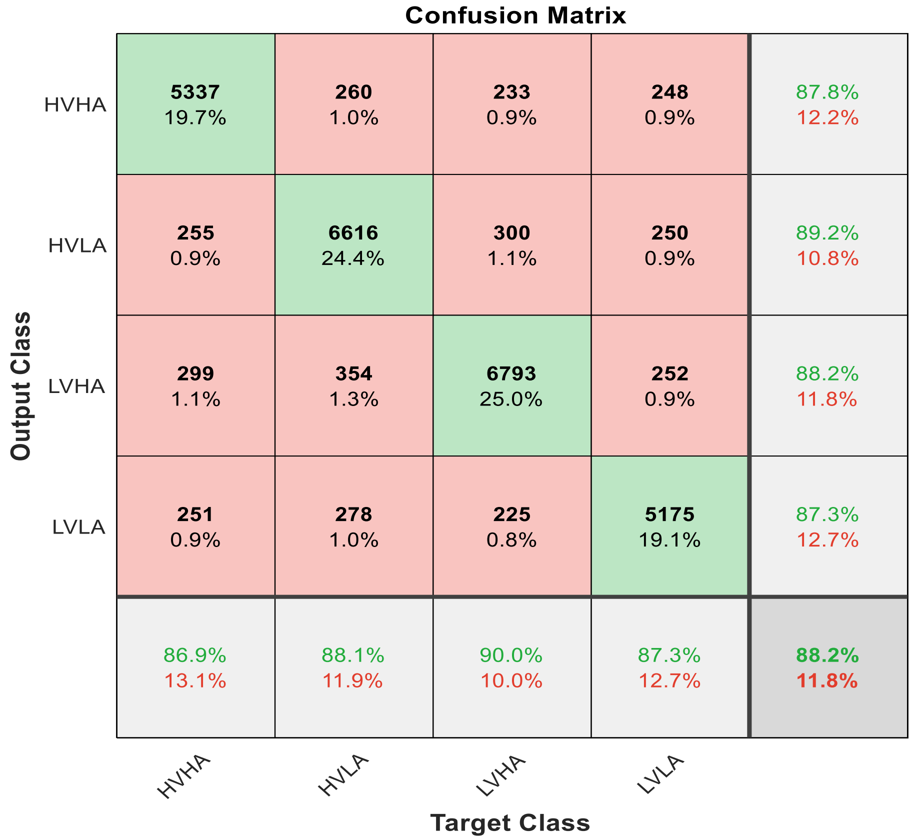



**Figure 5.14:** Confusion Matrix for Four Classes of Emotions with 50% Overlap + Linear Interpolation

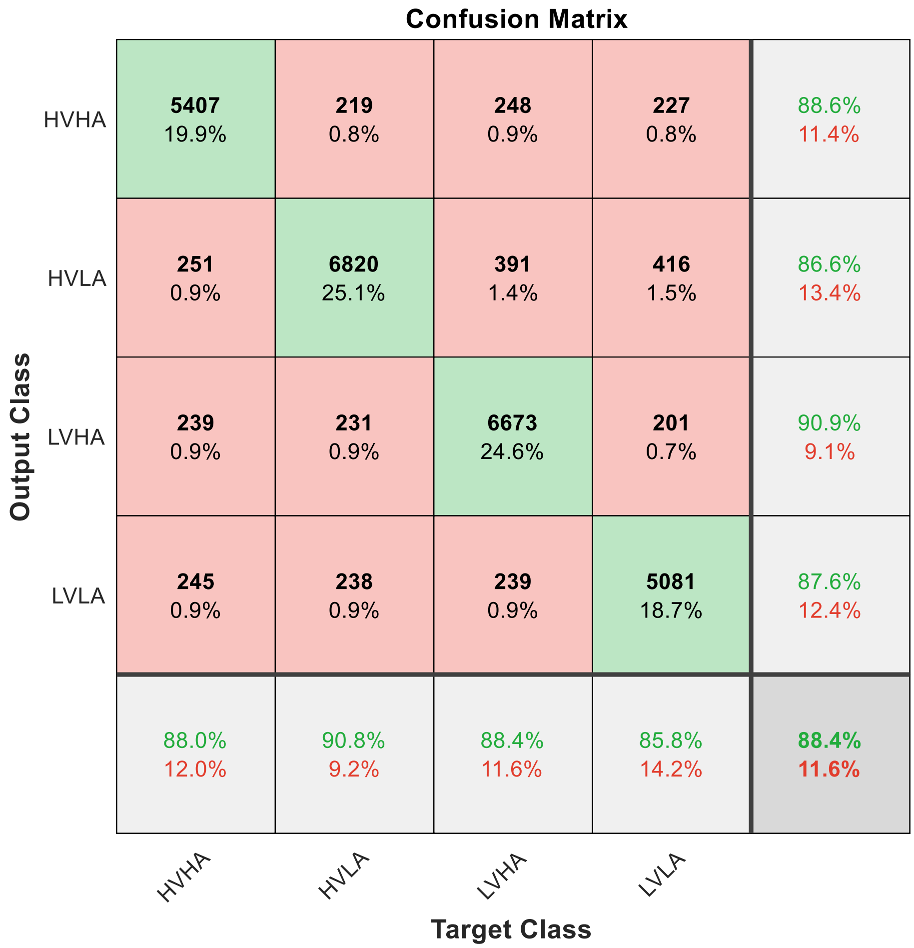



**Figure 5.15:** Confusion Matrix for Four Classes of Emotions with 50% Overlap + Linear Interpolation with Alexnet

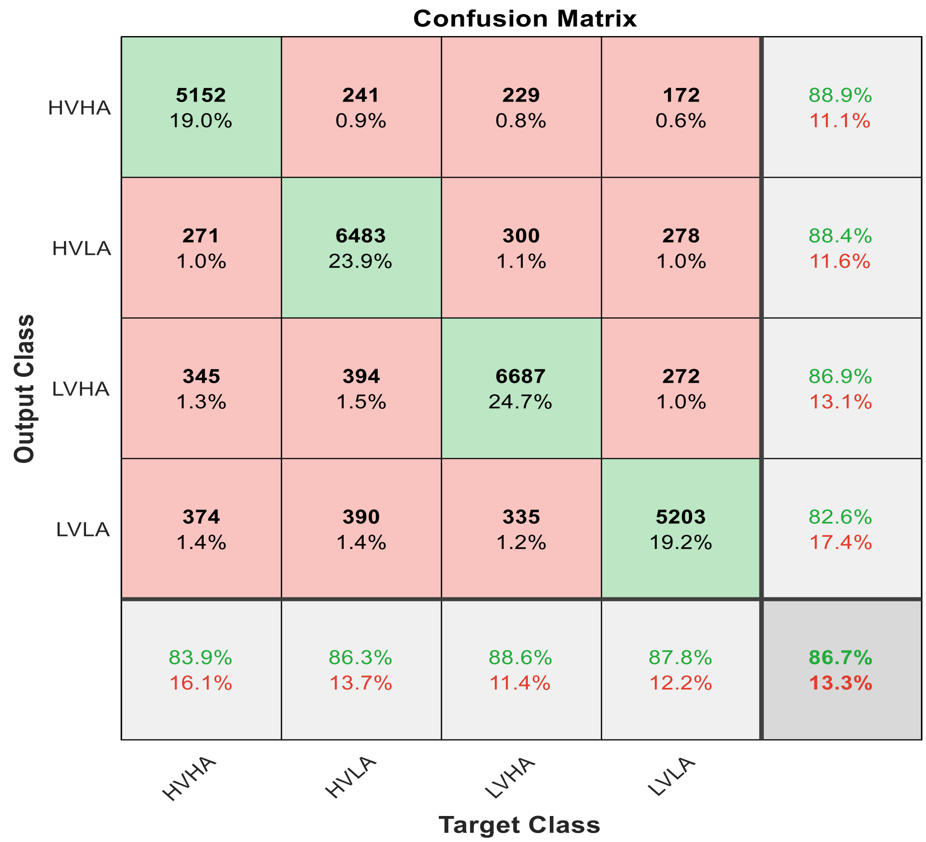



**Figure 5.16:** Confusion Matrix for Four Classes of Emotions with 50% Overlap + Nearest Neighbor Interpolation:

| Dimension | | Valence | Arousal |
|---|---|---|---|
| **2-Classes (Low, High)** | without overlap | 79.40% | 81.90% |
| | 50% overlap + linear interpolation | 90.0% | 90.40% |
| **3-Classes (Low, Neutral, High)** | without overlap | 78.50% | 78.20% |
| | 50% overlap + linear interpolation | 88.7% | 88.4% |
| **4-Classes (LVLA, LVHA, HVLA, HVHA)** | without overlap | 76.40% | |
| | 50% overlap | 81.30% | |
| | 50% overlap + linear interpolation | 88.20% | |
| | 50% overlap + nearest neighbor interpolation | 86.70% | |
| | Alexnet without overlap | 77.80% | |
| | Alexnet with 50% overlap + linear interpolation | 88.4% | |

**Table 5.8: Classification Accuracies of Different Classes**

## 5.3 Classification with Scalograms

A scalogram represents the distribution of spectral energy in a signal [39]. The frequency range with high concentration of energy is represented by hot color of the specific pixels in a scalogram. For the classification with scalograms two approaches are adopted. In the first approach the 1D EEG signals of all channels are segmented into segments of 1s length each. The average of data from all channels in one segment is computed and then the scalogram is computed for each segment. The mother wavelet used computing continuous wavelet transform (CWT) is ‘morse’ wavelet. The computed scalograms are then sent to the same 2D CNN for classification into four classes. Figure 5.17 shows some of the scalograms formed for different segments. The classification results are shown in figure 5.18.

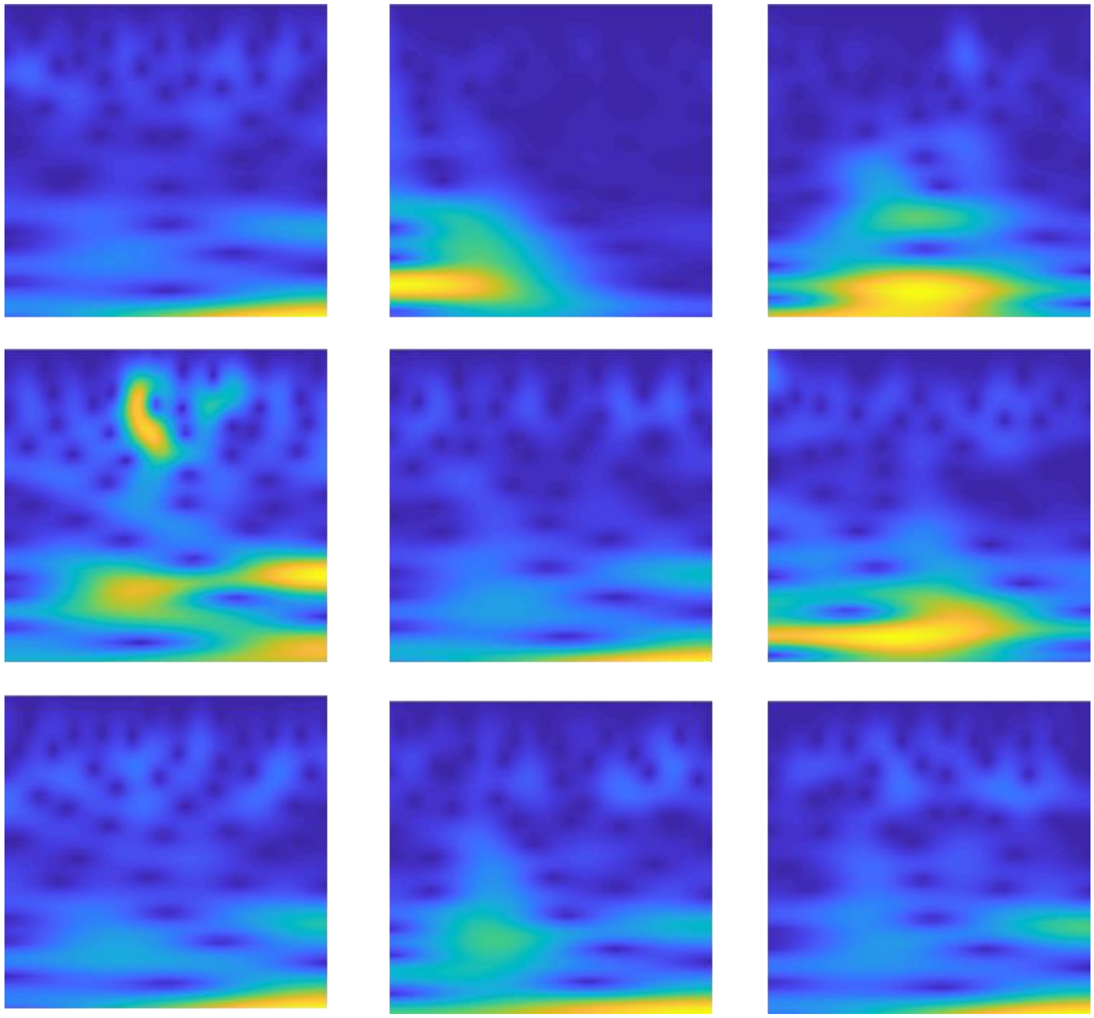


**Figure 5.17:** Scalograms of Different Segments of EEG signals:

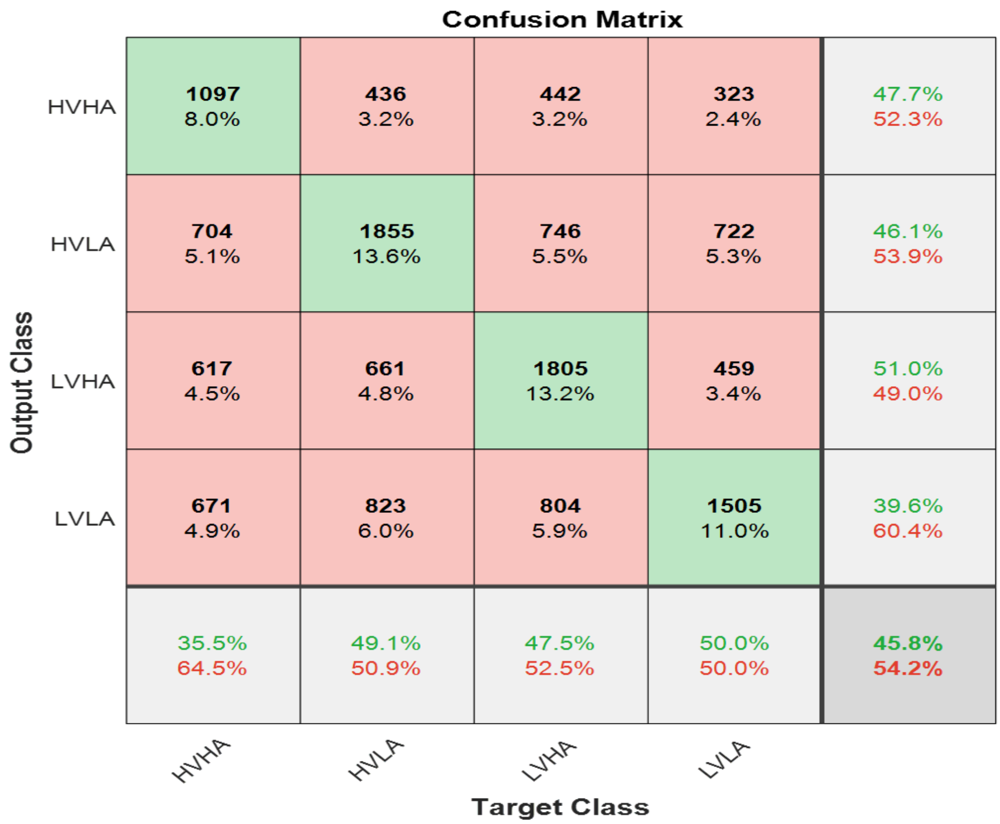



**Figure 5.18:** Classification Results with Scalograms with Average of Channels' Data Samples:

The second approach used involves the computation of complete 1D signals of individual channels. Afterwards the elements of scalogram of one channel are added up during one time window along the time axis to get 1D vector. Similar 1D vectors from scalograms of all channels are stacked together to obtain a 2D matrix. The time window is moved ahead and similar procedure is repeated to get 2D matrix of next time window. Same process is adopted to get 2D matrices corresponding to all time windows of one trial containing elements of scalograms from all channels' signals. The mother wavelet chosen for computing CWT is 'Db-4'. The 2D frame representations are classified using Alexnet. Figure 5.19 shows some of the 2D matrices containing elements of scalogrms from one trial. Figure 5.20 shows the classification results. Both the approaches have not shown good result shows that scalograms are not suitable for classification of multi-channel EEG signals.

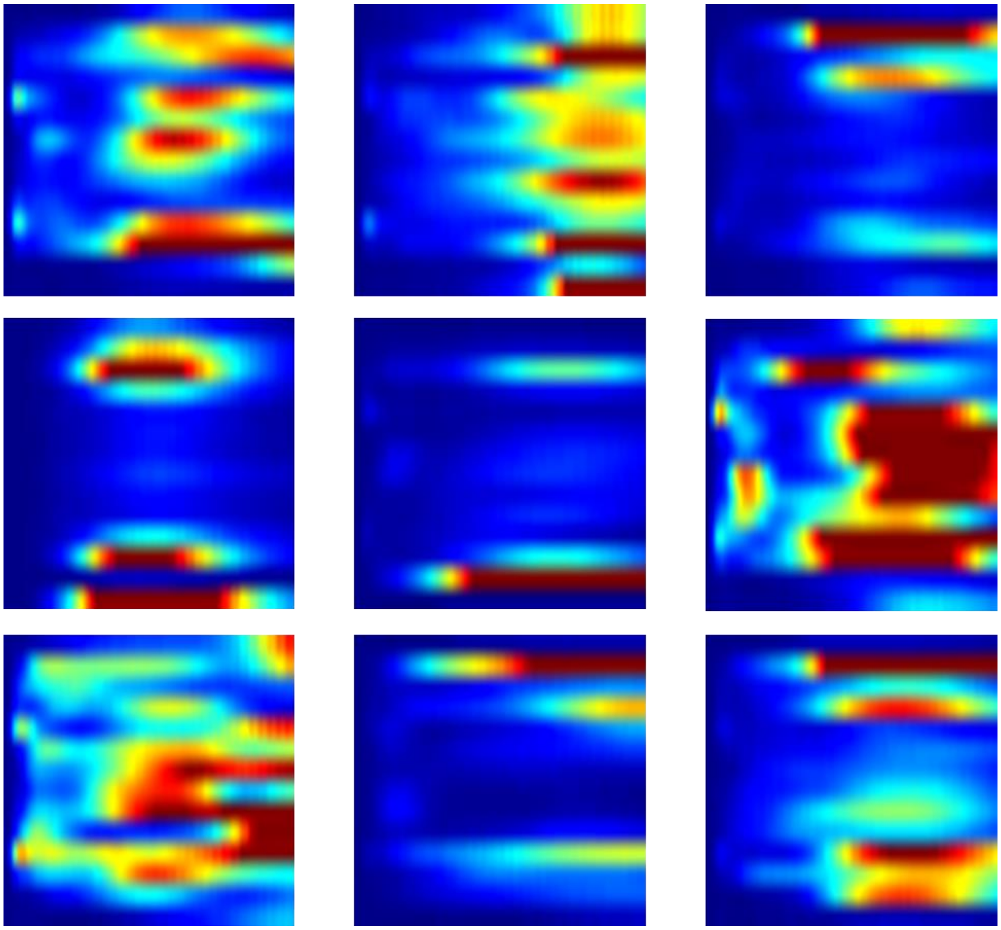


**Figure 5.19:** 2D Representation of Elements of Scalograms:

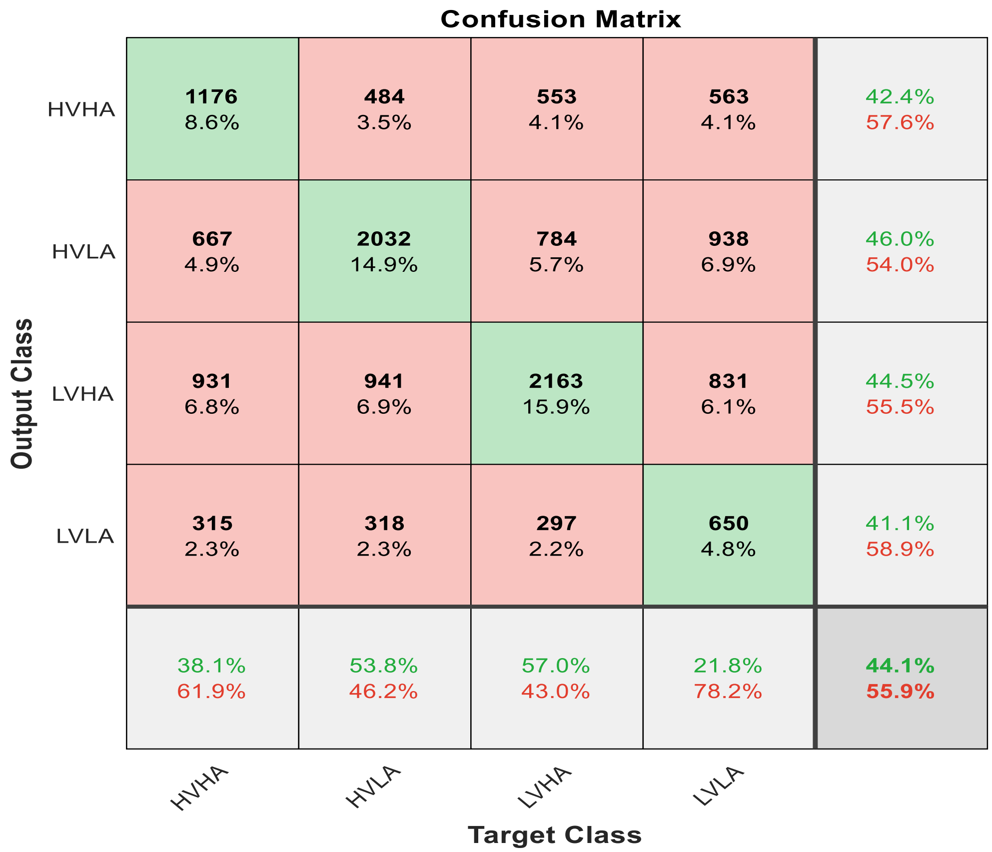



**Figure 5.20:** Classification results with 2D Representation of Elements of Scalograms:

## 5.4 Discussion

The experimental results are obtained by first creating the 2D topological images for each segment of 1D signal of 1sec length each. These images when sent to CNN give accuracy of 76.4% for four classes when there is no overlap sliding windows. There is a slight improvement in accuracy (77.5%) when the same images are classified using Alexnet. The number of images are then doubled with 50% overlap of sliding window and the accuracy gets improved to 81.3% and this improvement in accuracy is due to the fact that the overlapping sliding window approach helps in removing the abrupt changes. The 2D spatial matrices which are created using the location map of electrodes is very sparse as only 14 out of 81 indices of matrices are filled with data values. This sparsity in matrices is reduced by making them more dense using linear interpolation. The dense matrices are then converted to 2D topological images and these images improve the accuracy further to 88.2%. When the same images with 50% overlap and linear interpolation are sent to Alexnet then the accuracy is 88.4%. The matrices are also interpolated using nearest neighbor interpolation and the classification accuracy in this case is 86.7%. The overlapping sliding window approach is applied to 2 and 3 class classification tasks as well. The same procedure of making the images from more dense matrices is also adopted for 2 and 3 classes of Valence and arousal and the accuracies for both dimensions are improved from 79.4% to 90.4% for Valence (2classes), from 81.9% to 90.4% for Arousal (2classes), from 78.5% to 88.7% for Valence (3classes) and from 78.2% to 88.4 % for Arousal (3classes). The comparison with related studies that have used the same dataset AMIGOS is presented in table.

| **Study** | **Dataset** | **Features** | **Classifier** | **Accuracy** | | |
|---|---|---|---|---|---|---|
| | | | | **Arousal (2 classes)** | **Valence (2 classes)** | **4 classes** |
| [70] | AMIGOS | Statistical Features | Variational Auto-encoder (VAE) | 68.8% | 67% | - |
| [51] | AMIGOS | Power spectral density, spectral power asymmetry | SVM | 59.2% | 57.6% | - |
| [50] | AMIGOS | Spatio-temporal features | 3D-CNN | - | - | **95.86 %** |
| [72] | AMIGOS | Spectrogram | LSTM-RNN | 83.3% | 79.4% | - |
| [71] | AMIGOS | power spectral density | CNN-VGG-16 | 79.13% | 83.02% | 55.71% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Exp setup 1 | AMIGOS | Scalograms | 2D CNN | - | - | 45.8% |
| Exp Setup 2 (without overlap) | AMIGOS | Spatio-temporal features | 2D CNN | 79.4% | 81.9% | 76.4% |
| Exp setup 2 (50% overlap) | AMIGOS | Spatio-temporal features | 2D CNN | - | - | 81.3% |
| Exp setup 2 (50% overlap + interpolation) | AMIGOS | Spatio-temporal features | 2D CNN | **90.4%** | **89.1%** | 88.2% |

**Table 5.9:** Comparison with Related Studies

# CHAPTER 6: CONCLUSION & FUTURE WORK

## 6.1 Conclusion

In the proposed methodology the EEG signals are sent to 2D CNN in the form of 2D topological images. The classification results show that although CNNs are known to be best suited for extracting features from images but they can also provide good results for classification of data other than images as in case of EEG signals. The baseline signals which correspond to no neural activity, are subtracted from the actual signals and this pre-processing step greatly improved the accuracy. EEG signals are sent directly to CNN architecture without extracting hand crafted features and still good results are achieved. The studies [40, 41, 42, 50, 66] that used raw EEG signals as input to CNN achieved higher accuracies as compared to studies [15, 22, 23, 39, 49] that used hand crafted features in time domain, frequency domain and both time-frequency domain as input to CNN. The overlapping sliding window approach results in improving the overall accuracy as it overcomes the abrupt variations. Dense spatio-temporal feature matrices provide better results as compared to sparse spatio-temporal matrices. The shallow 2D CNN can provide comparable results with deep architectures. Scalograms are not suitable for classification of multi-channel EEG signals

## 6.2 Contribution

- Emotion recognition system for classification of emotions from EEG signals in two, three and four classes using a 2D convolutional neural network.
- Review & comparison of recent developments in emotion recognition systems from electroencephalogram (EEG) signals and to see usefulness of deep learning for this application.

## 6.3 Future Work

The proposed study is based on 2D images being sent to 2D CNN. The scope of this study can be extended to 3D CNN as it can extract spatial and temporal features effectively because in 3D CNN both convolution and pooling operations are performed spatially as well as temporally.

The combination of CNN and LSTM networks can also be explored to better deal with spatio-temporal features. Also the scope of this study can be extended to data from stroke patients and other real-time applications including online BCI applications.

# REFERENCES:

[1] Ernst Niedermeyer, Fernando Lopes da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields - Page 140, Lippincott Williams & Wilkins, 2004 ISBN 0-7817-5126-8, ISBN 978-0-7817-5126-

[2] https://www.trans-cranial.com/docs/10_20_pos_man_v1_0_pdf.pdf

[3] The Scientific World Journal, vol. 2014, Article ID 627892, 10 pages, 2014. https://doi.org/10.1155/2014/62789232

[4] https://mentalhealthdaily.com/2014/04/15/5-types-of-brain-waves-frequencies-gamma-beta-alpha-theta-delta/

[5] Deak, Anita. (2011). Brain and Emotion: Cognitive Neuroscience of Emotions. Rev Psychol. 18. 71-80.

[6] Raschle N, Tshomba E, Menks W, Fehlbaum L and Stadler C (2016) Emotions and the Brain – Or How to Master “The Force”. Front. Young Minds. 4:16. doi: 10.3389/frym.2016.00016

[7] A Guide to the Anatomy of the Brain (verywellmind.com).

[8] Sigurdsson, T., Doyère, V., Cain, C. K., & LeDoux, J. E. (2007). Long-term potentiation in the amygdala: A cellular mechanism of fear learning and memory.

[9] Stangor, C., & Walinga, J. (2019). 5.2 Our Brains Control Our Thoughts, Feelings, and Behaviour. Introduction to Psychology.]

[10] Other content on MentalHealthDaily.com

[11] Hockenbury, D. H., & Hockenbury, S. E. (2010). Discovering psychology. Macmillan.

[12] Ekman P. Basic Emotions. *Handbook of Cognition and Emotion*. 2005:45-60. doi:10.1002/0470013494.ch3

[13] Plutchik, R. (1984). In search of the basic emotions. PsycCRITIQUES, 29(6), 511-513.

[14] Yang, Yilong & Wu, Qingfeng & Fu, Yazhen & Chen, Xiaowei. (2018). Continuous Convolutional Neural Network with 3D Input for EEG-Based Emotion Recognition

[15] T. Song, W. Zheng, P. Song and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," in IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 532-541, 1 July-Sept. 2020, doi: 10.1109/TAFFC.2018.2817622.

[16] J. Liu, H. Meng, A. Nandi and M. Li, "Emotion detection from EEG recordings," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, 2016, pp. 1722-1727, doi: 10.1109/FSKD.2016.7603437.

[17] Malmivuo, Jaakko&Plonsey, Robert. (1995). Bioelectromagnetism. 13. Electroencephalography.

[18] Hjorth B. EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. 1970;29:306–310. doi: 10.1016/0013-4694(70)90143-4

[19] Kroupi E., Yazdani A., Ebrahimi T. EEG correlates of different emotional states elicited during watching music videos; Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction; Memphis, TN, USA. 9–12 October 2011; Berlin/Heidelberg, Germany: Springer; pp. 457–466. [Google Scholar]

[20] Petrantonakis P.C., Hadjileontiadis L.J. Emotion recognition from EEG using higher order crossings. IEEE Trans. Inf. Technol. Biomed. 2010;14:186–197. doi: 10.1109/TITB.2009.2034649. [PubMed] [CrossRef] [Google Scholar]

[21] Petrantonakis P.C., Hadjileontiadis L.J. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. IEEE Trans. Affect. Comput. 2010;1:81–97. doi: 10.1109/T-AFFC.2010.7.

[22] Tripathi Samarth, Acharya Shrinivas, Sharma Dev Ranti, Mittal Sudhanshu and Bhattacharya Samit "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset."(2017)

[23] Alnafjan, Abeer & Hosny, Manar & Al-Wabil, Areej & Al-Ohali, Yousef. (2017). Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network. International Journal of Advanced Computer Science and Applications. 8. 10.14569/IJACSA.2017.080955.

[24] Jirayucharoensak, Suwicha & Pan-ngum, Setha & Israsena, Pasin. (2014). EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation. TheScientificWorldJournal. 2014. 627892. 10.1155/2014/627892.

[25] Pan, Chao; Shi, Cheng; Mu, Honglang; Li, Jie; Gao, Xinbo. 2020. "EEG-Based Emotion Recognition Using Logistic Regression with Gaussian Kernel and Laplacian Prior and Investigation of Critical Frequency Bands" *Appl. Sci.* 10, no. 5: 1619.

[26] Valenzi, Stefano & Islam, Tanvir & Jurica, Peter & Cichocki, Andrzej. (2014). Individual Classification of Emotions Using EEG. Journal of Biomedical Science and Engineering. 07. 604-620. 10.4236/jbise.2014.78061.

[27] R. Jenke, A. Peer and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," in IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 327-339, 1 July-Sept. 2014, doi: 10.1109/TAFFC.2014.2339834.

[28] Cho, Jungchan; Hwang, Hyoseok. 2020. "Spatio-Temporal Representation of an Electoencephalogram for Emotion Recognition Using a Three-Dimensional Convolutional Neural Network" *Sensors* 20, no. 12: 3491. https://doi.org/10.3390/s20123491

[29] Nasehi, S. & Pourghassem, Hossein. (2012). An optimal EEG-based emotion recognition algorithm using gabor features. WSEAS Transactions on Signal Processing. 8. 87-99.

[30] M. Murugappan, "Human emotion classification using wavelet transform and KNN," in Pattern analysis and intelligent robotics (ICPAIR), 2011 international conference on, 2011, vol. 1, pp. 148-153: IEEE.]

[31] Mohammadi, Z., Frounchi, J. & Amiri, M. Wavelet-based emotion recognition system using EEG signal. *Neural Comput & Applic* 28, 1985–1990 (2017). https://doi.org/10.1007/s00521-015-2149-8

[32] O. Bazgir, Z. Mohammadi and S. A. H. Habibi, "Emotion Recognition with Machine Learning Using EEG Signals," 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME), Qom, Iran, 2018, pp. 1-5, doi: 10.1109/ICBME.2018.8703559.

[33] Classification In Machine Learning | Classification Algorithms | Edureka

[34] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," Artificial Intelligence Review, vol. 26, no. 3, pp. 159–190, 2006.

[35] Gu, Xiaotong, Zehong Cao, A. Jolfaei, Peng Xu, D. Wu, Tzyy-Ping Jung and Chin-Teng Lin. "EEG-based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and their Applications." *IEEE/ACM transactions on computational biology and bioinformatics* PP (2021): n. pag.

[36] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng. 2019 Jun;16(3):031001. doi: 10.1088/1741-2552/ab0ab5. Epub 2019 Feb 26. PMID: 30808014

[37] Y. Gao, H. J. Lee and R. M. Mehmood, "Deep learninig of EEG signals for emotion recognition," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, 2015, pp. 1-5, doi: 10.1109/ICMEW.2015.7169796.

[38] H. Mei and X. Xu, "EEG-based emotion classification using convolutional neural network," 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Shenzhen, 2017, pp. 130-135, doi: 10.1109/SPAC.2017.8304263.

[39] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou and B. Hu, "Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network," 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, 2016, pp. 352-359, doi: 10.1109/BIBM.2016.7822545.

[40] Zhang, Dalin & Yao, Lina & Zhang, Xiang & Wang, Sen & Chen, Weitong & Boots, Robert. (2017). Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface.

[41] Elham S. Salama, Reda A.El-Khoribi, Mahmoud E.Shoman and Mohamed A.Wahby Shalaby, "EEG-Based Emotion Recognition using 3D Convolutional Neural Networks" International Journal of Advanced Computer Science and Applications(IJACSA), 9(8), 2018. http://dx.doi.org/10.14569/IJACSA.2018.090843]

[42] Yang, Yilong, Q. Wu, M. Qiu, Yingdong Wang and Xiaowei Chen. "Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network." *2018 International Joint Conference on Neural Networks (IJCNN)* (2018): 1-7.

[43] Chao H, Dong L, Liu Y, Lu B. Emotion Recognition from Multiband EEG Signals Using CapsNet. Sensors (Basel). 2019 May 13; 19(9):2212. doi: 10.3390/s19092212. PMID: 31086110; PMCID: PMC6540345.

[44] Zeng H, Wu Z, Zhang J, et al. EEG Emotion Classification Using an Improved SincNet-Based Deep Learning Model. *Brain Sci*. 2019;9(11):326. Published 2019 Nov 14. doi:10.3390/brainsci9110326.

[45] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang and Y. N. Zhang, "Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks," in IEEE Access, vol. 7, pp. 44317-44328, 2019, doi: 10.1109/ACCESS.2019.2908285.

[46] K. H. Cheah, H. Nisar, V. V. Yap and C. Lee, "Short-time-span EEG-based personalized emotion recognition with deep convolutional neural network," 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 2019, pp. 78-83, doi: 10.1109/ICSIPA45851.2019.8977786.

[47] M. A. Ozdemir, M. Degirmenci, O. Guren and A. Akan, "EEG based Emotional State Estimation using 2-D Deep Learning Technique," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895158.

[48] G. Cao, Y. Ma, X. Meng, Y. Gao and M. Meng, "Emotion Recognition Based On CNN," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8627-8630, doi: 10.23919/ChiCC.2019.8866540.

[49] H. Donmez and N. Ozkurt, "Emotion Classification from EEG Signals in Convolutional Neural Networks," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 2019, pp. 1-6, doi: 10.1109/ASYU48272.2019.8946364.

[50] Zhao, Yuxuan, X. Cao, Jinlong Lin, Dunshan Yu and Xixin Cao. "Multimodal Emotion Recognition Model using Physiological Signals." *ArXiv* abs/1911.12918 (2019): n. pag.

[51] Miranda, Juan & Khomami Abadi, Mojtaba & Sebe, Nicu & Patras, Ioannis. (2017). AMIGOS: A dataset for Mood, personality and affect research on Individuals and GrOupS. IEEE Transactions on Affective Computing. PP. 10.1109/TAFFC.2018.2884461.

[52] Dar, Muhammad N.; Akram, Muhammad U.; Khawaja, Sajid G.; Pujari, Amit N. 2020. "CNN and LSTM-Based Emotion Charting Using Physiological Signals" *Sensors* 20, no. 16: 4551. https://doi.org/10.3390/s20164551

[53] Yousef Rezaei Tabar and Ugur Halici, "A novel deep learning approach for classification of EEG motor imagery signals", *Journal of neural engineering*, vol. 14, no. 1, 2016

[54] Li, Youjun; Huang, Jiajin; Zhou, Haiyan; Zhong, Ning. 2017. "Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks" *Appl. Sci.* 7, no. 10: 1060. https://doi.org/10.3390/app7101060

[55] Harper, R.; Southern, J. End-To-End Prediction of Emotion From Heartbeat Data Collected by a Consumer Fitness Tracker. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; pp. 1–7.

[56]. Zhang, J.; Chen, M.; Zhao, S.; Hu, S.; Shi, Z.; Cao, Y. ReliefF-based EEG sensor selection methods for emotion recognition. Sensors **2016**, 16, 1558. [CrossRef]

[57]. Zhuang, N.; Zeng, Y.; Tong, L.; Zhang, C.; Zhang, H.; Yan, B. Emotion recognition from EEG signals using multidimensional information in EMD domain. BioMed Res. Int. **2017**, 2017. [CrossRef]

[58]. Cimtay, Y.; Ekmekcioglu, E. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. Sensors **2020**, 20, 2034.

[59] Arevalillo-Herráez, M.; Cobos, M.; Roger, S.; García-Pineda, M. Combining Inter-Subject Modeling with a Subject-Based Data Transformation to Improve Affect Recognition from EEG Signals. Sensors **2019**, 19, 2999.

[60] P. Ekman and W. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues. Oxford: Prentice-Hall, 1975.

[61] YiWang and Zhiyi Huang. Emotionet: A 3-d convolutional neural network for eeg-based emotion recognition. In 2018 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2018.

[62] Lecun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44

[63] Yamashita, Rikiya & Nishio, Mizuho & Do, Richard & Togashi, Kaori. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging. 9. 10.1007/s13244-018-0639-9.

[64]https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj1keHPlODvAhXTTsAKHUUNB9kQFjAGegQIBhAD&url=https%3A%2F%2Fmedium.com%2Fdata sciencebootcamp%2Funderstand-the-softmax-function-in-minutes f3a59641e86d&usg=AOvVaw2tcgBpo5Su1bSkpXAfoHgg

[65] https://medium.datadriveninvestor.com/why-are-convolutional-neural-networks-good-for-image-classification-146ec6e865e8

[66] Jason Brownlee A Gentle Introduction to Batch Normalization for Deep Neural Networks. 2019 in Deep Learning Performance

[67] https://www.inf.ufpr.br/todt/IAaplicada/CNN_Presentation.pdf

[68] Basha, S.S.; Dubey, S.R.; Pulabaigari, V.; Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. Neurocomputing 2020, 378, 112–119. [CrossRef]

[69] Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, Massachusetts, 2012

[70] H. Yang and C. Lee, "An Attribute-invariant Variational Learning for Emotion Recognition Using Physiology," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 1184-1188, doi: 10.1109/ICASSP.2019.8683290.

[71] Siddharth, S.; Jung, T.P.; Sejnowski, T.J. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. arXiv **2019**, arXiv:1905.07039.

[72] Li, C.; Bao, Z.; Li, L.; Zhao, Z. Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. Inf. Process. Manag. 2020, 57, 102185.

[73] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, “Emotion representation, analysis and synthesis in continuous space: A survey,” in IEEE FG, 2011, pp. 827–834.

[74] Wang, Dan, and Yi Shang. “Modeling Physiological Data with Deep Belief Networks.” International journal of information and education technology (IJIET) vol. 3,5 (2013): 505-511. doi:10.7763/IJIET.2013.V3.326

[75] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, Jianhua Zhang, Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, Computer Methods and Programs in Biomedicine, Volume 140, 2017, Pages 93-110, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2016.12.005.

[76] Atefeh Goshvarpour, Ataollah Abbasi, Ateke Goshvarpour, An accurate emotion recognition system using ECG and GSR signals and matching pursuit method, Biomedical Journal, Volume 40, Issue 6, 2017, Pages 355-368, ISSN 2319-4170, https://doi.org/10.1016/j.bj.2017.11.001.

[77] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," in IEEE Access, vol. 7, pp. 57-67, 2019, doi: 10.1109/ACCESS.2018.2883213.

[78] Yang YX, Gao ZK, Wang XM, Li YL, Han JW, Marwan N, Kurths J. A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG. Chaos. 2018 Aug;28(8):085724. doi: 10.1063/1.5023857. PMID: 30180618.

[79] P. Zhong, D. Wang and C. Miao, "EEG-Based Emotion Recognition Using Regularized Graph Neural Networks," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2020.2994159.

[80] Zeng H, Wu Z, Zhang J, et al. EEG Emotion Classification Using an Improved SincNet-Based Deep Learning Model. *Brain Sci*. 2019;9(11):326. Published 2019 Nov 14. doi:10.3390/brainsci9110326

[81] M. A. Ozdemir, M. Degirmenci, O. Guren and A. Akan, "EEG based Emotional State Estimation using 2-D Deep Learning Technique," 2019 Medical Technologies Congress (TIPTEKNO), 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895158.

[82] Barrett LF, Mesquita B, Ochsner KN, Gross JJ. The experience of emotion. Annu Rev Psychol. 2007;58:373-403. doi: 10.1146/2Fannurev.psych.58.110405.085709

[83] Pessoa L. Emotion and cognition and the amygdala: from "what is it?" to "what's to be done?".Neuropsychologia. 2010;48(12):3416-29. doi: 10.1016/j.neuropsychologia.2010.06.038

[84] Ressler KJ. Amygdala activity, fear, and anxiety: modulation by stress. Biol Psychiatry. 2010;67(12):1117-9. doi: 10.1016/j.biopsych.2010.04.027

[85] [P S, SREEJA & G S, Mahalakshmi. (2017). Emotion Models: A Review. International Journal of Control Theory and Applications. 10. 651-657]

[86] P.J. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications," Technology in mental healthcare delivery systems, pp.119-137, 1980.

[87] R. Cowie, E. Douglas-Cowie, S. Savvidou*, E. McMahon, M.Sawey, and M. Schröder, " 'FEELTRACE': An instrumentfor recording perceived emotion in real time," In ISCA tutorial and research workshop (ITRW) on speech and emotion,pp. 19-24, 2000.

[88] S. D'mello, A. Graesser, " Mind and body: Dialogue and posture for affect detection in learning environments," Frontiersin Artificial Intelligence and Applications, vol.158, 161-168, 2007.

[89] https://en.wikipedia.org/wiki/Emotion_recognition

[90] [http://neurosky.com/2015/04/reading-your-brainwaves-understanding-the-basics-of-eeg/]

[91] Böcker, Koen &Avermaete, Jurgen& Berg-Lenssen, M. (1994). The international 10–20 system revisited: Cartesian and spherical co-ordinates. Brain topography. 6. 231-5. 10.1007/BF01187714.

[92] http://neurosky.com/2015/05/greek-alphanet-soup-making-sense-of-eeg-bands/