İZMİR
KÂTİP ÇELEBİ
ÜNİVERSİTESİ
2010
GRADUATE SCHOOL OF NATURAL
AND APPLIED SCIENCES

Multi-Channel EEG-Based Emotion Classification Using Deep Neural Networks

Submitted to the Graduate School of Natural and Applied Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

in Biomedical Technologies

by

TUĞBA ERGİN

ORCID 0000-0002-1040-4848

October, 2021

This is to certify that we have read the thesis **Multi-Channel EEG-Based Emotion Classification Using Deep Neural Networks** submitted by **Tuğba Ergin**, and it has been judged to be successful, in scope and in quality, at the defense exam and accepted by our jury as a MASTER'S THESIS.

APPROVED BY:

Advisor: **Assist. Prof. Dr. Onan Güren** ……………..
İzmir Kâtip Çelebi University

Committee Members:

Assoc. Prof. Dr. Aytuğ Onan ……………..
İzmir Kâtip Çelebi University

Assist. Prof. Dr. Şenay Mihçin ……………..
İzmir Institute of Technology

Date of Defense: October 27, 2021

Declaration of Authorship

I, **Tuğba Ergin**, declare that this thesis titled **Multi-Channel EEG-Based Emotion Classification Using Deep Neural Networks** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for the Master's degree at this university.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. This thesis is entirely my own work, with the exception of such quotations.

- I have acknowledged all major sources of assistance.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signature:

_____

Date: 27.10.2021

_____

Multi-Channel EEG-Based Emotion Classification Using Deep Neural Networks

Abstract

Emotions are the impressions that a particular object, event or moment creates in the inner world of a person. Emotions are an important part of our lives. The aim of this study is to classify emotions from multi-channel EEG signals using visual and auditory stimuli based on multivariate synchrosqueezing transform. In this study, EEG signals were obtained from 13 women and 12 men. The graphical user interface (GUI) is designed to display visual and auditory stimuli at specific time intervals. The 60 visual stimuli consisting of 48 images selected from the International Affective Image System (IAPS) and 12 images selected with the same characteristics were used. In addition, 60 auditory stimuli consisting of 48 voices selected from the International Affective Digitized Sounds System (IADS) and 12 voices selected with the same characteristics were used. Preprocessing is done after the EEG signals are obtained and long signals are segmented to obtain the time intervals in which the stimuli are shown. With the multivariate synchrosqueezing transform method, 32 channels were processed simultaneously. Then the 2D images obtained were given as input to the deep learning architecture. In this study, AlexNet, VGG-19, GoogLeNet, ResNet-50 and Inception-v3 architectures were used, which are among the most known convolutional neural network architectures. The best AlexNet came out and therefore other classification scenarios were executed over AlexNet. The k-fold cross validation was adopted to evaluate the robustness of the models proposed in this study and the k

value was chosen as 3. Performance measurements were made for Fold-1, Fold-2, Fold-3, and average fold in valence, arousal and dominance. By training the AlexNet architecture in visual arousal, an average accuracy value of 71.60% was reached, while an average accuracy value of 70.58% was achieved in auditory arousal. The average accuracy value for visual valence was 67.93%, while the average accuracy value for auditory valence was 70.58%. While the average accuracy value for visual dominance was 65.40%, the average accuracy value was 72.60% for auditory dominance. The results demonstrated that the proposed method achieved promising performance to classify the emotions. The proposed method reduced the computational cost, and significantly improved the classification performance.

Derin Sinir Ağları Kullanarak Çok Kanallı EEG Tabanlı Duygu Sınıflandırması

Öz

Duygular, belirli bir nesnenin, olayın veya anın bir kişinin iç dünyasında yarattığı izlenimlerdir. Duygular hayatımızın önemli bir parçasıdır. Bu çalışmanın amacı, çok değişkenli senkrosıkıştırma dönüşümüne dayalı olarak görsel ve işitsel uyaranları kullanarak çok kanallı EEG sinyallerinden duyguları sınıflandırmaktır. Bu çalışmada 13 kadın ve 12 erkekten EEG sinyalleri alınmıştır. Grafik kullanıcı arayüzü (GUI), belirli zaman aralıklarında görsel ve işitsel uyaranları görüntülemek için tasarlanmıştır. Uluslararası Afektif Görüntü Sisteminden (IAPS) seçilen 48 resim ve aynı özelliklerde seçilen 12 resimden oluşan 60 görsel uyaran kullanılmıştır. Ayrıca Uluslararası Duyuşsal Sayısallaştırılmış Sesler Sisteminden (IADS) seçilen 48 ses ve aynı özelliklerde seçilen 12 sesten oluşan 60 işitsel uyaran uyaran kullanılmıştır. EEG sinyalleri elde edildikten sonra ön işleme yapılır ve uyaranların gösterildiği zaman aralıklarını elde etmek için uzun sinyaller segmentlere ayrılır. Çok değişkenli senkrosıkıştırma dönüşümü yöntemi ile 32 kanal eş zamanlı olarak işlenmiştir. Daha sonra elde edilen 2 boyutlu görüntüler derin öğrenme mimarisine girdi olarak verilmiştir. Bu çalışmada en bilinen evrişimli sinir ağı mimarilerinden AlexNet, VGG-19, GoogLeNet, ResNet-50 ve Inception-v3 mimarileri kullanılmıştır. En iyi AlexNet çıkmıştır ve bu nedenle diğer sınıflandırma senaryoları AlexNet üzerinden yürütüldü. Bu çalışmada önerilen modellerin sağlamlığını değerlendirmek için k-katlamalı çapraz doğrulama uygulanmış ve k değeri 3 olarak seçilmiştir. Uyarılma, valence ve baskınlıkta Kat-1, Kat-2, Kat-3 ve ortalama kat için performans ölçümleri yapılmıştır. AlexNet mimarisini görsel uyarılmada eğiterek, %71,60 ortalama doğruluk değerine,

işitsel uyarılmada ise %70,58 ortalama doğruluk değerine ulaşılmıştır. Görsel değerlik için ortalama doğruluk değeri %67.93 iken işitsel değerlik için ortalama doğruluk değeri %70.58 idi. Görsel baskınlık için ortalama doğruluk değeri %65,40 iken, işitsel baskınlık için ortalama doğruluk değeri %72,60 olarak bulunmuştur. Sonuçlar, önerilen yöntemin duyguları sınıflandırmak için umut verici bir performans gösterdiğini kanıtlamıştır. Önerilen yöntem, hesaplama maliyetini azaltmış ve sınıflandırma performansını önemli ölçüde iyileştirmiştir.

Anahtar Kelimeler: EEG sinyali, duygu tanıma, derin sinir ağları, evrişimli sinir ağları, AlexNet, çok değişkenli senkrosıkıştırma dönüşümü

Acknowledgment

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACC	Accuracy
AUC	Area Under Curve
BCI	Brain-Computer Interface
BCI LLMR	Brain-computer Interface Lower-Limb Motor Recovery
CNN	Convolutional Neural Networks
CWD	Choi-Williams Distribution
DBN	Deep Belief Network
DEAP	A Database for Emotion Analysis using Physiological Signals
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EMD	Emperical Mode Decomposition
EMG	Electromyogram
EOG	Electrooculogram
FPR	False Positive Rate
FSST	Fourier based Synchrosqueezed Transform
GUI	Graphical User Interface
HHT	Hilbert Huang Transform
ID	Identification Number
IAPS	International Affective Picture System
IADS	International Affective Digitized Sounds
LDA	Linear Discriminant Analysis

MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MSST	Multivariate Synchrosqueezing Transform
NHH	Negative Valence - High Arousal - High Dominance
NHL	Negative Valence - High Arousal - Low Dominance
NLH	Negative Valence - Low Arousal - High Dominance
NLL	Negative Valence - Low Arousal - Low Dominance
PCC	Pearson Coefficient of Correlation
PHH	Positive Valence - High Arousal - High Dominance
PHL	Positive Valence - High Arousal - Low Dominance
PLH	Positive Valence - Low Arousal - High Dominance
PLL	Positive Valence - Low Arousal - Low Dominance
PRE	Precision
ReLu	Rectified Linear Unit
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SAM	Self-Assessment Manikin
SEED	SJTU Emotion EEG Dataset
SEN	Sensitivity
SGD	Stochastic Gradient Descent
SFT	Spatial-Frequency-Temporal
SPE	Specificy
SST	Synchrosqueezing Transformations
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
SWT	Synchrosqueezed Wavelet Transforms
TFRs	Time–Frequency Representations

TN	True Negatives
TP	True Positives
TPR	True Positive Rate
UIHBCI	User-Independent Hybrid Brain-computer Interface
WSST	Wavelet based Synchrosqueezed Transforms
WT	Wavelet Transforms
WVD	Wigner Ville Distribution

List of Symbols

α	Alpha
β	Beta
γ	Gamma
θ	Theta
μV	Microvolt
Δ	Delta
Ω	Instantaneous frequencies
A	Instantaneous amplitudes
Hz	Hertz
l	Frequency index
$s_+(t)$	Multivariate analytical signal
T_k^{multi}	Multivariate time-frequency coefficient

Chapter 1

Introduction

Emotions are the impressions that a particular object, event or individual creates in the inner world of a person. Emotions are an important part of our lives. It has a huge impact on people's personality and mental state, provides interpersonal language and emotional communication [1]. It is also an important motivator of human behavior, that is, a source of motivation. Analyzing emotions is an important interdisciplinary research topic in the fields of psychology, medicine, neuroscience, computer science, and artificial intelligence [2].

People can express their emotions with words [3], tones, facial expressions [4], and body language. However, the inner emotion can be different. To understand the inner emotion, the electroencephalogram (EEG) method of emotional state detection is used. The EEG signal is commonly used to detect different emotions. EEG is one of the methods used in investigation of functions of living human brain [5]. The EEG uses the International 10/20 system [6]. Neurological problems [7] can be detected with signals from the EEG device and data about psychiatric diseases can be obtained. For understanding the inner feeling, EEG emotion recognition method is also used. Different stimuli are used to stimulate individuals to reveal different emotions. It is possible to use music [8], text, video [9] and images [10], and thus emotion signals can be recorded. External stimuli received through the senses come to the brain as nerve signals. Emotion signals can be used to diagnose specific emotions using various stimuli, and data can be used for disease diagnosis.

Thanks to EEG recording, abnormal conditions in the brain, functional disorders, such as differences in brain energy are detected. It can make very serious determinations about the brain that works so intricately. It does not cause any side effects for the

patient; it does not contain radiation. It is also a great advantage that EEG can be applied easily and quickly. EEG devices are portable, they are easy to apply and have good temporal resolution [11]. Due to the complexity of EEG signals and the variability of human emotions from person to person, studies in this area are still being conducted for EEG-emotion recognition systems that can recognize emotions with high accuracy.

Nowadays the researches, that aim to detect of human emotion with using EEG based on the relationship between emotional states and brain signals, significantly increased but it isn't reached enough success due to obtained insufficient emotion states or inappropriate selection of different signal processing algorithms. The success of emotion recognition with EEG signal studies will contribute to Brain-Computer Interface (BCI) studies aiming to increase the human-machine interaction with the BCI design that can fully perceive human needs. But they have not yet reached the desired level in interpreting the emotions of people.

In this study, it is aimed to perform emotion recognition by using specific signal processing algorithms to EEG signals obtained from the 32-channel EEG device with visual stimuli shown to individuals and auditory stimuli respectively. The 60 visual and 60 auditory stimuli will be applied to 32 people participating in the research and EEG signals will be recorded during this time. In addition, effective features in emotion analysis will be investigated and the applied algorithms will be compared by applying Machine Learning and Deep Learning algorithms to multi-channel EEG signals and inferences will be made from the results obtained Deep Learning techniques.

1.1 Brain

The nervous system, due to its functions, is the body's most complex system. The nervous system consists of two main parts as the central and peripheral nervous system. The central nervous system consists of two main structures called the brain and spinal cord. The brain is located in the skull bone system, and the spinal cord is in the spine and extends from the neck to the coccyx. Thanks to the central nervous system, we move, feel, taste and see. An adult's brain weighs an average of 1300-1400

grams. Located in the skull cavity, the brain consists of 100 billion neurons and trillions of support cells called "glia". The brain is well protected within the skull bone system. The skull protects the entire brain from trauma and impact. The Figure 1.1 shows the whole brain, it consists of the cerebrum, cerebellum, and brainstem.
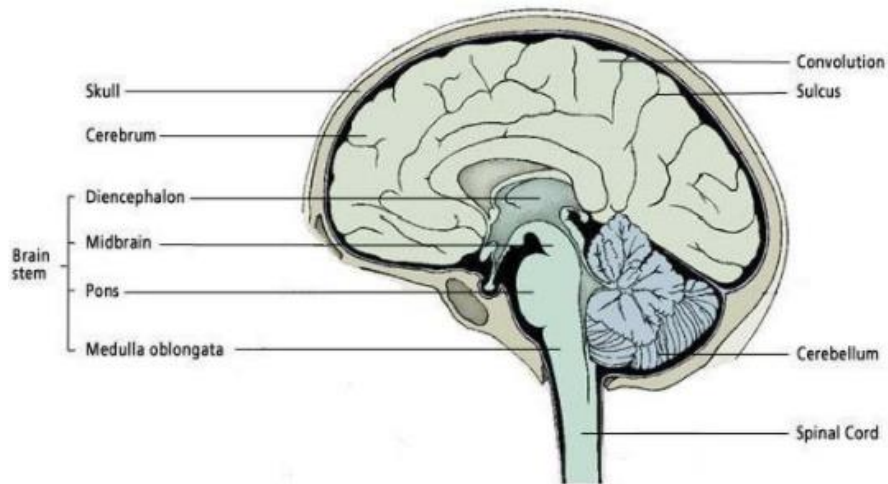
Figure 1.1: The brain structure [12]

The brain is a complex structure. Millions of times a second, signals and messages are sent to our brain from various parts of our body. The brain realizes these by evaluating and separating them through neurons, comparing them with the old information in the memory, and establishing new connections between neurons. The nervous system is the body's electrochemical communication network. All kinds of simple or complex behaviors take place through the activity of nerve cells called neurons located in different parts of the brain. Communication between neurons is essential for the brain to function.

The working system of the brain is based on nerve cells (neurons) and the communication between them. The neuron is the information-processing and information-transmitting component of the nervous system and mainly includes that, cell body or soma, dendrites, axon, terminal buttons. Hundreds or thousands of dendrites can be found in a nerve cell. Soma includes structures related to the vital processes of the nucleus and the cell. Dendrites have a branched-out structure and receive information from axon terminals of other cells. The axon has a thin cylindrical

structure that transmits information from the right to the terminal buttons. The terminal buttons are located at the end of the axon branch, it creates synapses with another neuron and sends information to that neuron. The Figure 1.2 shows the structure of biologic nerve cell. A: cell body, B: axon, C: dendrites, D: terminal buttons.
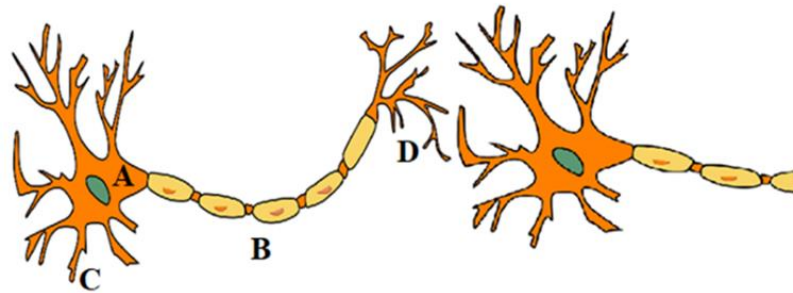
Figure 1.2: The structure of biologic nerve cell

1.1.1   Lobes of The Brain and Functions

The brain is an organ located in the skull that controls all functions in our body. The front part of the brain is called the "frontal", the middle part "parietal", the posterior "occipital" and the lateral "temporal" lobe. These lobes are located in both brain hemispheres. Each section has a specific function. The Figure 1.3 shows the lobes of the brain.
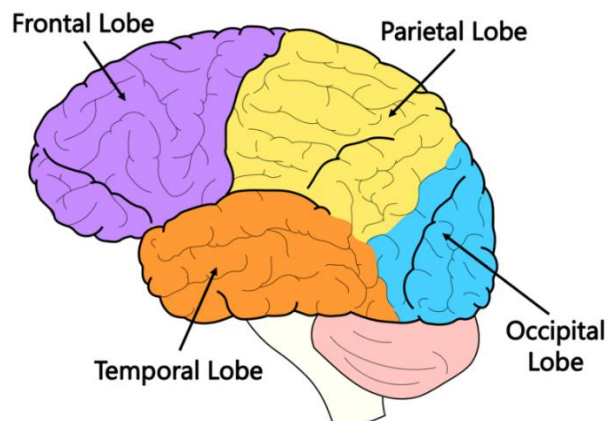
Figure 1.3: Lobes of the brain [13]

- Frontal Lobe: Cognitive tasks such as analytical thinking, decision-making, planning, joy, happiness, conscious thinking, change of feeling, problem solving, reasoning, and speaking skills are the functions of the frontal lobe. If it is damaged, there may be a change in mood and feeling.

- Parietal Lobe: It is responsible for combining and interpreting various sensations and signals from sensory neurons. This lobe governs literacy and numeracy skills. Damage to any nerve cell in these lobes can cause Alzheimer's disease.

- Temporal Lobe: This lobe detects all external factors regarding sound and smell. It is located at ear level in the two hemispheres of the brain. This lobe performs auditory tasks, identifies and remembers sounds. At the same time, complex stimuli such as faces and places are processed by this lobe.

- Occipital Lobe: It is the brain's center for processing visual information. A large part of it is the visual cortex. Receiving information through the eye retina, the lobe interprets it instantly. Any damage to this lobe leads to visual impairment.

1.1.2 Limbic System

The limbic system is the common name for the brain regions deep in the brain that are mainly responsible for the regulation of emotions and behavior, long-term memory, motivation, and the processing of the sense of smell [14]. The limbic system is the part of our brain that is associated with emotions and impulses. This system consists of many interconnected brain structures and includes various functions. This makes it difficult to know precisely the structures that make up the system and the concrete functions of each. In general, the limbic system consists of four main structures called the hypothalamus, hippocampus, thalamus and amygdala. The Figure 1.4 shows the important parts of limbic system.

- Hypothalamus: This is the center where emotional behaviors and inner glands are arranged. The autonomic nervous system is also controlled by the hypothalamus.

- Amygdala: It receives information from all senses. It regulates the emotion mechanism, especially fear and aggression. Amygdala has been shown to play a role in the understanding of emotional facial expressions. The amygdala recognizes happy facial expressions, particularly expressions of fear, distress and

danger, and it is stated that it is important in evaluating the threatening and threatening emotional states [15].

- Talamus: It collects nerve stimuli from the eyes and the ear, conducts initial processing and transmits it to the areas of vision and hearing in the brain shell [15]. If the thalamus works incorrectly, there are difficulties in processing sensory information.

- Hippocampus: Stores memories in long-term memory. Saves short memories by placing them in various parts of the brain. People who suffer from hippocampus have difficulty remembering new events, faces, places and conversations because they have problems in placing them in their long-term memories [15].
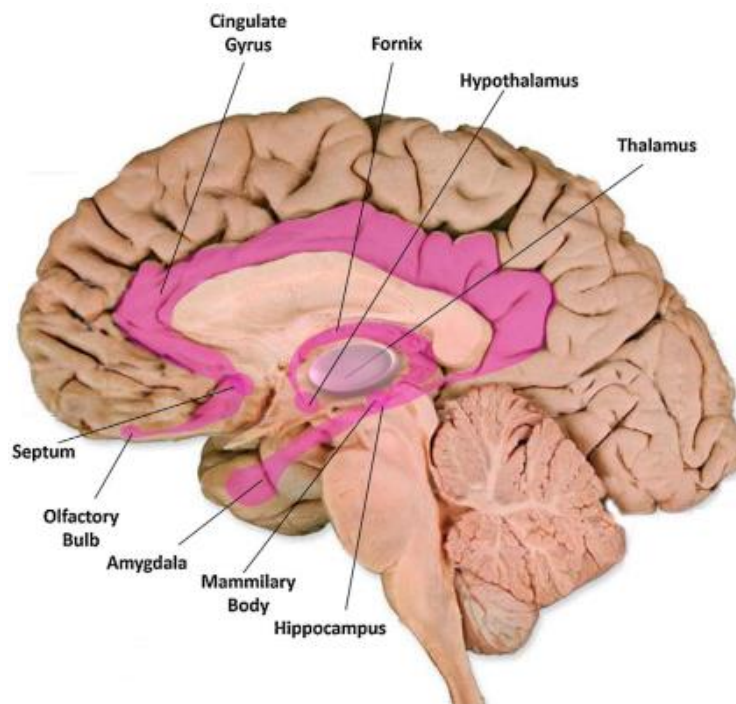
Figure 1.4: Parts of limbic system [16]

1.2 EEG

The brain produces a very low intensity continuous electrical current and spreads the waves regularly, EEG is the process of recording these waves in a computer environment. Electroencephalography is one of the methods used in the detection and

investigation of normal and abnormal functions of living human brain. They are often used to understand structural disorders in brain research. EEG analysis means electrical activity in the human brain. Measuring electrical activity reflects how the different neurons in the brain move and communicate with each other.
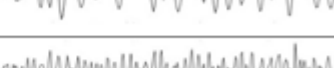
EEG is obtained with the help of electrodes placed around the skull. EEG waves are not physically full periodic but rhythmic waves. The frequency of observed potential EEG oscillations according to the activity of the brain varies in the range of 0.5 - 70 Hz and amplitudes of observed potential EEG oscillations varies in the range of 5 - 400 µV. As the activity level of the brain increases, the frequency of EEG waves increases and amplitude decreases. The most important parameter in the evaluation of EEG waves is frequency and the second parameter is amplitude.

Brain waves are irregular. Different waves occur depending on the situation. EEG waves in normal subjects, it is classified as alpha (α), beta (β), gamma (γ), theta (θ) and delta (Δ).

1.2.1 EEG Signals and Subbands of EEG Signals

According to frequency EEG oscillations can be divided into 5 categories [17]. These are briefly shown in Table 1.1.

Table 1.1: The five frequency bands of EEG signal

Frequency Band Name	Symbol of Band Name	Frequency Bandwidth	Example of EEG Signals
Delta	Δ	1–4 Hz	
Theta	θ	4–8 Hz	
Alpha	α	8–12 Hz	
Beta	β	12–30 Hz	
Gamma	γ	30-100+ Hz	

Delta Waves: creates the lowest frequency band and occurs during deep sleep.

Theta Waves: reflects the state of awakening and sleep. This wave is seen in the case of thoughtful and dreaming.

Alpha Waves: occurs when a person is awake. It indicates comfort and calmness. Alpha rhythm is best seen when the person is awake, at rest, relax and eyes closed.

Beta Waves: Beta activity is a partially low-voltage normal activity seen in frontal and central regions with a frequency of 12-30 Hz. This wave is active when the eyes are open, thinking, decision making and problem solving.

Gamma Waves: This wave is seen during extreme mental activities. Gamma waves are considered to be the optimum operating frequency of the brain. Gamma waves are often associated with increased sense of compassion and happiness and optimal brain functioning. Gamma waves are associated with increased awareness and mental skills.

1.2.2   Measurement of EEG Signals

Accurate measurement of EEG signals is very important for accurate analysis of brain signals. The EEG uses the International 10/20 system. The most commonly used of these standardized coordinate systems is the international 10/20 system. Placing the electrodes with the International 10-20 system ensures that all the scalp is coated identically. In EEG recordings, electrodes are placed in standard coordinates [18]. The EEG signal consists of different brain waves reflecting the electrical activity of the brain in certain brain regions. According brain regions, electrode placements are labelled: F (frontal), C (central), T (temporal), P (posterior) and O (occipital). The Figure 1.5 shows the electrode placements according to International 10/20 system.
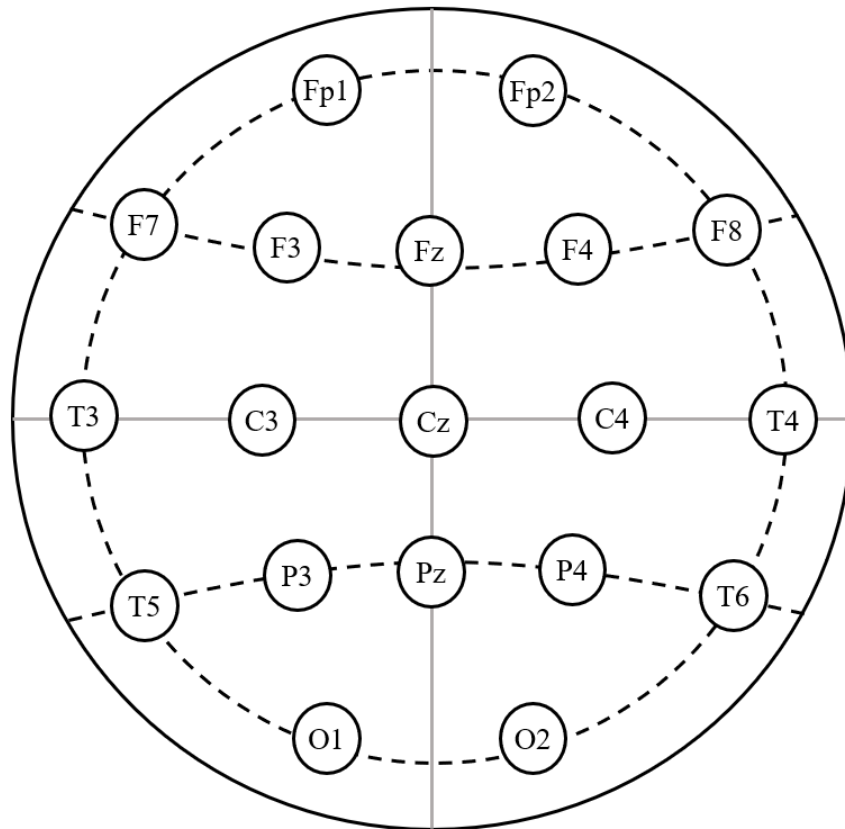
Figure 1.5: Electrode placements according to International 10/20 system [19]

1.3   Emotional Phenomena

Emotion phenomena can vary from person to person. Genetic factors and socio-cultural mechanisms and unique personal experiences are very important determinants of the personality functioning and human relations of individuals [20]. Not only a person's culture, but also neuro physiological processes are effective in identifying people's behavior and in different states of emotion. Emotion is a subjective experience and it tends to create images and cognitions that affect subsequent effects and behaviors. Neuro-physiologically, emotion is associated with innate neural firing in a situation.

1.3.1   The Classification of Emotions

Two different types are used for emotion models. One of them is discrete and the other is dimensional. Discrete models consist of basic emotions such as joy, anger, sadness,

fear, disgust and surprise. These basic emotions are described as "discrete," believed to be distinguishable by an individual's facial expression and biological processes [21].

In dimensional model, emotions are expressed according to their regions in the coordinate plane and it is a universal model. Dimensional emotion models describe human emotions by describing where emotions are in 2D or 3D [22]. The 2D model includes the arousal and valence dimensions of emotions. In general, arousal is shown on the y-axis and valence on the x-axis. In general, valence refers to the state of being happy or unhappy. Arousal refers to the state of calmness or excitement. The 3D model is affected by 3 different factors. These are Valence, Arousal and Dominance [24]. Dominance means control of emotion. These form the VAD model. The Figure 1.6 shows the 2D model consisting of valence and arousal dimensions. The Figure 1.7 shows the 3D emotion classification model consisting of arousal, valence and dominance.
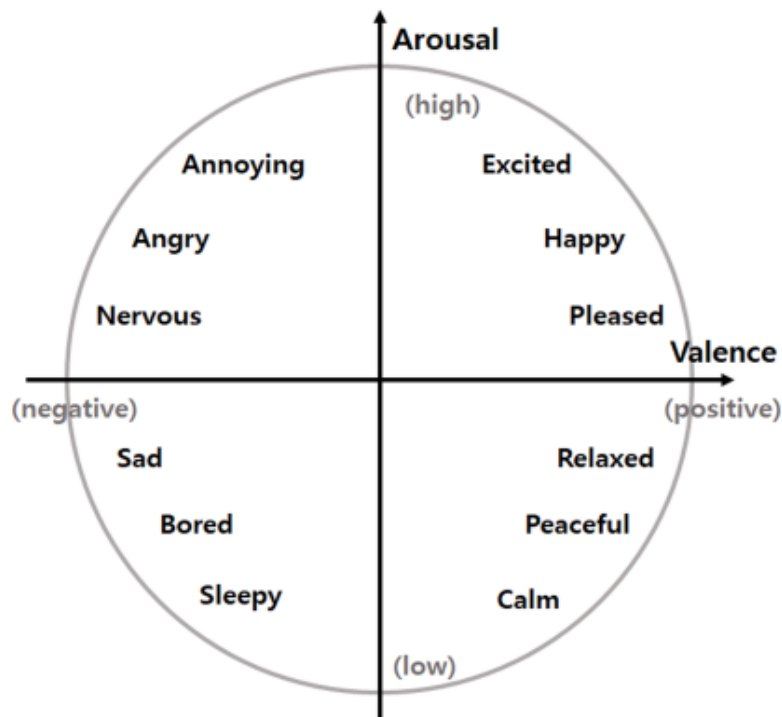
Figure 1.6: The 2D model consisting of valence and arousal dimensions [23]
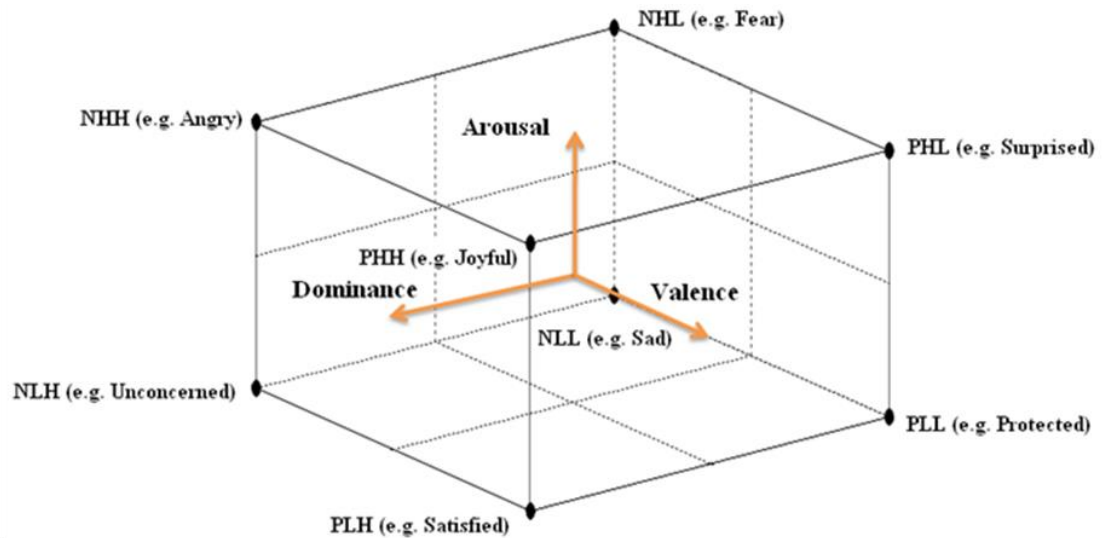
Figure 1.7: 3D emotion classification model [24]

The VAD model consists of 8 different emotion states [24]. Their expression is given below:

• Positive Valence - Low Arousal - Low Dominance (PLL) refers to "protected".

• Positive Valence - High Arousal - Low Dominance (PHL) refers to "surprise".

• Positive Valence - Low Arousal - High Dominance (PLH) refers to "satisfied".

• Positive Valence - High Arousal - High Dominance (PHH) refers to "happy".

• Negative Valence - Low Arousal - Low Dominance (NLL) refers to "sad".

• Negative Valence - High Arousal - Low Dominance (NHL) refers to "fear".

• Negative Valence - Low Arousal - High Dominance (NLH) refers to "unconcerned".

• Negative Valence - High Arousal - High Dominance (NHH) refers to "angry".

1.3.2 Influences on Emotion

Emotion is a complex psycho-physiological change that occurs in the individual's mood with the interaction of internal and environmental influences. It plays a central role in human daily life. Emotion is universal. They actually have many sources.

Emotions occur with the perception of a stimulus such as a situation, object or thought. There is a series of bodily changes that accompany or prepare it and result in an action or reaction of the individual. In other words, an emotion has both cognitive, physiological and behavioral qualities. People around the world express basic emotions in much the same way, regardless of their cultural differences, race, gender, or education. It is possible to conclude that basic emotions are common across cultures and are an innate biological tendency.

R. Plutchik [25] suggested in the 1980s that there are 8 basic emotions. Fear, surprise, sadness, disgust, anger, hope, joy, and acceptance each enable us to adapt to our environment in different ways. Emotions are not just psychological abstract concepts. They are also biological. So, some of the factors that affect your biology also affect your emotions. For this, factors such as exercise, age and sleep are important sources that affect our emotions and moods. The sources that affect our emotions can be listed as follows: personality, weather, stress, sleep, age and exercise.

Personality: Individuals with different personality structures also have different degrees of feeling emotion.

Weather: Research shows that the weather has little effect on your emotions. On the contrary, there are studies that argue that the weather and the time spent outside affect the mood of the people.

Stress: Stress is a state of tension that occurs in an individual due to extraordinary demands, pressures or opportunities. A little stress helps you get things done because it stimulates the brain. In addition, too much stress is harmful for people and the work to be done.

Sleep: Insomnia affects our emotions. According to a study, less sleep reduces job satisfaction and makes people more sensitive.

Exercise: A healthy life requires a certain amount of movement and effort. This is why exercises are important. Studies have shown that regular exercise helps people feel better.

Age: According to a study conducted with people between the ages of 18-94, the likelihood of negative moods decreases as individuals get older, and positive moods of older people last longer and negative moods are shorter than younger ones.

Contrary to rational thought, emotions do not occur voluntarily, they can exist outside of one's consciousness. And physiological responses also occur outside of consciousness. The most important task in emotional mechanisms belongs to the brain structures that make up the limbic system. A significant portion of the limbic system's responses to stimuli is hereditary, according to researchers.

1.4   Emotional Brain-Computer Interfaces

Computer-based systems that detect, analyze and transmit signals related to the desired action in the brain are called brain-computer interface (BCI) or brain-machine interface [6]. In principle, it can be used with any brain signal. It is possible to capture signals through a chip implant in the brain or electrodes placed on the scalp. This activity is usually measured with EEG. The EEG device is not considered an interface by itself. Because this device only records brain signals and does not produce any output around the user when used alone.

The BCI systems are computer-based systems that collect and analyze brain signals and convert these signals into commands to transfer them to output devices in order to perform the desired actions as a result of this analysis. Brain computer interfaces are divided into three classes: invasive, partial invasive, and non-invasive.

In invasive techniques, special devices must be used to collect data (brain signals), these devices are implanted directly into the human brain through critical surgery. In partial invasive techniques, the devices are placed on the skull from the top of the human brain. In general, non-invasive devices [26] are considered the safest and least costly device type. However, these devices can only capture "weak" human brain signals due to skull occlusion. In this method, the detection of brain signals is provided by electrodes placed on the scalp. EEG has unique usability advantages over other types of brain signal recording recommended for commercial use. It is easy to use, portable and inexpensive. In addition, EEG provides high temporal resolution and

various solutions to improve signal localization. The Figure 1.8 shows the non-invasive brain-computer interface.
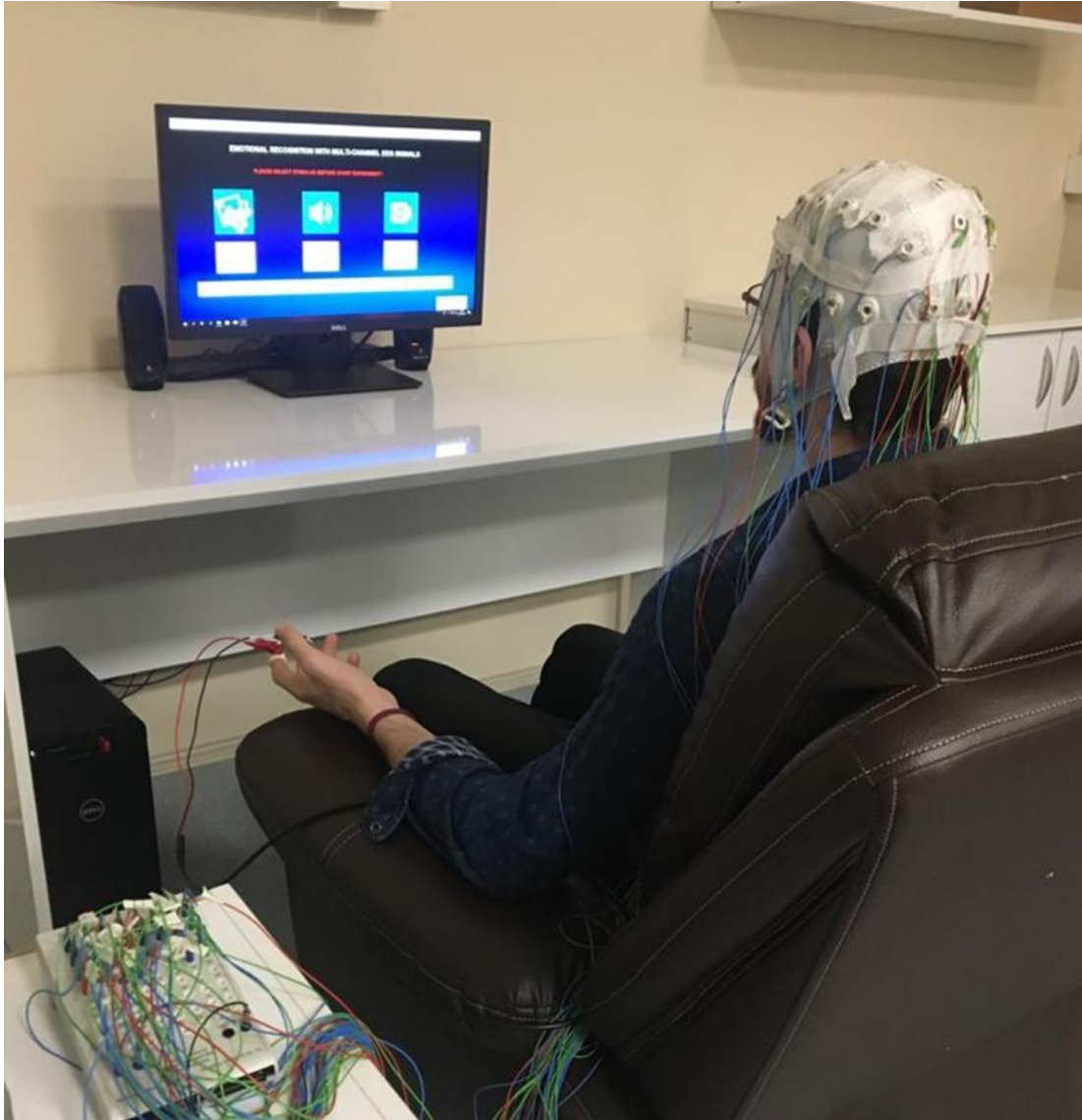
Figure 1.8: Brain-Computer Interface

The function of BCIs is based on the detection of the brain signals showing the purpose of the user, and the conversion of the appropriate commands to the device to be managed after the necessary measurements and analysis are made. This system basically consists of four main components: signal acquisiton, feature determination, conversion and output.

Signal acquisition process involves measuring brain signals using a specific sensor technology [27]. In feature determination step, purpose-relevant signals are separated from foreign content and analyzed. This step requires a strong relationship and harmony between the user's purpose and signals. Relevant signals obtained are transferred to the algorithm. The algorithm to be used must be dynamic in order to adapt to instant changes and to cover the full control of the device. Commands make the external device work. As a result, feedback is provided to the user and the loop is closed.

BCIs bring human convenience and comfort through mind control of machines. BCI only requires the inclusion of brain signals so that a set of commands can be performed, and this does not require muscle intervention. With BCI, robots can support people with disabilities in their lives. BCI systems contribute to the detection of diseases such as mental state monitoring function, abnormal brain structure, seizure disorder, sleep disorder and brain swelling. It can be used for patients who are unable to regain previous levels of mobility to help the patient regain normal functionality. Brain-computer interfaces can also be leveraged to provide greater security. BCI connectivity provides us the ability to communicate telepathically with another person without the need for speech.

Chapter 2

Literature Review

A wide variety of studies have been conducted within the scope of EEG analysis in the literature. Based on the literature review, there have been a lot of studies about emotion recognition from EEG signals by using stimulus, emotion classification using stimulated EEG signals applying machine learning algorithms and deep learning algorithms for EEG signal classification. Classical machine learning methods have limitations in terms of processing very large data. As the number of classes increases in machine learning applications, the problem becomes more complex and the classification success of the system decreases. In the literature, there are studies in which deep learning methods are applied to various classification problems in recent years. In the classification of EEG signals, using deep learning models has become widespread. Deep learning networks are very important in terms of both time and cost. It provides an effective solution for solving complex classification problems. The advantages of deep learning networks constitute the source of motivation for this thesis.

Isah Salim Ahmad and Shuai Zhang et al. [28] offered the convolutional neural network (CNN) model for the learning feature and recognizing the emotion of positive, neutral, and negative of pure EEG signals single model based on the SEED dataset with ResNet50 and Adam optimizer. As stimulus fifteen video clips were selected for the experiment. The fifteen Chinese participants participated in the experiment and the 60-channel EEG cap was used to record the EEG signals. The dataset is a shuffle and then split into two, 70% for training and 30% for testing and fed the signal into the model. The model was trained, predictions were made and the model performance accuracy was evaluated. Negative emotion has the accuracy rate as 94,86%, neutral

emotion has the accuracy rate as 94,29% and positive emotion has the accuracy rate as 93,25%. The average accuracy for the SEED dataset found as 94.13%.

Rofiqoh H. Dien Haqque and Esmeralda Contessa Djamal et al. [29] used 2D Convolutional Neural Networks (CNN) for emotion recognition and Wavelet filters are used to obtain a frequency range of 4-45 Hz through the decomposition stage. In this study, SJTU Emotion EEG Dataset (SEED) was used and 12 channels, which are more effective FT7, FT8, T7, T8, C5, C6, TP7, TP8, CP5, CP6, P7 and P8, were used. In this study, the first experiment was conducted to measure the use of one- and two-dimensional CNN types. The two-dimensional CNN (2D CNN) method produces an accuracy rate of 83.44%, with 75.97% accuracy than one-dimensional CNN (1D CNN). Two different optimization models were used as Maximum Adaptive Moment Estimation (AdaMax) and Stochastic Gradient Descent (SGD). As a result, eight layers with SGD weight correction resulted in the best accuracy of 83.44% and the lowest Loss of 0.769.

Nesma E. Elsayed and Ahmed S. Tolba et al. [30] offered a new method that the User-Independent Hybrid Brain-computer Interface (UIHBCI) model. A fourteen-channel EEG device was used to record the responses of nine participants' brain signals. They described the relevant multi-sensory features of multi-channel EEG by audio or video and male or female.

The Deep Belief Network (DBN) was applied and for evaluation four models were tested. These models are Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Brain-computer Interface Lower-Limb Motor Recovery (BCI LLMR) and Hybrid Steady-State Visual Evoked Potential Rapid Serial Visual Presentation Brain-computer Interface (Hybrid SSVEP-RSVP BCI). At the end of the study, 94.44% success rate was achieved in the proposed model, while SVM, LDA, BCI Lower Limb motor recovery and Hybrid SSVEP-RSVP BCI achieved 61.11%, 66.67%, 83.33% and 89.67%, respectively. When the results are examined, it is shown that the proposed DBN is the most successful classifier, promising and improvable.

Jingxia Chen and Dongmei Jiang et al. [31] proposed a method for EEG data representation. This representation method converts 1D chain-like EEG vector sequences into 2D mesh-like matrix sequences. For this, the researchers used the

sliding window method to divide 2D matrix arrays into segments, which are the EEG sample that integrates the temporal and spatial correlation of the raw EEG recordings. Furthermore, cascade and parallel hybrid convolutional recurrent neural networks have been proposed to accurately predict the emotional category of EEG signals. Result of this study, they achieved that the recognition accuracy of the two hybrid models reached 93% for both valence and arousal.

Minmin Miao and Wenjun Hu et al. [32] investigated a novel spatial-frequency-temporal (SFT) 3D CNN model. They proposed MI EEG 3D for feature representation and SFT 3DCNN model was proposed for classification. Weight value visualization is applied to investigate the accuracy of the SFT-3DCNN model. There are 5 subjects and the effectiveness of the proposed method was evaluated on two general MI EEG datasets. At the end of the study, the average accuracy rate of the five subjects was 86.6% for BCI Contest III Dataset, while the average accuracy rate for the BCI Contest III dataset IIIa was 91.85%.

Vaishali M. Joshi and Rajesh B. Ghongade [33] aimed to perform emotion detection using Electroencephalography (EEG) signal based on Linear Formulation of Differential Entropy feature extractor and BiLSTM network classifier. A method has been proposed to distinguish between positive, negative and neutral emotions in the SEED database, and valence and arousal in the DEAP database. Subject conditional, unconditional, and interdependent experiments were conducted to evaluate the proposed model in the SEED database, and the average accuracy of emotion detection was improved by 4.12% for the subject conditional approach, 4.5% for the unconditional approach, and 1.3% for the interdependent approach. It was found that feature extractor LF-DfE was better with BiLSTM network than other methods.

Anjana K. and Ganesan M. [34] proposed a method of EEG data classification of SEED database having three emotions like positive, neutral and negative. Image encoding and classification of emotions were done in this study. First, the data is converted into an image that is used to analyze the EEG signal, and then deep learning algorithms are used to understand the emotions encountered during the converted image EEG signal generation. In the proposed model, various image coding approaches such as Spectrogram, Scalogram and HHT (Hilbert Huang Transform)

were used. Experimental results showed that the image coding scalogram provides the best classification accuracy of 98%. The spectrogram and HHT have classification accuracy of 78% and 75%, respectively.

Ante Topic and Mladen Russo [35] proposed a model that based on the generation of feature maps based on topographic (TOPO-FM) and holographic (HOLO-FM) representation of EEG signal features. Deep learning was used as a feature extraction method in feature maps, and then the obtained features were classified to recognize different emotion types. Experiments were performed on four emotion data: DEAP, SEED, DREAMER, and AMIGOS. As a result of the study, it is seen that the proposed method is successful in different datasets and has a high success rate when compared to other studies.

Md. Rabiul Islam and Md. Milon Islam [36] proposed a deep machine learning-based model. One-dimensional EEG data acquired using the DEAP dataset were converted to Pearson Coefficient of Correlation (PCC) featured images of channel correlation of EEG subbands. Using the CNN model, images are given to this structure. The experiment, called protocol-1, was used to describe the two levels, and the experiment, called protocol-2, was used to recognize the emotional trivalent and arousal level. As a result of the study, they reached 78.22% maximum ACC on valence and 74.92% maximum ACC on arousal.

Chapter 3

Materials and Methods

The aim of this study is to perform emotion classification by using signal processing algorithms to EEG signals obtained from the 32-channel EEG device with visual and and auditory stimuli shown to individuals respectively. In the proposed method, it is aimed to obtain success rates in the valence, activation and dominance emotion dimensions from multi-channel EEG signals with the multivariate synchrosqueezing transform method, which gives better results than most methods in the literature. In addition, effective features in emotion classification will be investigated and the applied algorithms will be compared by applying Machine Learning and Deep Learning algorithms to multi-channel EEG signals. The characteristics of the data set used for this project and the details of the methods used in the thesis for the classification of EEG signals are explained.

In this study, emotion classification system is done step by step. First, pictures and audios are used to stimulate subjects to reveal emotions that are represented on the 3D valence-arousal-dominance emotion plane. At the same time, biological EEG and EDA signals are recorded while they are under effects of stimuli and the subjects mark the SAM (Self-Assessment Manikin) form after each stimulus according to their emotions. The experiment protocol is designed in MATLAB GUI and lasted about 45 minutes for each subject. After the obtaining EEG signal data, EEG data is segmented and EEG signals are preprocessed. An output of 32-channel EEG signals is obtained for each stimulus using the multivariate synchrosqueezing transform method. A 3D TF (time frequency) output is obtained and this output is converted to 2D. The resulting 2D picture is given as input to the deep learning architecture, the emotions are classified and the results are obtained.

A 32-electrode EEG device and SAM will be used for data collection. With the EEG device, signals existing in the human body can be received harmlessly without the need for doctor's supervision. EEG signals generated in individuals against different stimuli will be recorded with the device, and performance analysis will be made by comparing the SAM form after each stimulus and the results of various algorithm studies applied on the individual's EEG signals. The Figure 3.1 shows the steps of emotion classification system.
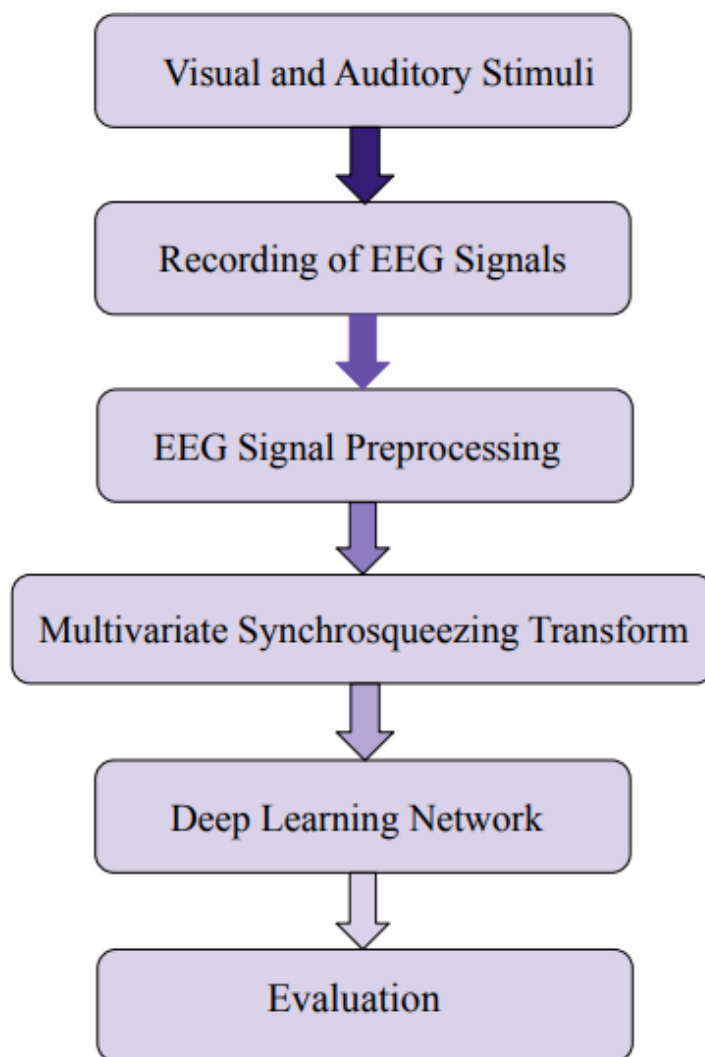
Figure 3.1: Steps of emotion classification system

3.1 Data Acquisition

In this thesis, two types of stimuli were used to obtain EGG signals, as visual and auditory stimuli and the study was carried out in two stages. EEG recordings of 25 healthy participants were used, including 13 female and 12 male volunteers for two stages. The EEG signals were taken from a 32-channel Brain Vision Recorder EEG device by using 60 visual and 60 auditory stimuli that could stimulate various emotions. Two databases are set up. From IAPS database [37], the 48 images are used to evoke emotions. Then from IADS database [38], the 48 sounds are used as the auditory stimuli. Visual and auditory stimuli were selected from IAPS and IADS according to certain criteria and other stimuli 12 images and 12 sounds were selected by us in the same criterion. EEG data of participants were recorded by Brain Vision Recorder EEG device. The details of the participants are given in Table 3.1.

Table 3.1: Information about the age of the participants

INFORMATION		Female	Male
Number of Subjects	For visual and auditory experiment	13	12
Age Range of Subjects	For visual and auditory experiment	[19 – 24]	[20 – 27]
Average of Age	For visual and auditory experiment	22,5	23,1

3.2 Matlab – Graphical User Interface (GUI)

Graphical user interface (GUI) allows users to use electronic devices more easily with icons, windows, buttons and panels. GUI isn't just for computers. GUIs are found in smartphones, portable music players, refrigerators, printers, gadgets, and more. Any screen that displays information and accepts user input on a screen benefits from a graphical user interface [39]. In this study, GUI was designed thanks to many codes

written in MATLAB. Specific visuals, commands and warnings are used to make it easier for users to understand and complete the experiment more easily.

The 3 different parts are designed in GUI as visual, auditory and video. However, only the visual and auditory parts were used in the experiment. The Figure 3.2 shows the design window of GUI and the Figure 3.3 shows the window presented to users.
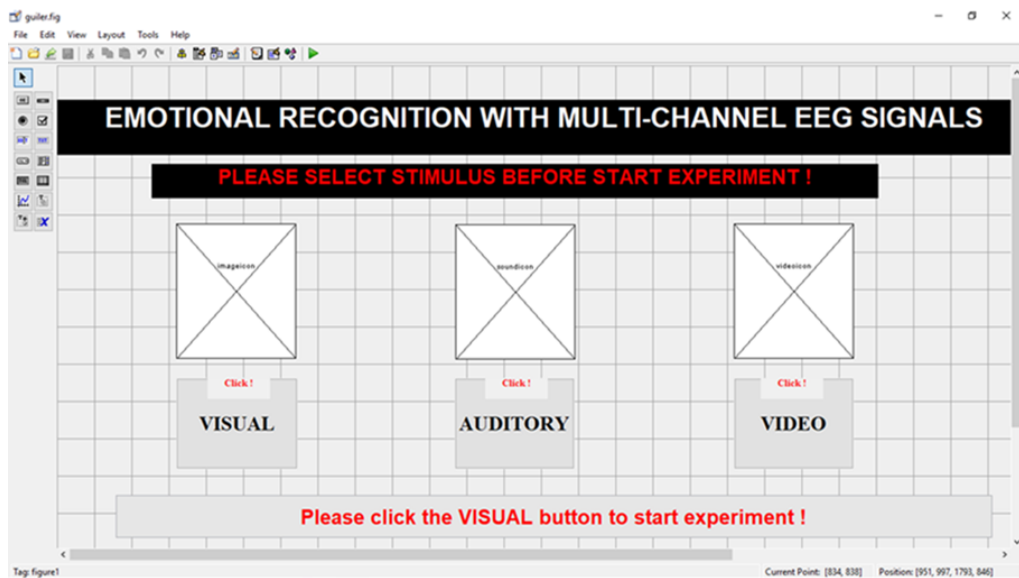
Figure 3.2: Design window of graphical user interface in fig format
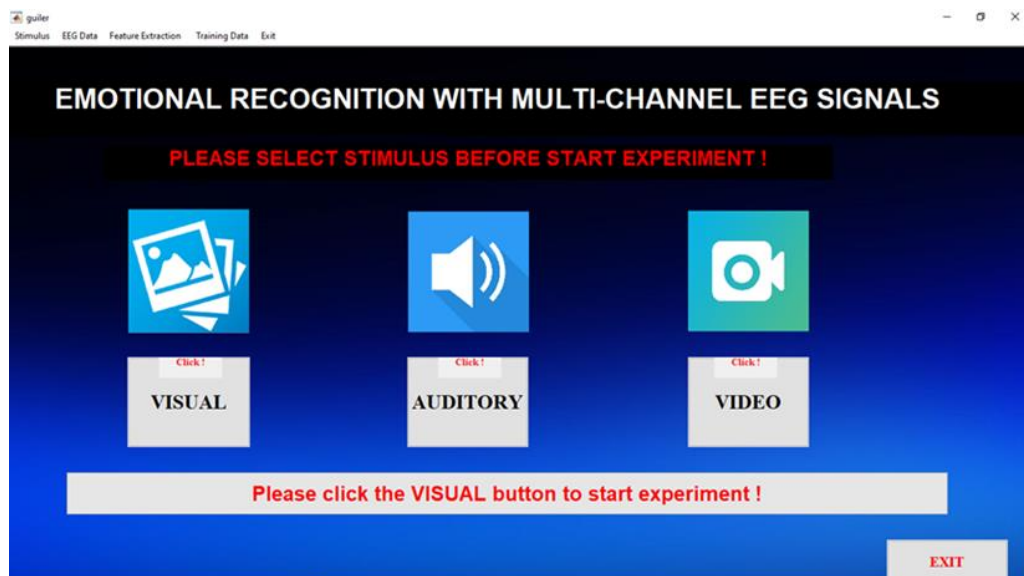
Figure 3.3: Presented window of graphical user interface to users

Stimulus were provided to users by the designed GUI. After each stimulus, the questionnaire (SAM) was designed and the results of this survey were recorded in the text file. At certain time intervals, stimuli were shown to the participants and a specific time was designed to fill the questionnaire. The delays during the experiment were also recorded in the text file. Thanks to the GUI, people who participated in the experiment successfully completed the experiment with the given commands.

3.3   Determination of Stimuli

3.3.1   The International Affective Picture System (IAPS)

IAPS is an image set and database designed to study the emotions and attention of people used in psychological research [25]. In 2005, National Institute of Mental Health Center for Emotion and Attention at the University of Florida develops the IAPS [37]. This picture set consists of a lot of colorful different pictures stimulating various emotions. Also, this dataset includes images of daily events, objects, home furniture, family, landscapes, disabled people, food and animals. IAPS images have been used in EEG studies and many other psychophysiological measurements. In addition, it is forbidden to distribute and share of these images.

3.3.1.1   Determination of Visual Stimuli

The 48 pictures selected from IAPS and 12 selected images with the same characteristics are shown as visual stimuli to 25 people who participate in this study. A total of 60 visual stimuli were used. The images selected from IAPS are in JPEG format and the maximum size is 1024 x 768 pixels. The 12 images we selected are also in JPEG format and the size is 1024 x 768 pixels. In the visual experiment, according to the images' location in the Arousal, Valence and Dominance 3 - dimensional map, 48 images were chosen to induce eight emotional states. The experiment composed of 8 session and these are Positive-Low Arousal-Low Dominance (PLL), Positive-High Arousal-Low Dominance (PHL), Positive-Low Arousal-High Dominance (PLH), Positive-High Arousal-High Dominance (PHH), Negative-Low Arousal-Low Dominance (NLL), Negative-High Arousal-Low Dominance (NHL), Negative-Low Arousal-High Dominance (NLH), Negative-High Arousal-High Dominance (NHH). 6

images are selected for each section. ID numbers from 1 to 48 are assigned to the IAPS images used in the experiment, respectively. ID numbers from 49 to 60 are defined for the remaining 12 images. ID number for each image is given and these images are shown in the specified order. The pictures we selected are also given ID number, added to the end of the IAPS images and shown after the IAPS images. The details of IAPS images used to target emotions in each session, their ID number, Valence, Arousal and Dominance mean values of IAPS are given in Table 3.2 for all subjects according to whether they are high or low value and positive or negative value. If the mean values are greater than 5, they are taken as high or positive, if they are less than 5, they are taken as low or negative.

Table 3.2: The details of IAPS images according to Valence, Arousal and Dominance value

Valence				Arousal				Dominance			
Positive		Negative		High		Low		High		Low	
ID	Value	ID	value	ID	value	ID	value	ID	value	ID	value
1	6.52	25	1.95	7	6.23	1	4.45	13	6.05	1	4.55
2	5.28	26	2.20	8	6.02	2	4.88	14	6.00	2	4.75
3	6.67	27	3.32	9	6.99	3	4.60	15	5.83	3	4.15
4	6.54	28	3.10	10	7.35	4	4.44	16	6.12	4	4.70
5	5.19	29	3.69	11	6.61	5	4.97	17	6.21	5	4.26
6	5.22	30	3.16	12	7.31	6	4.78	18	6.56	6	4.40
7	6.65	31	1.79	19	5.69	13	4.61	19	6.49	7	4.29
8	7.35	32	1.91	20	6.28	14	4.31	20	5.67	8	4.96
9	6.48	33	1.80	21	6.28	15	3.73	21	6.14	9	4.73
10	7.33	34	1.90	22	6.14	16	2.90	22	6.17	10	4.70
11	7.09	35	1.90	23	6.07	17	3.20	23	5.51	11	4.63
12	7.21	36	1.89	24	6.35	18	3.55	24	6.17	12	4.63
13	8.19	37	4.71	31	5.25	25	4.53	37	5.55	25	3.22
14	8.21	38	4.22	32	6.76	26	4.88	38	5.70	26	3.75
15	6.56	39	4.89	33	6.82	27	3.95	39	5.39	27	4.72
16	7.14	40	4.98	34	7.29	28	3.67	40	5.64	28	3.47
17	7.36	41	4.06	35	6.54	29	3.93	41	5.32	29	4.85
18	7.06	42	3.28	36	6.01	30	4.48	42	5.17	30	4.22
19	7.12	43	3.59	43	5.11	37	3.74	43	5.21	31	3.7
20	7.38	44	3.17	44	5.39	38	3.77	44	5.03	32	3.69
21	8.10	45	4.69	45	5.28	39	4.33	45	5.06	33	2.81
22	7.74	46	3.76	46	5.17	40	2.77	46	4.43	34	2.73
23	7.63	47	2.95	47	4.61	41	3.93	47	6.04	35	3.21
24	7.54	48	2.67	48	4.63	42	4.35	48	4.40	36	3.22

The valence, arousal and dominance values are obtained by taking the average value of 25 participants for the remaining 12 pictures added to the experiment by us. The details of these values are given in Table 3.3 according to whether they are high or low value and positive or negative value for 25 participants.

Table 3.3: The details of Valence, Arousal and Dominance values for the remaining 12 images

Valence				Arousal				Dominance			
Positive		Negative		High		Low		High		Low	
ID	Value	ID	Value	ID	value	ID	value	ID	value	ID	value
49	7.20	53	4.20	49	5.92	50	4.92	49	5.72	-	
50	6.64	-		51	5.36	52	4.08	50	6.12	-	
51	6.92	-		60	5.12	53	4.44	51	5.64	-	
52	6.0	-		-		54	4.60	52	5.72	-	
54	7.20	-		-		55	4.72	53	5.04	-	
55	5.84	-		-		56	4.72	54	5.46	-	
56	7.52	-		-		57	4.32	55	6.12	-	
57	5.72	-		-		58	3.60	56	5.60	-	
58	5.28	-		-		59	4.84	57	5.04	-	
59	6.92	-		-		-		58	6.92	-	
60	7.52	-		-		-		59	5.75	-	
-		-		-		-		60	5.60	-	

The 2D valence-arousal emotion model describes emotional states. The 2D model can provide variations of emotion changes. Different emotions can be found in various positions in the 2D plane. There are 4 different groups in the 2D model. These are Positive Valence – Low Arousal, Positive Valence – High Arousal, Negative Valence – Low Arousal and Negative Valence – High Arousal. Classification of stimuli with ID numbers from 1 to 60 according to their valence and arousal values as a 2D model is given in Table 3.4.

The 3D emotion classification model contains the valence, arousal and dominance values. According to valence, arousal and dominance values, the 8 different combinations are occurred. The 8 groups for 3D model are given in Table 3.5.

According to arousal, valence and arousal values of the other 12 images selected by us, 3D emotion classification is applied and the details of this are given in Table 3.6.

Table 3.4: The details of image stimulus according to 2D model

	ID	Valence	Arousal
	1	6.52	4.45
	2	5.28	4.88
	3	6.67	4.60
	4	6.54	4.44
	5	5.19	4.97
	6	5.22	4.78
	13	8.19	4.61
	14	8.21	4.31
First Group	15	6.56	3.73
	16	7.14	2.90
Positive Valence	17	7.36	3.20
-	18	7.06	3.55
Low Arousal	50	6.64	4.92
	52	6.0	4.08
	54	7.20	4.60
	55	5.84	4.72
	56	7.52	4.72
	57	5.72	4.32
	58	5.28	3.60
	59	6.92	4.84
	7	6.65	6.23
	8	7.35	6.02
	9	6.48	6.99
	10	7.33	7.35
	11	7.09	6.61
Second Group	12	7.21	7.31
	19	7.12	5.69
Positive Valence	20	7.38	6.28
-	21	8.10	6.28
High Arousal	22	7.74	6.14
	23	7.63	6.07
	24	7.54	6.35
	49	7.20	5.92
	51	6.92	5.36
	60	7.52	5.12
	25	1.95	4.53
	26	2.20	4.88
Third Group	27	3.32	3.95
	28	3.10	3.67
Negative Valence	29	3.69	3.93
-	30	3.16	4.48
Low Arousal	37	4.71	3.74
	38	4.22	3.77
	39	4.89	4.33
	40	4.98	2.77

Table 3.5 (continued)

Third Group	41	4.06	3.93
	42	3.28	4.35
	53	4.20	4.44
Forth Group Negative Valence - High Arousal	31	1.79	5.25
	32	1.91	6.76
	33	1.80	6.82
	34	1.90	7.29
	35	1.90	6.54
	36	1.89	6.01
	43	3.59	5.11
	44	3.17	5.39
	45	4.69	5.28
	46	3.76	5.17
	47	2.95	4.61
	48	2.67	4.63

Table 3.6: Details of stimuli used in visual experiment for 48 IAPS images

			Mean values of IAPS		
	IAPS Number	ID	Valence	Arousal	Dominance
Positive Valence- Low Arousal- Low Dominance (PLL)	1810	1	6.52	4.45	4.55
	1908	2	5.28	4.88	4.75
	5890	3	6.67	4.60	4.15
	5990	4	6.54	4.44	4.70
	7497	5	5.19	4.97	4.26
	7632	6	5.22	4.78	4.40
Positive Valence- High Arousal- Low Dominance (PHL)	1650	7	6.65	6.23	4.29
	5470	8	7.35	6.02	4.96
	8179	9	6.48	6.99	4.73
	8030	10	7.33	7.35	4.70
	8400	11	7.09	6.61	4.63
	8492	12	7.21	7.31	4.63
Positive Valence- Low Arousal- High Dominance (PLH)	1440	13	8.19	4.61	6.05
	1460	14	8.21	4.31	6.00
	2358	15	6.56	3.73	5.83
	2370	16	7.14	2.90	6.12
	5200	17	7.36	3.20	6.21
	7325	18	7.06	3.55	6.56
Positive Valence- High Arousal- High Dominance (PHH)	4599	19	7.12	5.69	6.49
	7405	20	7.38	6.28	5.67
	8190	21	8.10	6.28	6.14
	8470	22	7.74	6.14	6.17
	8499	23	7.63	6.07	5.51
	8200	24	7.54	6.35	6.17

Table 3.7 (continued)

	2205	25	1.95	4.53	3.22
Negative	2375.1	26	2.20	4.88	3.75
Valence-Low	2490	27	3.32	3.95	4.72
Arousal-Low	9001	28	3.10	3.67	3.47
Dominance	2399	29	3.69	3.93	4.85
(NLL)	9471	30	3.16	4.48	4.22
	2095	31	1.79	5.25	3.7
Negative	3030	32	1.91	6.76	3.69
Valence-High	3530	33	1.80	6.82	2.81
Arousal-Low	6350	34	1.90	7.29	2.73
Dominance	9635.1	35	1.90	6.54	3.21
(NHL)	9325	36	1.89	6.01	3.22
	2279	37	4.71	3.74	5.55
Negative	2280	38	4.22	3.77	5.70
Valence-	2309	39	4.89	4.33	5.39
Low Arousal-	2397	40	4.98	2.77	5.64
High Dominance	2525	41	4.06	3.93	5.32
(NLH)	2715	42	3.28	4.35	5.17
	7360	43	3.59	5.11	5.21
Negative	1274	44	3.17	5.39	5.03
Valence-High	2458	45	4.69	5.28	5.06
Arousal-High	9594	46	3.76	5.17	4.43
Dominance	9831	47	2.95	4.61	6.04
(NHH)	2276	48	2.67	4.63	4.40

Table 3.8: Details of 12 images stimuli used in visual experiment with 25 participants

			Mean values for All Subject		
	Number	ID	Valence	Arousal	Dominance
PHH	1	49	7.20	5.92	5.72
PLH	2	50	6.64	4.92	6.12
PHH	3	51	6.92	5.36	5.64
PLH	4	52	6.0	4.08	5.72
NLH	5	53	4.20	4.44	5.04
PLH	6	54	7.20	4.60	5.46
PLH	7	55	5.84	4.72	6.12
PLH	8	56	7.52	4.72	5.60
PLH	9	57	5.72	4.32	5.04
PLH	10	58	5.28	3.60	6.92
PLH	11	59	6.92	4.84	5.75
PHH	12	60	7.52	5.12	5.60

3.3.2 The International Affective Digitized Sound System (IADS)

To provide a set of normative emotional stimuli for experimental investigations of emotion recognition and attention, the International Affective Digitized Sounds (IADS) was developed [38]. The IADS is developed and distributed by the National Institute of Mental Health Center for Emotion and Attention at the University of Florida. Sound stimulants set includes different sounds. Thanks to these sound stimuli, people's emotion and attention can be examined. Arousal, valence and dominance values of emotions can be used with this dataset. The audio dataset is used for many research. It is mainly used to stimulate emotions and to detect emotion status with EEG signal. In addition, it is used as a stimulus for many studies that measure heart rate, skin conductivity and electrodermal activity. It is forbidden to share and publish this audio data anywhere like IAPS.

3.3.2.1 Determination of Auditory Stimuli

In the auditory experiment, the 48 sounds selected from IADS and 12 selected sounds with the same characteristics are shown as auditory stimuli to 23 people who participate in this study. The sounds selected from the IADS are in WAV format and are 6 seconds long. These sounds have different arousal, valence and dominance values. They are voices that appeal to different emotions. In addition, the 12 sounds we selected are in the form of wav and are 6 seconds long. A total of 60 auditory stimuli were used. EEG recordings were obtained by listening to the auditory stimuli consisting of 48 sounds selected from IADS and 12 sounds selected by us.

The auditory stimuli were given to the participants according to the ID number. According to the sounds' location in the arousal, valence and dominance 3-dimensional map, 48 sounds were chosen to induce eight emotional states. Different sounds have been chosen for each eight emotional states to stimulate different emotions. The details of IADS sounds used to target emotions in each session, their ID number, Valence, Arousal and Dominance mean values of IADS are given in Table 3.7 and the Table 3.8 shows that details of remaining 12 sounds according to whether they are high or low value and positive or negative value.

Table 3.9: The details of IADS sounds according to Valence, Arousal and Dominance value

Valence				Arousal				Dominance			
Positive		Negative		High		Low		High		Low	
ID	value	ID	value	ID	value	ID	value	ID	value	ID	value
1	5.31	25	3.54	6	6.04	1	4.6	13	6.07	1	4.53
2	5.26	26	4.01	7	5.51	2	2.88	14	5.91	2	4.87
3	5.15	27	4.83	8	7.54	3	4.75	15	6.21	3	4.6
4	5.99	28	4.83	9	5.62	4	3.77	16	6.36	4	4.85
5	5.18	29	4.34	10	5.38	5	4.12	17	6.29	5	4.85
6	5.01	30	4.52	11	6.55	13	4.46	18	6.07	6	4.56
7	5.23	31	2.05	12	5.15	14	4.38	19	6.14	7	4.78
8	6.94	32	1.65	19	6.03	15	3.36	20	6.09	8	4.73
9	5.19	33	1.68	20	7.12	16	4.51	21	5.77	9	4.83
10	6.02	34	1.68	21	7.07	17	3.36	22	6.41	10	4.86
11	6.46	35	2.04	22	6.72	18	4.18	23	6.86	11	4.8
12	5.09	36	2.42	23	6.85	25	4.94	24	6.44	12	4.67
13	6.84	37	4.63	24	7.15	26	4.75	37	5.36	25	4.08
14	6.95	38	4.83	31	8.16	27	4.97	38	5.07	26	4.33
15	6.62	39	4.88	32	7.61	28	4.65	39	5.19	27	4.66
16	6.82	40	4.68	33	7.95	29	3.51	40	5.62	28	4.53
17	7.44	41	4.86	34	7.88	30	4.42	41	5.76	29	4.64
18	7.51	42	4.72	35	7.99	37	4.91	42	5.4	30	4.93
19	7.64	43	4.52	36	7.98	38	4.65	43	5.04	31	2.55
20	7.65	44	4.45	44	5.37	39	4.6	44	5.23	32	2.89
21	7.17	45	4.3	45	5.79	40	4.03	45	5.33	33	2.3
22	7.33	46	4.86	46	5.89	41	4.18	48	5.06	34	2.31
23	7.9	47	3.65	47	5.33	42	4.35			35	2.29
24	7.67	48	4.96	48	5.37	43	4.87			46	4.59
										47	4.37

Table 3.10: The details of Valence, Arousal and Dominance values for the remaining 12 sounds for 25 participants

Valence				Arousal				Dominance			
Positive		Negative		High		Low		High		Low	
ID	value	ID	Value	ID	value	ID	value	ID	value	ID	value
49	6.09	60	4.83	50	6.83	49	4.35	49	5.87	50	4.87
50	6.61	-		51	6.43	52	4.74	51	5.52	-	
51	7.65	-		53	5.17	58	4.48	52	5.52	-	
52	6.0	-		54	5.52	59	4.96	53	5.52	-	
53	6.3	-		55	5.39	60	3.87	54	5.35	-	
54	7.61	-		56	5.22	-		55	6.04	-	
55	6.7	-		57	6.0	-		56	6.09	-	
56	7.04	-		-		-		57	5.96	-	
57	5.83	-		-		-		58	5.22	-	
58	5.04	-		-		-		59	5.87	-	
59	6.26	-		-		-		60	5.43	-	

According to 2D emotions classification, there are Positive Valence – Low Arousal, Positive Valence – High Arousal, Negative Valence – Low Arousal and Negative Valence – High Arousal groups. Classification of sounds with ID numbers from 1 to 60 according to their valence and arousal values as a 2D model is given in Table 3.9.

Table 3.11: The details of all sound stimulus according to 2D model

	ID	Valence	Arousal
First Group Positive Valence - Low Arousal	1	5.31	4.6
	2	5.26	2.88
	3	5.15	4.75
	4	5.99	3.77
	5	5.18	4.12
	13	6.84	4.46
	14	6.95	4.38
	15	6.62	3.36
	16	6.82	4.51
	17	7.44	3.36
	18	7.51	4.18
	49	6.09	4.35
	52	6.0	4.74
	58	5.04	4.48
	59	6.26	4.96
Second Group Positive Valence - High Arousal	6	5.01	6.04
	7	5.23	5.51
	8	6.94	7.54
	9	5.19	5.62
	10	6.02	5.38
	11	6.46	6.55
	12	5.09	5.15
	19	7.64	6.03
	20	7.65	7.12
	21	7.17	7.07
	22	7.33	6.72
	23	7.9	6.85
	24	7.67	7.15
	50	6.61	6.83
	51	7.65	6.43
	53	6.3	5.17
	54	7.61	5.52
	55	6.7	5.39
	56	7.04	5.22
	57	5.83	6.0

Table 3.12 (contunied)

	25	3.54	4.94
	26	4.01	4.75
	27	4.83	4.97
	28	4.83	4.65
Third Group	29	4.34	3.51
	30	4.52	4.42
Negative Valence	37	4.63	4.91
-	38	4.83	4.65
Low Arousal	39	4.88	4.6
	40	4.68	4.03
	41	4.86	4.18
	42	4.72	4.35
	43	4.52	4.87
	60	4.83	3.87
	31	2.05	8.16
	32	1.65	7.61
Forth Group	33	1.68	7.95
	34	1.68	7.88
Negative Valence	35	2.04	7.99
-	36	2.42	7.98
High Arousal	46	4.86	5.89
	47	3.65	5.33
	44	4.45	5.37
	45	4.3	5.79
	48	4.96	5.37

The 3D emotion classification model contains the valence, arousal and dominance values. According to these values of sounds, the 8 combinations are occurred. The 8 groups of the combinations for 3D model are given in Table 3.10. According to arousal, valence and arousal values of the other 12 sounds selected by us, 3D emotion classification is applied and the details of this are given in Table 3.11.

Table 3.13: Details of stimuli used in sound experiment for 48 IADS sounds

			Mean values of IADS		
	IADS Number	ID	Valence	Arousal	Dominance
	170	1	5.31	4.6	4.53
Positive Valence	262	2	5.26	2.88	4.87
Low Arousal	368	3	5.15	4.75	4.6
Low Dominance	602	4	5.99	3.77	4.85
(PLL)	698	5	5.18	4.12	4.85

Table 3.14 (contunied)

	114	6	5.01	6.04	4.56
	152	7	5.23	5.51	4.78
Positive Valence	360	8	6.94	7.54	4.73
High Arousal-	364	9	5.19	5.62	4.83
Low Dominance	400	10	6.02	5.38	4.86
(PHL)	415	11	6.46	6.55	4.8
	425	12	5.09	5.15	4.67
	112	13	6.84	4.46	6.07
Positive Valence	150	14	6.95	4.38	5.91
Low Arousal	172	15	6.62	3.36	6.21
High Dominance	726	16	6.82	4.51	6.36
(PLH)	809	17	7.44	3.36	6.29
	810	18	7.51	4.18	6.07
	110	19	7.64	6.03	6.14
Positive Valence	311	20	7.65	7.12	6.09
High Arousal	352	21	7.17	7.07	5.77
High Dominance	367	22	7.33	6.72	6.41
(PHH)	815	23	7.9	6.85	6.86
	817	24	7.67	7.15	6.44
	250	25	3.54	4.94	4.08
Negative Valence	252	26	4.01	4.75	4.33
Low Arousal	722	27	4.83	4.97	4.66
Low Dominance	627	28	4.83	4.65	4.53
(NLL)	708	29	4.34	3.51	4.64
	723	30	4.52	4.42	4.93
	275	31	2.05	8.16	2.55
	290	32	1.65	7.61	2.89
Negative Valence	279	33	1.68	7.95	2.3
High Arousal	286	34	1.68	7.88	2.31
Low Dominance	424	35	2.04	7.99	2.29
(NHL)	712	36	2.42	7.98	2.84
	410	46	4.86	5.89	4.59
	280	47	3.65	5.33	4.37
	102	37	4.63	4.91	5.36
	246	38	4.83	4.65	5.07
Negative Valence	376	39	4.88	4.6	5.19
Low Arousal	700	40	4.68	4.03	5.62
High Dominance	720	41	4.86	4.18	5.76
(NLH)	728	42	4.72	4.35	5.4
	358	43	4.52	4.87	5.04
Negative Valence	702	44	4.45	5.37	5.23
High Arousal- High	729	45	4.3	5.79	5.33
Dominance (NHH)	104	48	4.96	5.37	5.06

Table 3.15: Details of 12 stimuli used in auditory experiment with 23 participants

			Mean values for All Subject		
	Number	ID	Valence	Arousal	Dominance
PLH	1	49	6.09	4.35	5.87
PHL	2	50	6.61	6.83	4.87
PHH	3	51	7.65	6.43	5.52
PLH	4	52	6.0	4.74	5.52
PHH	5	53	6.3	5.17	5.52
PHH	6	54	7.61	5.52	5.35
PHH	7	55	6.7	5.39	6.04
PHH	8	56	7.04	5.22	6.09
PHH	9	57	5.83	6.0	5.96
PLH	10	58	5.04	4.48	5.22
PLH	11	59	6.26	4.96	5.87
NLH	12	60	4.83	3.87	5.43

3.3.3 Evaluation of Visual - Auditory Stimuli

People who participated in the experiment filled in a questionnaire to evaluate their feelings after every image they saw and every sound they heard. After each visual and auditory stimulus, the subjects completed the questionnaire. Thanks to these stimuli, these questionnaires were made in order to express the emotional state of the participants. The SAM images were presented to the participants to evaluate their emotions. The SAM is a non-verbal pictorial questionnaire that directly measures emotional response [40]. GUI is designed in MATLAB, and the SAM scale is shown on the screen after each stimulus and the emotions are evaluated by the participant. Before the experiment, SAM User Instructions were explained to the participants.

In this SAM form, there are 4 scales rated from 1 to 9 and each stimulus is evaluated according to the degree of the emotions evoked by stimuli. Valence, arousal and dominance show three different types of emotions: happy and unhappy, excited and calm, controlled and uncontrolled. There is also liking scale according to the liking or dislike of stimuli. Briefly the valence scale refers to unhappiness and happiness. Arousal scale refers to the state of calmness and excitement. Dominance refers to domination and control of emotions. Liking scale refers to dislike of dislike. The Figure 3.4 shows the SAM scale designed in GUI.
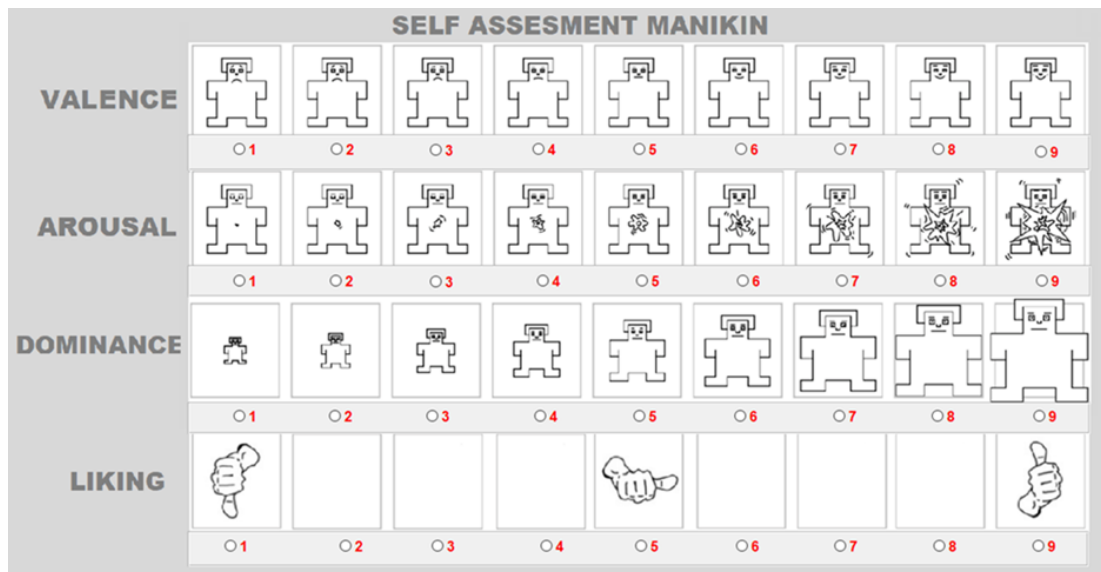
Figure 3.4: SAM for valence, arousal, dominance and liking in GUI

Valence, the first SAM scale, is the rating of happiness. At one end of the happy and unhappiness scale, happy, satisfied, contented and hopeful feelings are represented. If you feel completely happy while viewing / listening to the stimulus, you should click the button on the left. The other end of the scale is when you feel completely unhappy, annoyed, unsatisfied, melancholic, hopeless and bored, you should click on the button on the right side. If you feel completely neutral, you should click on the fifth figure in the middle. According to the decision and emotions of the participants, they can mark the option they want from 1 to 9.

Arousal, the second SAM scale is the rating of the dimensions of emotions of excitement and calmness. At one end of the scale, you have an excited, excited, enthusiastic, awakened feeling. If you feel completely excited while viewing / listening to the stimulus, you should click the button on the left. On the other hand, at the other end of the scale, completely relaxed, calm, sluggish, dull, sleepy feelings are represented, you should click on the button on the right side. If you're not at all excited or calm at all, you should click on the fifth figure in the middle. If you want to make a more precise assessment of how excited or calm you feel, you can mark the option you want from 1 to 9.

Dominance, the third SAM scale is the rating of controllable emotions and emotions that is under control. At one end of the scale, you have the feeling of being fully

controlled, influenced and directed. So, if your emotions are under control by the stimuli, you should click the button on the left. At the other end of this scale, you felt completely effective, important and dominant, you should click on the button on the right side. When the figure is large, you feel important and effective and when figure is small you feel, controlled and directed. If you feel that your emotions aren't under your control or you're not in control, you should click on the middle image. You can use the ratings between the figures to represent your feelings.

The liking assessment scale is related to whether the stimulus is liked or not. If the stimulus is liked, the choice should be made between 6 and 9, and between 1 and 4 if it is unlikely. If there is no liking or dislike, then the fifth button should be selected.

3.4   Tools Used

Some programs need to be used to record EEG and EDA signals. Also, we need to use some software as a platform to process recorded signals. In this section, the tools that we used throughout this project will be explained.

3.4.1   EEG Device (BrainVision Recorder and Analyzer)

The BrainVision Recorder is used to record EEG signals in this study. BrainVision Recorder is a powerful and practical recording program that used for recording neuro-/electrophysiological signals (e.g. EEG, EMG, ECG, EOG). BrainVision Recorder is a program usually used for psychological and neurophysiological research. It is used for recording, viewing and filtering of the signals. The BrainVision Recorder has an unlimited number of channels.

BrainVision Analyzer is used for the analysis of EEG data. Also it is used to process neuro-/electrophysiological signals. The BrainVision Analyzer program is used to convert the EEG signals that obtained in the experiment to the mat and vhdr format. The signals obtained in the vhdr format were used in MATLAB EEGLAB. In this study to process EEG signals, MATLAB was used. In this study 32 channels of EEG device are used to record EEG signals from participants. The setup of EEG device assembly is shown in Figure 3.5.
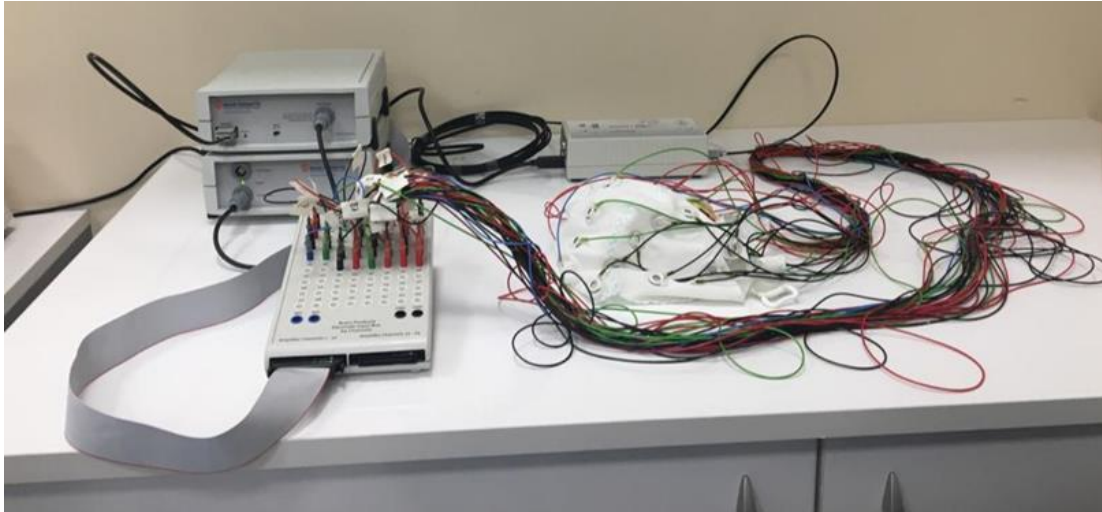
Figure 3.5: The setup of EEG device

In the EEG records, an EEG brain cap suitable for international 10-20 electrode placement system is used. The BrainCap is used for electrode contact and signal quality. The cap for the EEG recording is selected as appropriate according to the participants. Also gel is used for the best contact between electrode and scalp. The Figure 3.6 (a) shows the top view of BrainCap and (b) shows the right view of BrainCap. Ear-lobes are selected for reference and AFz was selected as ground.
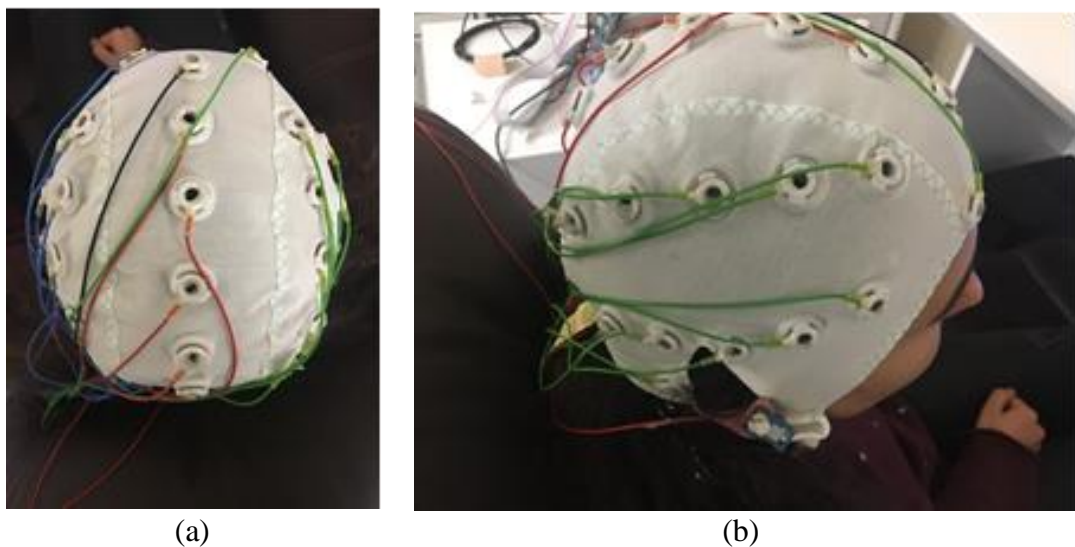

| (a) | (b) |

Figure 3.6: (a) Top view of BrainCap, (b) Right view of BrainCap

During the recording, all 32 channels are used for an active EEG signal. Ear-lobe is selected for reference and AFz is selected as ground. The Figure 3.7 displays the 32 electrodes positions on the scalp and the channel numbers and channel labels of the electrodes positioned for measuring the EEG signals are listed in Table 3.12.
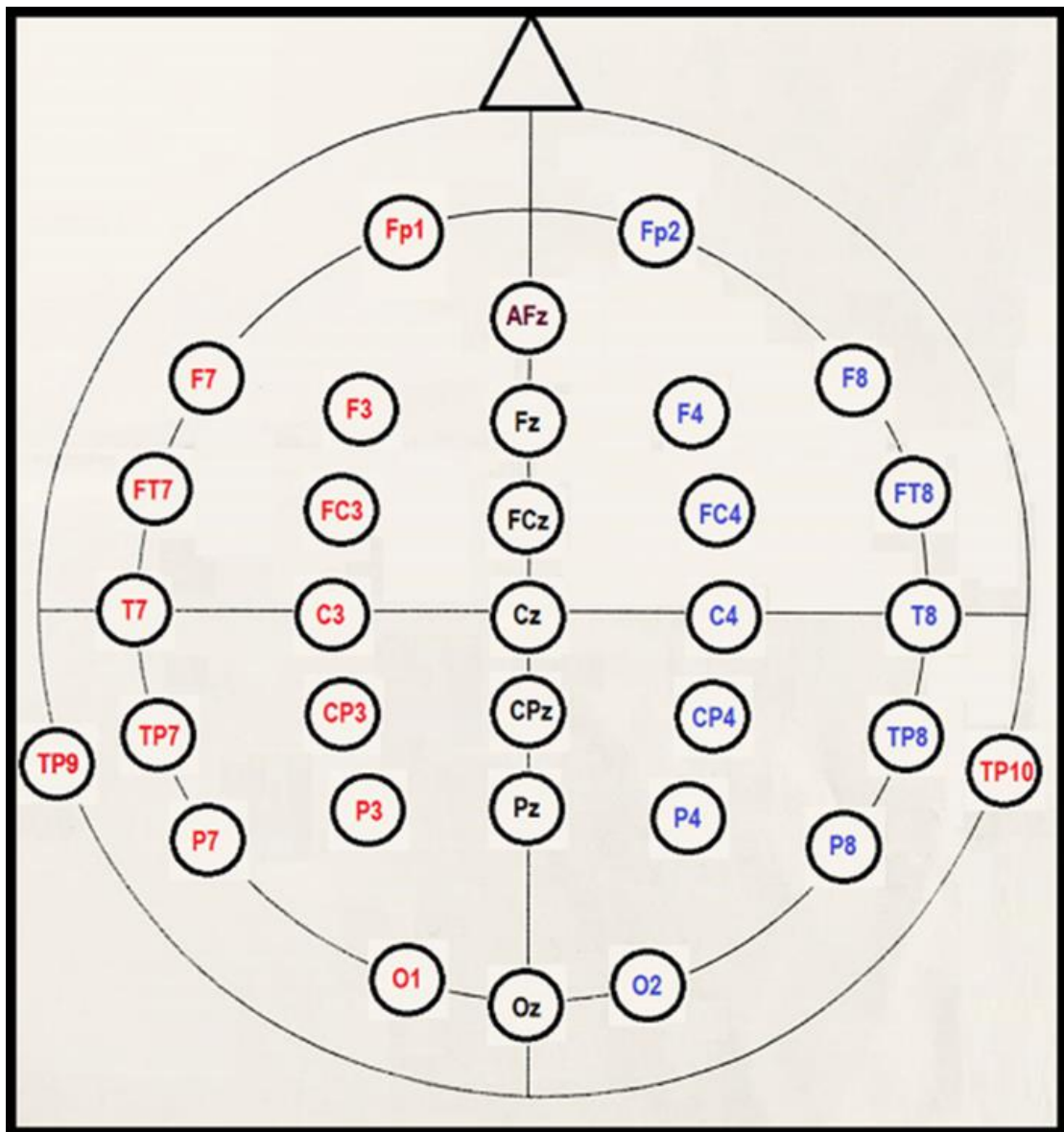

Figure 3.7: Displaying the 32 electrodes positions on the scalp

Table 3.16: EEG channel numbers and location of electrodes

Channel Numbers	Location of Electrodes
1	Fp1
2	Fp2
3	F7
4	F3
5	Fz
6	F4
7	F8
8	FT7
9	FC3
10	FCz
11	FC4
12	FT8
13	T7
14	C3
15	Cz
16	C4
17	T8
18	TP7
19	CP3
20	CPz
21	CP4
22	TP7
23	P7
24	P3
25	Pz
26	P4
27	P8
28	O1
29	Oz
30	O2
31	TP9
32	TP10

After the cap is attached to the top of the participants, the gels are applied to 32 channels. Here, impedance check is performed to observe the effectiveness of the gels with the BrainVision Recorder program. The impedance of ground, references and 32 channels as data is checked. In Figure 3.8, impedance check window is showed. If all the channels, ground and references shapes expressed as seen on the page appear to be

green, yellow or orange, the impedance control is completed. But if the colors of these shapes are red, it is not accepted. By applying the gel or by contacting the electrode with the full scalp, these shapes are provided to be green, yellow or orange to accept.
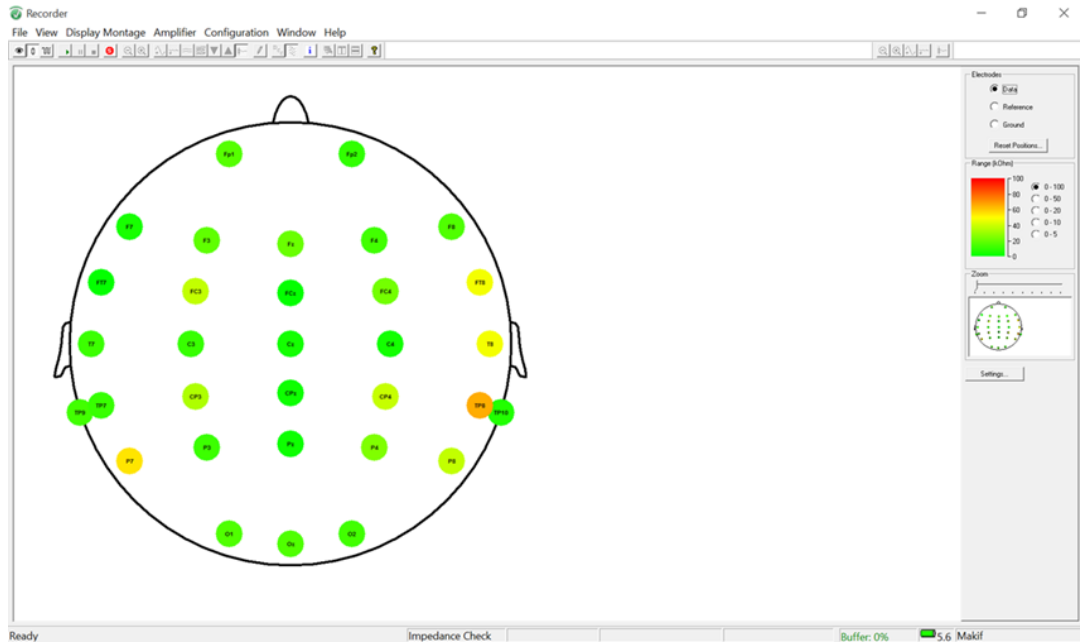

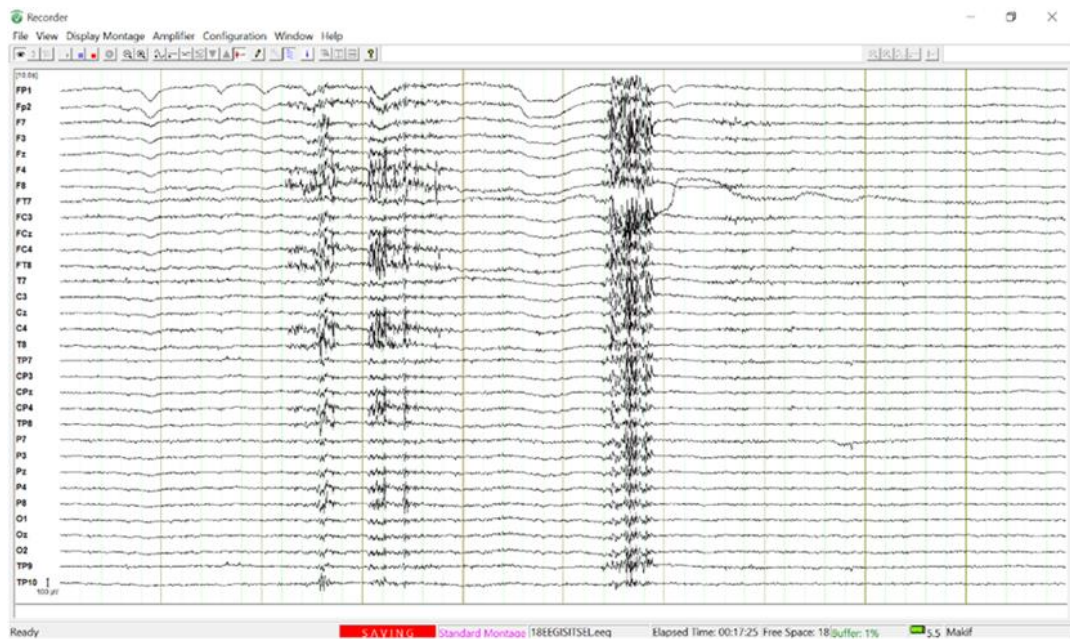Figure 3.8: Impedance check window

Figure 3.9: The taken EEG signals from participant during experiment

After all controls are performed, the experiment is started. EEG signals are recorded with the BrainVision Recorder during the experiment. The EEG signal taken at the same time when the stimulus is given to the participant in the experiment is shown in Figure 3.9.

3.4.2 EDA Device (BIOPAC Systems)

For recording electrodermal activity in the experiment, BIOPAC systems are used. BIOPAC Electrode Lead Test (SS57LA), two BIOPAC disposable electrodes for each subject. The electrodes used to measure electrodermal activity are different from the electrodes used to measure other common physiological signals, such as ECG. The electrodes used for EDA are rectangular and have a different concentration of electrode gel (0.5% chloride wet gel). The ion concentration in the gel is designed to be similar to the ion concentration in the sweat.

The EDA is typically measured on the palms and fingertips of the hands. These regions are chosen due to the density of eccrine sweat glands, which means an increased ability to detect changes in skin conductivity. The disposable electrodes used in experiment is shown in Figure 3.10 and Figure 3. 11 shows the the EDA signals received from the participant during the experiment.

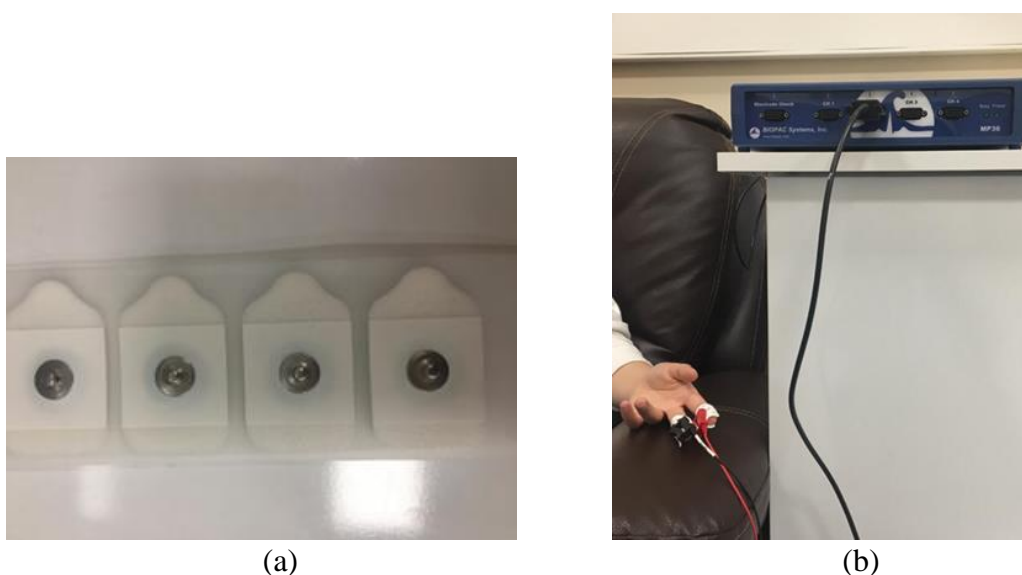(a)   (b)

Figure 3.10: (a) Disposable electrodes used in the experiment, (b) A photo taken from the participant during EDA recording
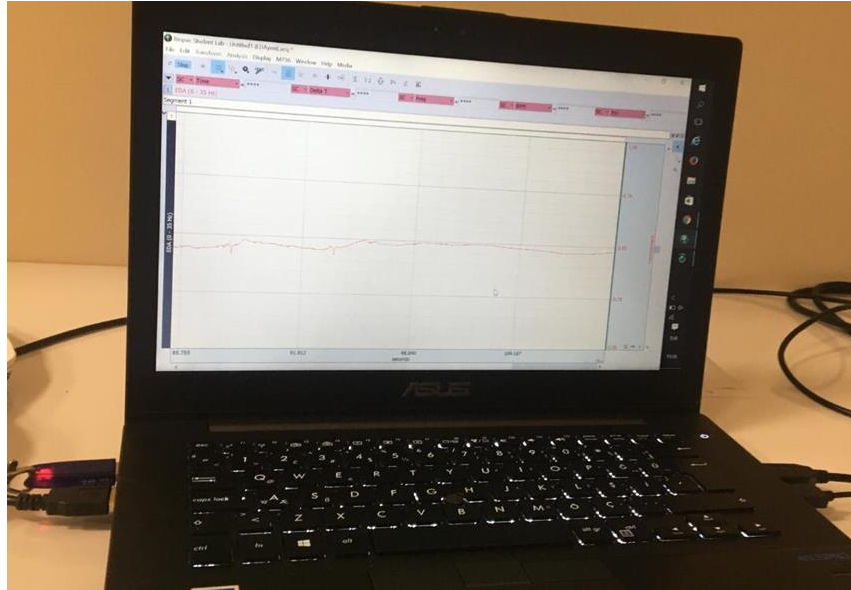
Figure 3.11: The EDA signals received from the participant during the experiment

3.4.3   MATLAB – EEGLAB

EEGLAB is a MATLAB toolbox used to process data from electrophysiological signals. It is an open source. In this study, EEG signals recorded by BrainVision Recorder were opened in EEGLAB in .vhdr format. The Figure 3.12 show the EEGLAB window and the Figure 3.13 shows 32-channel EEG signal in EEGLAB. The EEGLAB window for .vhdr file is shown in Figure 3.14. These EEG signals are recorded in .mat format to process signals in MATLAB.
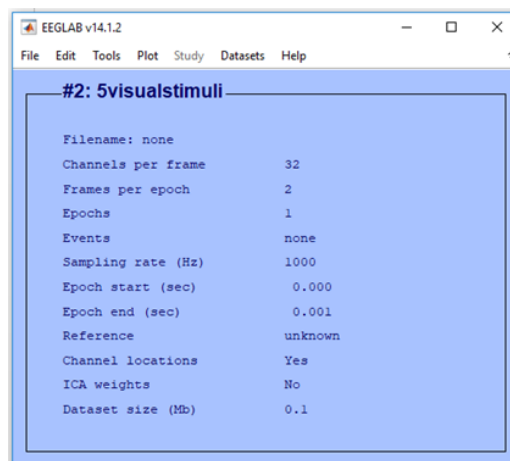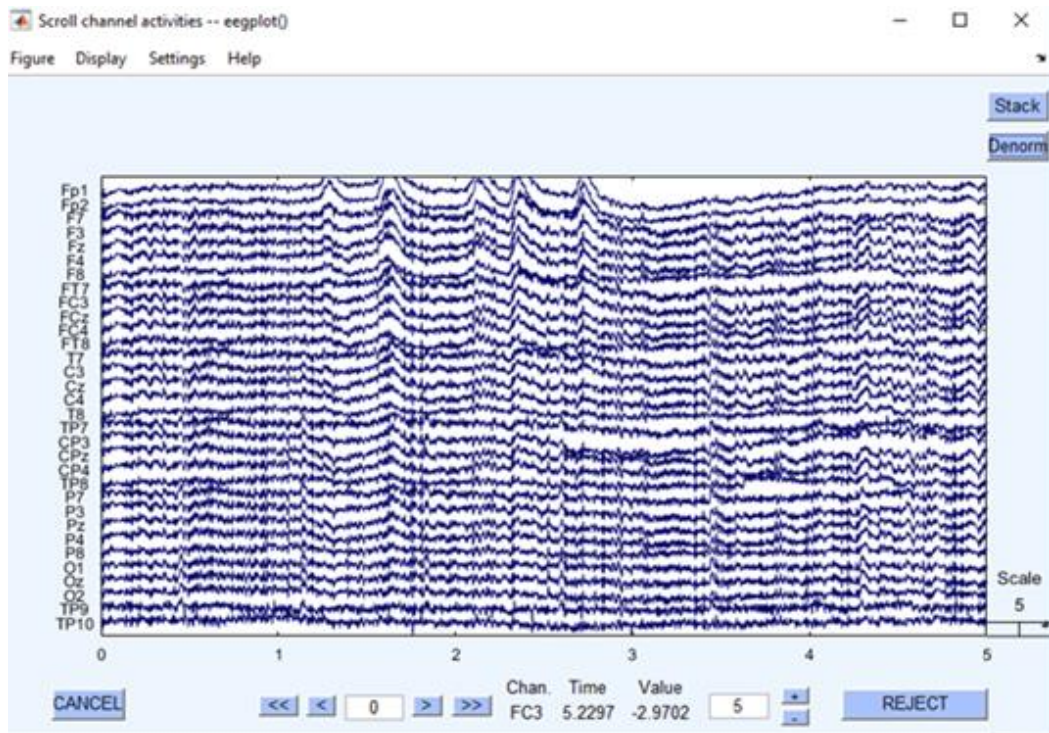

Figure 3.12: The EEGLAB window

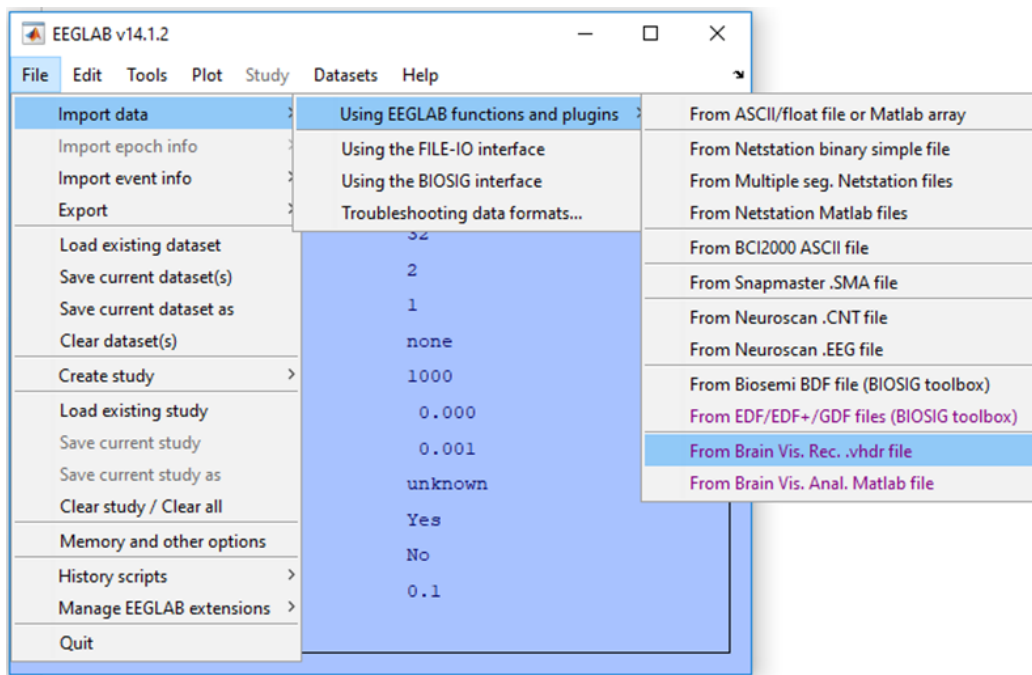Figure 3.13: Plotting 32-channel EEG signal in EEGLAB

Figure 3.14: EEGLAB window for .vhdr file

3.5   Experimental Data

3.5.1   Experiment Protocol

The study included 25 healthy individuals (19 to 27 years). Before the experiment was started, the experimental user instructions were given to participants (each step in the experimental phase), the experiment was explained and the ethical form was signed by the participants. First, the EEG signal is recorded using 32 electrode channels (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP7, P7, P3, Pz, P4, P8, O1, Oz, O2, TP9, TP10) in which the raw signal is stored during the experiment [41]. The sampling rate of the EEG signal used is 1000 Hz. A detailed description of the experimental protocol is given in Table 3.13.

Table 3.17: Experimental protocols

Number of Subjects	Number of Chanels	Sampling Frequency (Hz)	Total Experiment Time (Each person)
25	32	1000	49 minutes

The graphical user interface is designed for the stimulation of emotions in this study. The experiment consists of two parts. The first section is the part where visual stimuli are used. The 60 pictures were used, including 48 IAPS pictures and 12 similar images. Examples from used IAPS images are shown in Figure 3.15. Each picture is shown for 6 seconds. The survey that lasts 14 seconds after the display of each picture is finished appears on the screen. Participants must complete the questionnaire within 14 seconds. After this survey, 1 second black screen is shown. The purpose of selecting a black color is that the black color does not generate any stimuli and is a neutral color. Then the experiment continues in the order of picture, survey and black screen. The visual experiment lasts 22 minutes. After the visual experiment, there is a 5-minutes rest time for the participants. The Figure 3.16 shows the model of the demonstration plan for the visual stimulus experiment.
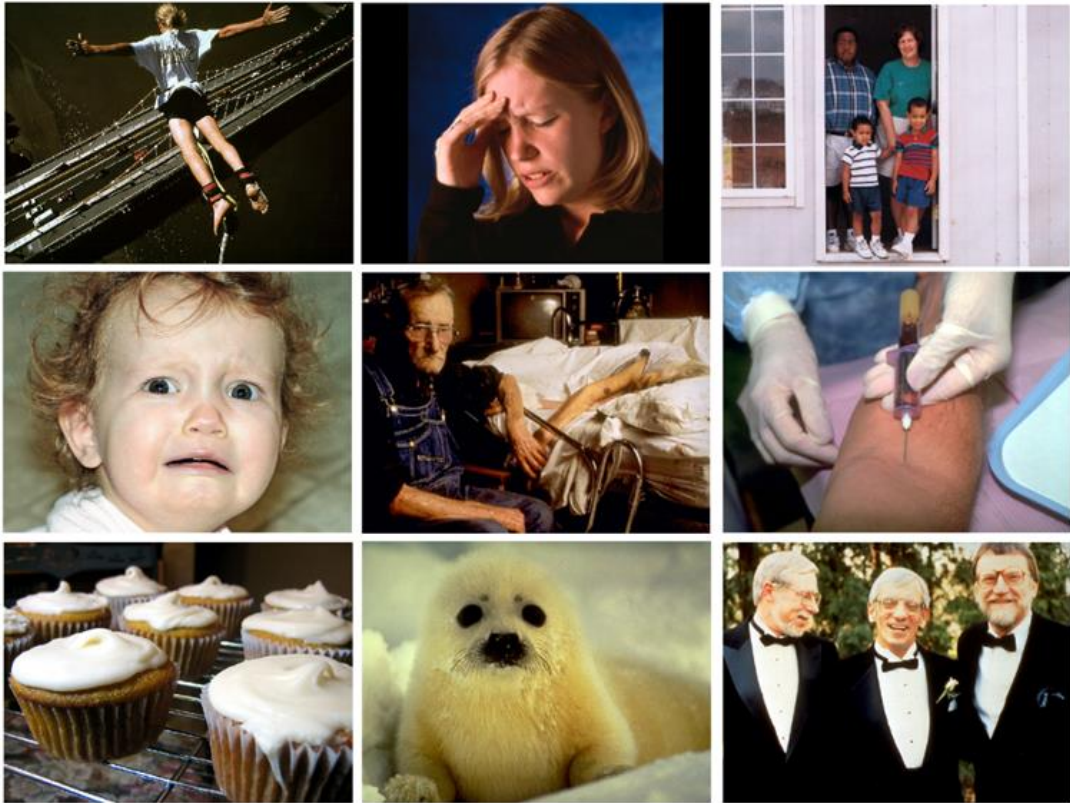
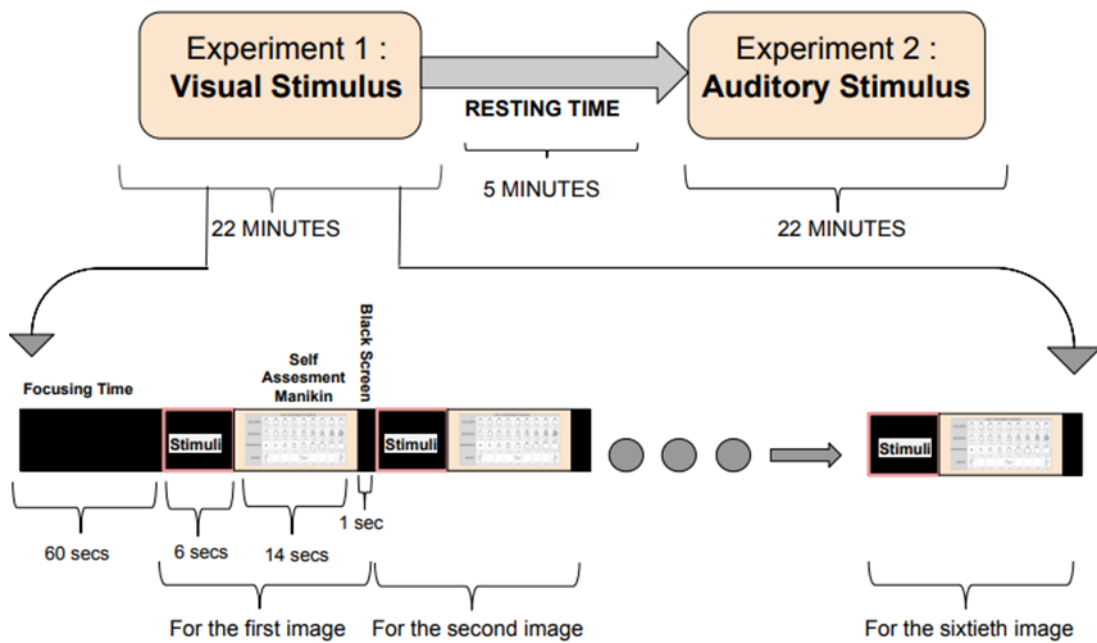Figure 3.15: Examples from used IAPS images

Figure 3.16: Demonstration plan for the visual stimulus experiment

In the second part, the selected sound from IADS and selected 12 similar sound is listened to the participants. The same plan in the visual plan applies here. Only the stimuli are different from the first experiment. The black screen appears for 6 seconds on the screen while the auditory stimuli are given. The other parts are the same. The Figure 3.17 shows the model of the demonstration plan for the auditory stimulus experiment.
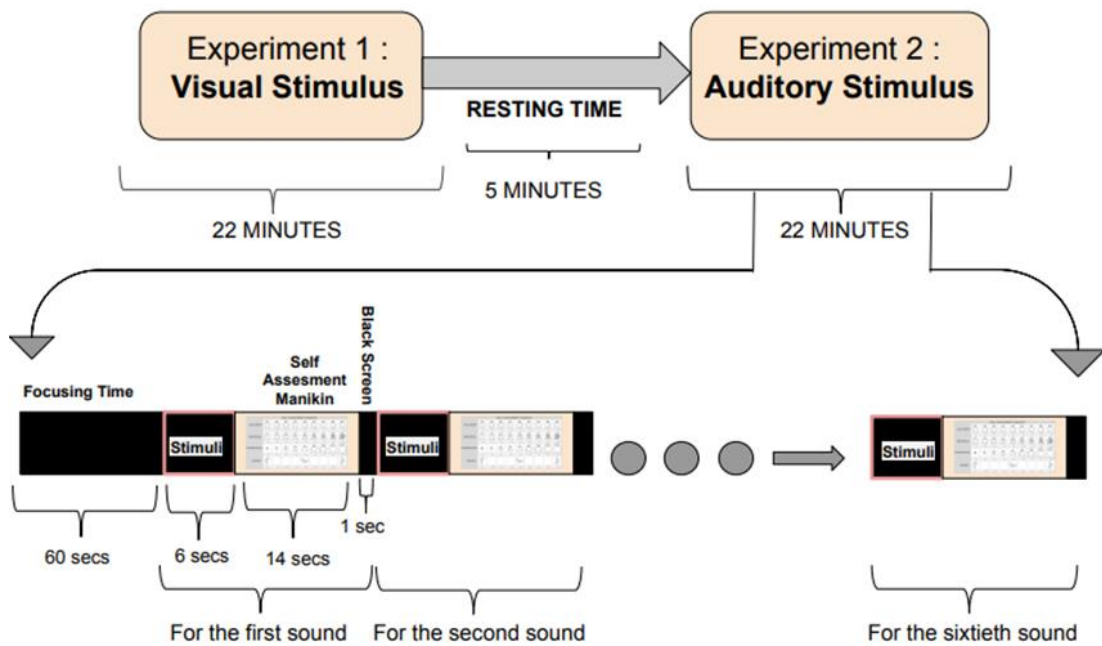
Figure 3.17: Demonstration plan for the auditory stimulus experiment

In the experiment, the volunteers are seated in a comfortable position and the pictures are viewed from the DELL branded desktop computer screen. The EEG signal data of the participants who had errors during the experiment are not used in the study.

3.5.2 Experimental Setup

This study consists of two parts including visual and auditory stimuli. The Figure 3.18 shows a photo taken from the participant during experiment.

Before starting the experiment;

1. Participants are informed about the experiment.
2. The ethical form is signed by participants.
3. The participant must remove the metal items and the phones must be switched off.
4. The participant sits in a comfortable chair and the head is measured for the cap of the brain.
5. The appropriate brain cap is selected, placed on the participant's head and the ears to be used as reference with alcohol are cleaned.
6. The gel is applied to 32 electrodes to increase conductivity. The gel is also applied to the reference electrode for the ear.
7. For the EDA recording, the index finger and middle finger of the non-dominant hand of the person are cleaned with alcohol. The electrodes are placed on these two fingers.
8. The user interface is opened and the recording room must be quiet and dimly lit.
9. When the preparatory phase is completed, the participant is informed by saying 'Experiment is starting'.

During the experiment;

1. First of all, we will ask the participant to start the experiment by counting down from 3. The EEG record and the stopwatch is started simultaneously.
2. At this stage, before the visual stimuli start, the black screen and white dot appear on the screen for people to focus for 60 seconds.
3. Approximately 10 seconds after the EEG recording starts, the EDA recording is started and the stopwatch is started before the visual stimuli start.
4. Then visual stimuli are displayed on the screen for 6 seconds. SAM survey is completed for 14 seconds on the computer by participants and after 1 second black screen are displayed, the second visual stimulus continues. The 60 visual stimuli displays. The experiment continues this way and lasts for 22 minutes. EEG signals and EDA signals are recorded.

5. After the visual experiment is completed, the participant rests for a while and the auditory experiment begins. In this part, it is an experiment that lasts 22 minutes with auditory stimuli.

After the experiment;

1. After the experiment is complete, the 'Experiment is done' warning appears on the screen.

2. The electrodes on the hand and head of the participant are removed.
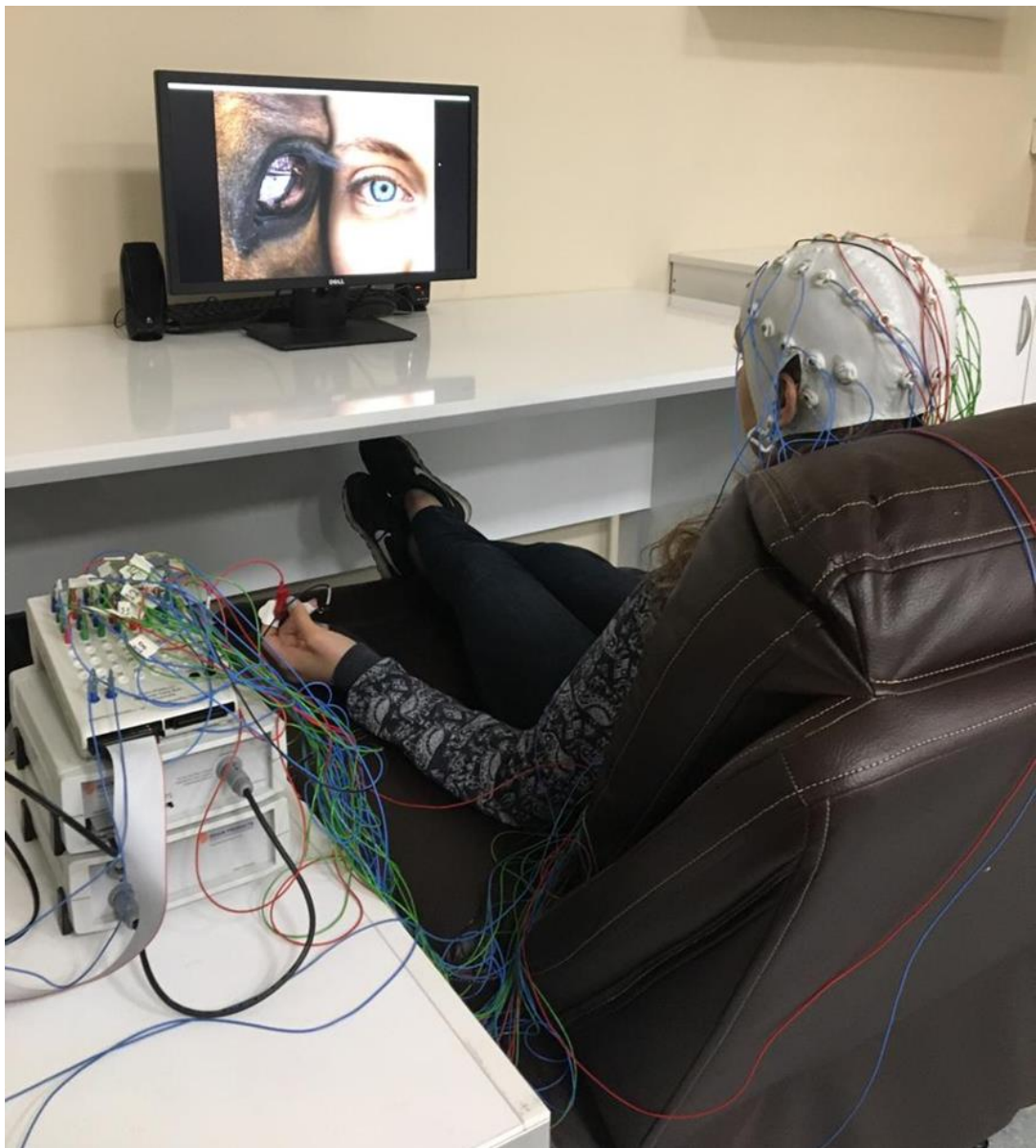
Figure 3.18: A photo taken from the participant during experiment

3.6 Data Analysis

After obtaining the EEG signals, EEG signals are recorded and these signals need to be processed. In this section, EEG signals analysis and signal processing methods are described step by step.

3.6.1 Proposed Method

Block diagram of the proposed method is shown in Figure 3.19.
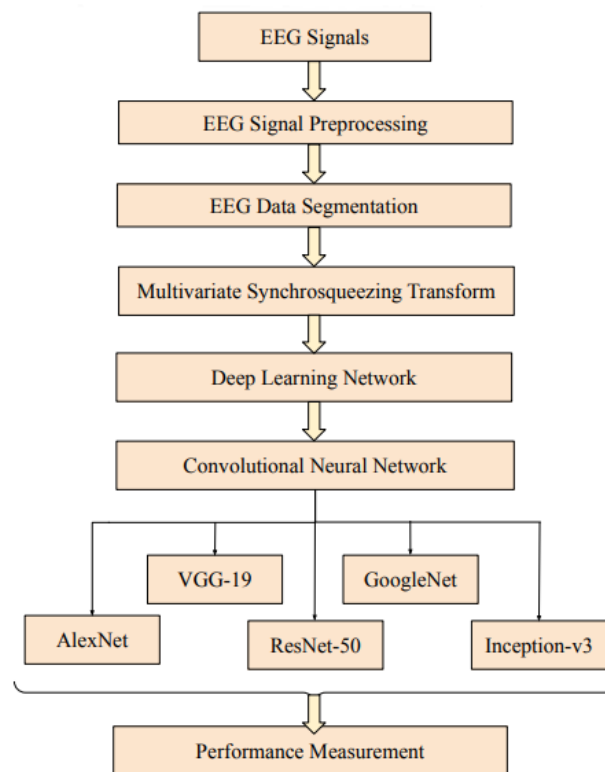
Figure 3.19: Block diagram of the proposed method

3.6.2 EEG Signal Preprocessing

Preprocessing is required after the EEG signals are obtained, because a lot of noise affects EEG signals. Blinking, heart beat, muscular contraction causes artifacts and affects the EEG spectrum. These noises can be effectively reduced by appropriate

bandpass filtering. As a filter in the recording program of the EEG signals, we have applied the 0.7 Hertz (Hz) frequency low pass filter, 100 Hz frequency high pass filter and notch filter. Notch filters are generally used in biomedical engineering applications to suppress 50 Hz noises caused by mains voltage. The low pass filter helps to prevent high frequency environmental noise from outside the recording system, allowing lower frequencies to pass. The high pass filter blocks low frequency signals and allows higher frequencies to pass.

3.6.3   EEG Data Segmentation

When we get the filtered signals, long signals are segmented to obtain the time intervals in which the stimuli are shown. We wanted to examine the emotions of people in the time intervals of stimuli. Recording time is available for each participant and recorded in a text file. So, we know the timing of every second of the experiments. Firstly, the focusing part which was 60 seconds from all filtered signal was removed. Then segmentation was performed. The remaining signals were segmented for 6 seconds, and the 14 seconds survey sections and 1 second black screen portion were removed. 60 visual stimuli were used. Each of the visual stimuli was shown to people for 6 seconds. From the 6-seconds EEG signals obtained, 1 second was subtracted from the beginning and the end, since there may be any artifact, and 4 seconds EEG signals were obtained clearly. In other words, a total of 60 pieces of 4-seconds signals were obtained from a person and 25 people have been participitaed EEG signals have been segmented into 60x25 .mat files by cutting stimulus intervals from recorded signals.

3.6.4   Time–Frequency Representations (TFRs)

It is well known that EEG signals have intrinsic nonstatinary characteristics that manifest itself as variations in spectrum by time. The time-frequency representation methods [42] promise an effective way to analyze and extract adequate information from an EEG signal. Conventional methods such as Short-time Fourier transform (STFT), wavelet transforms (WT), Choi-Williams Distribution (CWD) and Wigner Ville Distribution (WVD) methods have been widely used in analyzing of EEG. The basic drawback of these methods is producing dissatisfactory time -frequency resolution. Emperical Mode Decomposition (EMD) and Hilbert Huang Transforms

(HHT) methods depend on representing the signal in terms of finite number of intrinsic mode functions (IMF). Each IMF contains different frequency components of the original signal. In EMD method, the original signal is divided into IMFs and remainder. HHT uses the IMF information to find the instantaneous frequency information rather than constant frequency and constant amplitude.

Synchrosqueezing Transforms proposed by Daubechies has been got ahead off all these method by achieving the fine-detailed time-frequency representation.

3.6.5 Wavelet based Synchrosqueezed Transforms (WSST)

Synchrosqueezing Transformations (SST) is a reassignment technique applied to conventional time-frequency representations (TFR) of nonstationary signal. The main aims of SST are to provide instant frequency estimation and a sharper representation. The SST methods can be classified according to pre-processing method as Wavelet based Synchrosqueezed Transform (WSST) and Fourier based Synchrosqueezed Transform (FSST) [43].

3.6.6 Multivariate Synchrosqueezing Transform (MSST)

Existing SST methods is succesfull at analyzing of mono-component frequency signals whereas they fail to detect the instateneous frequencies of multi-channel signals. A. Ahrabian et al. [44] proposed the MSST [45] to overcome these issue. For a multivariate signal:

$$s(t) = \begin{bmatrix} a_1(t) \cos_1 \phi(t) \\ a_2(t) \cos_2 \phi(t) \\ \vdots \\ a_N(t) \cos_N \phi(t) \end{bmatrix} \tag{3.1}$$

with changing the instantaneous amplitude $a(t)$ and the instantaneous frequency $\phi(t)$ can be expressed as in the following form by applying Hilbert transform:

$$s_+(t) = \begin{bmatrix} a_1(t) e^{i\phi_1 t} \\ a_2(t) e^{i\phi_2 t} \\ \vdots \\ a_N(t) e^{i\phi_N t} \end{bmatrix} \tag{3.2}$$

To define multivariate single-component signals from multivariate signals, the time-frequency domain is divided into frequency band $\omega_{l,m} = [m/2^{(l+1)}, (m+1)/2^{(l+1)}]$ with 2^l equal-width frequency bands. If the corresponding frequency band level is expressed by L, the frequency index $l = 0, \dots, L$ will have value ranges $m = 0, \dots, 2^l - 1$. After finding the SST coefficients of each channel $T_n(\omega, b)$, instantaneous frequencies Ω and instantaneous amplitudes A of the multi-channel signal for each frequency band can found as given in following formulas:

$$\Omega_k^n(b) = \frac{\sum_{\omega \epsilon \omega_k} |T_n(\omega, b)|^2 \omega}{\sum_{\omega \epsilon \omega_k} |T_n(\omega, b)|^2} \tag{3.3a}$$

$$A_k^n(b) = \sqrt{\sum_{\omega \epsilon \omega_k} |T_n(\omega, b)|^2} \tag{3.3b}$$

The multi-variable instantaneous frequency is estimated by combining the common instantaneous frequencies across the N channel:

$$\Omega_k^{multi}(b) = \frac{\sum_{n=1}^{N} \left(A_k^n(b)\right)^2 \Omega_k^n(b)}{\sum_{n=1}^{N} \left(A_k^n(b)\right)^2} \tag{3.4}$$

Instantaneous amplitude will be as in Equation 3.5.

$$A_k^{multi}(b) = \sqrt{\sum_{n=1}^{N} \left(A_k^n(b)\right)^2} \tag{3.5}$$

The multivariate time-frequency coefficient will be written in as in:

$$T_k^{multi}(\omega, b) = A_k^{multi}(b)\delta\left(\omega - \Omega_k^{multi}(b)\right) \tag{3.6}$$

In our study we convert 32-Channel EEG signals to EEG images by closely following steps. An example of the MSST output of 32-Channel EEG signals in our dataset is shown in Figure 3.20.
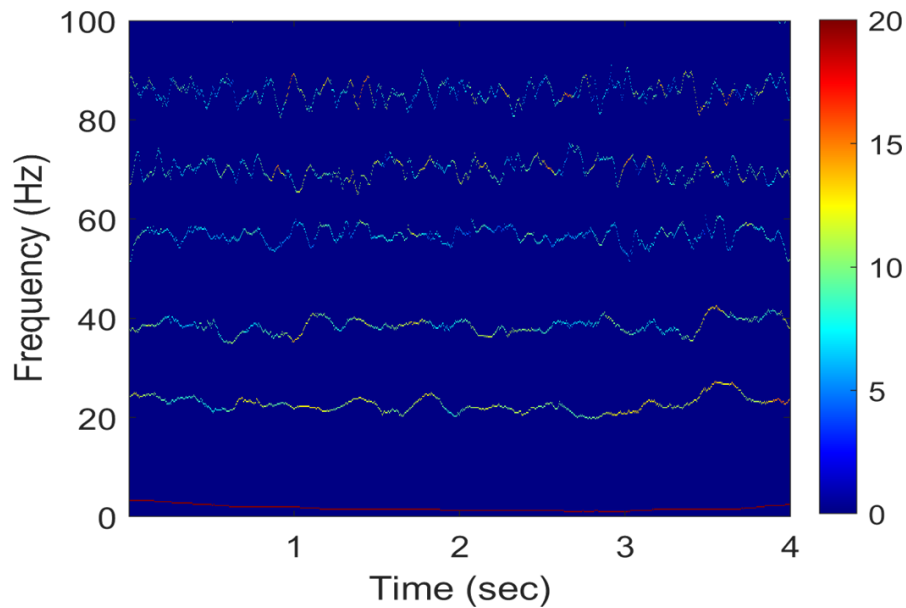
Figure 3.20: An example of the TFR of 32-Channel EEG signal in our dataset

3.6.7 Deep Neural Networks and Applications

Deep learning and machine learning are subsets of artificial intelligence that enable many new products and jobs to be created. Deep learning is a machine learning method consisting of multiple layers that predicts results with a given data set.

Deep learning is carried out with a wide variety of artificial neural networks. During the processing of a data with this model, learning is performed in all the processing layers. On the other hand, each new layer accepts the information obtained from the previous layer as output.

Artificial neural networks are made up of neurons, just like the human brain. All neurons are interconnected and affect the output [46]. A neural network has 3 layers; Input Layer, Hidden Layers and Output Layer. The Figure 3.21 shows the structure of the deep learning. Connections between neurons are associated with weight. Neurons use an activation function on the data. A large dataset is needed to train the neural network. A cost function is generated to iterate through the dataset and compare the outputs. After each iteration in the dataset, weights between neurons are adjusted using gradient descent to reduce the cost function. After each iteration in the dataset, weights between neurons are adjusted.
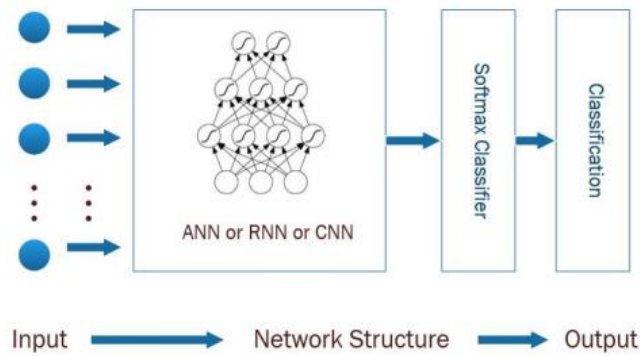
Figure 3.21: Structure of the deep learning [47]

There are many more deep machine learnings, especially LeNet, AlexNet and ZF Net. Today, deep learning architectures that emerged with artificial neural networks are used in many subjects such as image recognition and detection, signal processing, drug production, and dictionary creation. On the other hand, deep learning mechanisms are also used in fields such as defense industry, finance, architecture and smart phone applications. The success of the results reveals that this model is a much more qualified process than other technologies.

3.6.8 Convolutional Neural Network (CNN)

The algorithm, which is a deep learning computer vision model that can detect, classify and reconstruct images with high accuracy rates, is Convolutional Neural Networks (CNN) [48]. CNN renders the image with various layers. These layers are convolutional layer, non-linearity layer, pooling (downsampling) layer, and fully-connected layer. The CNN topology is shown in Figure 3.22.

Figure 3.22: An example of CNN topology [49]

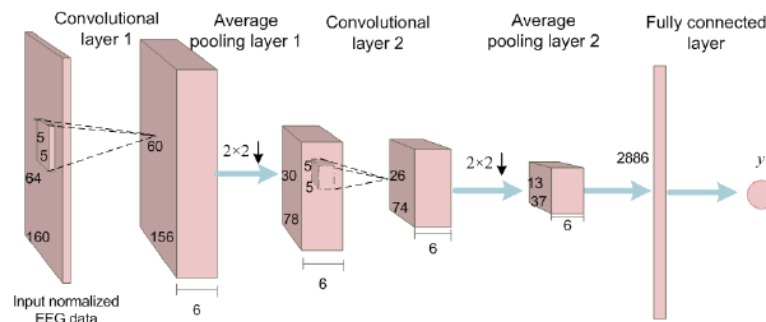Convolutional Layer: This layer handles the image in CNN algorithms. Images are matrices consisting of pixels with certain values in them. In the convolution layer, a filter smaller than the original image size hovers over the image and tries to capture certain features from these images [49], [50]. Its purpose is to detect the features of the picture. This layer applies some filters to the image. These filters can often be multidimensional. After the filtering process is finished, the output matrix is created. This matrix is often called a Feature Map. In CNN, when applying the initial filters, it is necessary to preserve as much information as possible for the other Convolutional Layers. Padding is used for this reason.

Non-linearity Layer: The activation layer uses one of the activation functions. Since the Rectified Linear Unit (ReLu) function gives the best results for the speed of Neural Network training, the ReLu function is generally used. ReLU, whose main purpose is to get rid of negative values, has a very important position in CNNs. The simple representation is as follows:
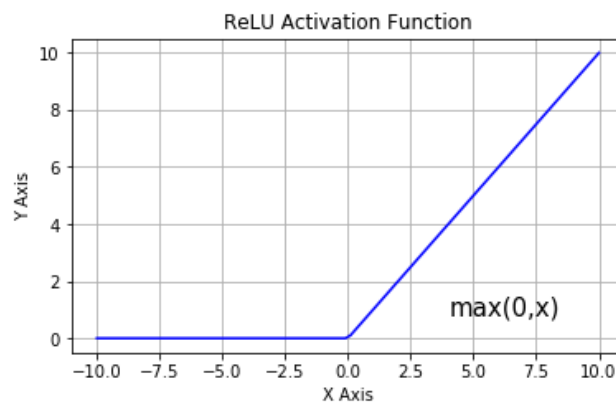
$$f(x) = max(0, x) \tag{3.7}$$

Figure 3.23: ReLu activation function

The Figure 3.23 shows ReLu activation function. Nonlinear functions such as tanh and sigmoid are used to prevent the model from learning negative values or not being able to grasp some features due to these negative values.

Pooling Layer: The pooling layer reduces dimensionality. The most commonly used pooling layer is max pooling. If you are applying max pooling, it takes the largest value in the area covered by the filter, and if you are applying average pooling, it takes the average of the values in the filter. This reduces size and retains important features.

Fully-Connected Layer: The matrix image passing through the convolution layer and the pool layer is converted to a flat vector at this stage [50]. When the input picture can be trained with neural networks, the working logic in classical neural networks is applied. At the nodes in the layers, the features are kept and the learning process begins by changing the weight and bias. The Figure 3.24 shows the example of fully-connected neural network.
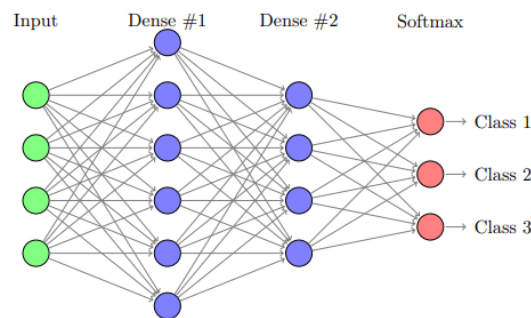
Figure 3.24: Example of fully-connected neural network [50]

We can list some of the CNN architectures as follows: LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, ZFNet. We used AlexNet, VGG-19, GoogLeNet, ResNet-50 and Inception-v3 architectures in this study.

3.6.8.1   AlexNet

AlexNet was developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton [51]. It has successive convolution and pooling layers [51]. While ReLu is used as the activation function, max-pooling is used in the pooling layers. AlexNet has eight weighted (learnable) layers [52]. Of these eight layers, the first five are convolutional layers and the last three are fully connected layers. The Figure 3.25 shows architecture of AlexNet.

57
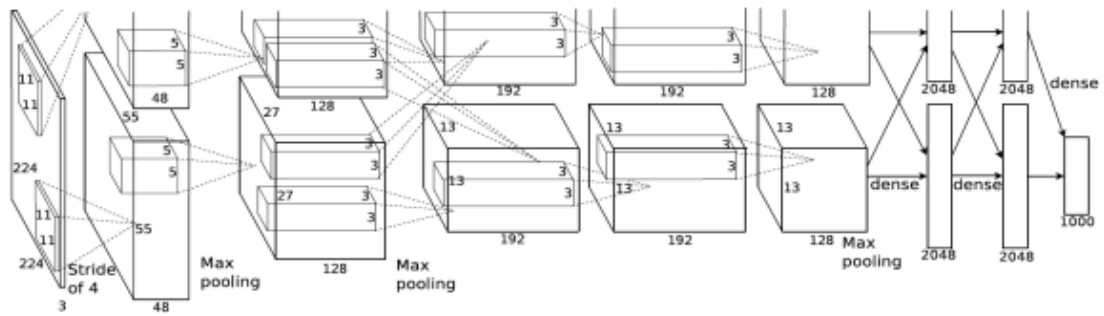
Figure 3.25: The architecture of AlexNet [51]

3.6.8.2   VGGNet-19

VGGNet [53] is one of the most popular CNN architecture, which is introduced by Simonyan and Zisserman in 2014. The authors introduced a total 6 different CNN configurations. The VGGNet-16 and VGGNet-19 are the most successful ones.

3.6.8.3   GoogLeNet

GoogLeNet appears to be faster than the VGG deep learning algorithm. GoogLeNet has 22 layers and has an error rate of 5.7%. It consists of approximately 100 layers in its architectural structure [54]. There are 144 layers in the GoogLeNet algorithm such as convolution, maxpooling, softmaxlayer, fullconnected layer, relulayer, input layer and output layer [54].

3.6.8.4   ResNet-50

It is residual neural networks. ResNet architectures were developed by the Microsoft research team to reduce the difficulty of training deep neural networks. Resnet adds shortcuts between layers to solve degradation problem of the CNN. Shortcut links do not contain extra parameters and do not cause computational complexity [55]. By using shortcut links, important information from the previous layer can be transferred to the next layers [55]. ResNet has several variants, consisting of 18, 34, 50, 101 and 152 weight layers.

3.6.8.5 Inception-v3

It is a kind of convolutional neural network model. This reduces the number of connections without reducing the efficiency of the network [56]. It consists of many convolutions and maximum pooling steps. The final stage contains a fully connected neural network [57].

In our study, we first tried 5 different deep learning networks for valence visual data: AlexNet, VGG-19, GoogLeNet, ResNet-50 and Inception-v3. The best results were obtained on the AlexNet deepnet, and therefore other classification scenarios were run on AlexNet.

3.6.9 Training Parameters of Deep Networks

In this study, Adam Optimizer [58] was used in the training phase of the proposed architecture due to efficient hyperparameter selection. Also, the batch size is fine-tuned [59] by parameter tuning. Different batch sizes 8, 16, 32, 64, 128 and 256 were tested during the training phase and optimized to batch size 16 to achieve the lowest error rate. Also, to ensure a lower error rate and avoid saturation of the model, different learning rates 0.01, 0.001, 0.0001 and 0.00001 were tested and the best value was obtained at 0.0001. At this rate, lowering the learning rate hyperparameter slightly increased the training cost, but fine-tuned it to avoid local minimums. Finally, the epochs adjust at 150 to more clearly compare and investigate the test results.

3.6.10 Performance Management

Performance evaluations are made for performance measurement of classifier according to accuracy, sensitivity, and specificity [60]. In this part, we have been simply defined and calculated the accuracy, sensitivity, and specificity [48].

Confison Matrix:

The two-dimensional confusion matrix is applied to measure system performance. A confusion matrix is a table commonly used to describe the performance of a

classification pattern on a set of test data, in which actual values are known. According to the this, 4 results can be extracted:

1. True Positives (TP): The actual label is 1 and the estimated label is 1.

2. True Negatives (TN): The actual label is 0 and the estimated label is 0.

3. False Positives (FP): The actual label is 0 and the estimated label is 1.

4. False Negatives (FN): The actual label is 1 and the estimated label is 0.

Thanks to the matrix, the estimated labels actual labels can be compared and interpreted. The Figure 3.26 shows the confusion matrix and the Figure 3.27 shows the example of obtained confusion matrixes in our study. Given in matrix, 1: TRUE and 0: FALSE, as follows:
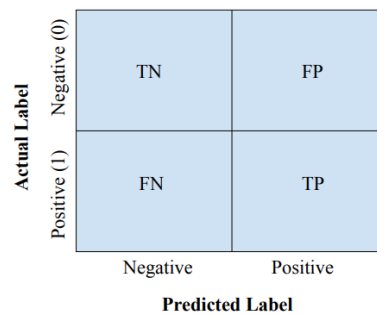

Figure 3.26: Confusion matrix
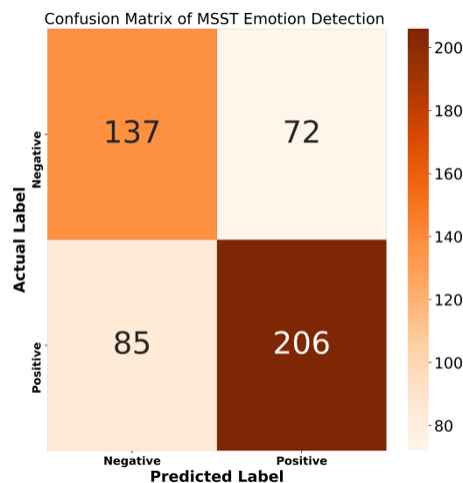
Figure 3.27: One of the obtained confusion matrixes in our study

60

Accuracy Rate (ACC): In general, it is a measure of how often the classifier guesses correctly.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \qquad (3.8)$$

Sensivity (SEN): It is a measure of how accurately all classes are estimated. Also known as Positive Predictive Value.

$$SEN = \frac{TP}{TP + FN} \qquad (3.9)$$

Specificy (SPE): is also called ratio of true negative.

$$SPE = \frac{TN}{TN + FP} \qquad (3.10)$$

Precision (PRE): It is a measure of how accurately predicted from all classes. It should be as high as possible.

$$PRE = \frac{TP}{TP + FP} \qquad (3.11)$$

Recall: It is the same as sensitivity. It only gives the proportion of correctly classified positive values.

$$Recall = \frac{TP}{TP + FN} \qquad (3.12)$$

F1: It obtains the harmonic average of the Precision and Recall values. Therefore, gives a combined idea of these two measurements. It is a better measure than accuracy. It is a measure of how well the classifier is performing.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (3.13)$$

ROC: This is a chart used to summarize the classifier's success over all possible values. ROC Curve is used to generate sensitivity, sensitivity, and specificity reports. The ROC curve is a very important performance measure for classification problems. ROC is a probability curve and the area under it, Area Under Curve (AUC), represents the degree or measure of separability. In ROC curve, FPR is the False Positive Rate on the

X axis and TPR is the True Positive Rate on the Y axis. TPR is Recall (Sensitivity) value. The Figure 3.28 shows a ROC curve obtained in this study.
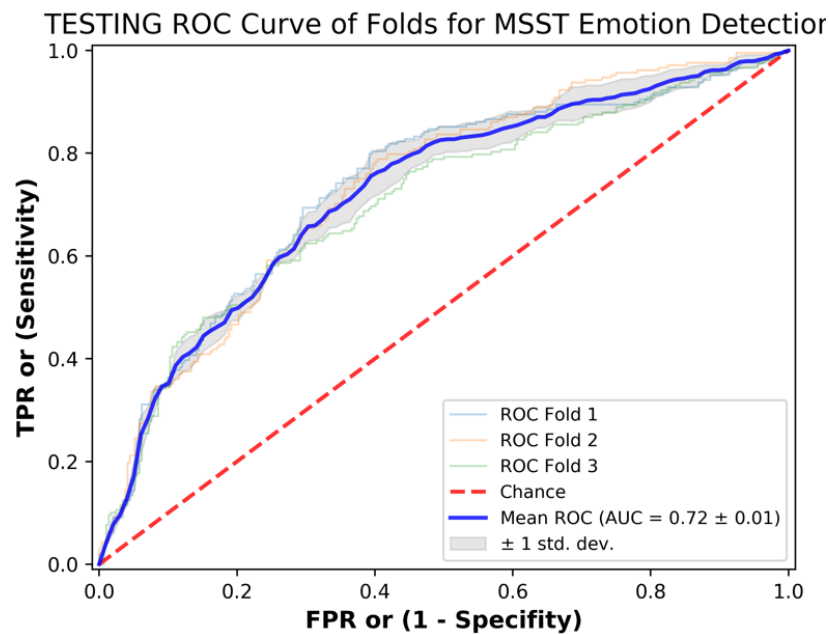
Figure 3.28: A ROC curve graph obtained in the study

As the area under the curve increases, the discrimination performance between classes increases. In order to create the best ROC curve, the best threshold value distinguishing between the classes must be found. When the best F1 value is obtained by trying different threshold values, the best ROC curve is also created. The ROC curve showed that the area under the AUC increased as the F1 value increased.

AUC (Area Under Curve) Value: The AUC, ranging between 0 and 1, tells how well the model can distinguish classes. The higher the AUC value, the better the model is at predicting zeros as zeros and ones as one. If the AUC value is 1, the model predicted 100% correctly. When we look at Figure 3.27, the AUC value is seen as 0.72, the success rate is 72% based on the AUC metric.

Mean Squared Error (MSE): The desired outcome is to create the prediction model with the smallest value MSE statistic. The prediction performance of the model that minimizes the error is considered the best. It gives a real number to compare with other model results and allows choosing the best regression model. The MSE statistic takes

62

squared errors, large deviations it gives exaggerated results. Root Mean Square Error (RMSE) is the square root of the MSE. The lower the MSE value, the better it fits the predictive data. The general MSE formula:

$$MSE = \frac{1}{2} \sum_{i=1}^{n} (Y_i - Y'_i)^2 \qquad (3.14)$$

<u>Log (Logarithmic) Loss or Cross Entropy Loss:</u> It is one of many possible loss functions and an important criterion for classification based on probability values in the estimation.

It is difficult to interpret the log loss in the observation size, but the lower the value, the higher the model success. A low value of the output indicates better model success.

<u>Cross-validation:</u> It is a method used to objectively and accurately evaluate the performance of a machine learning model on data it has not seen. Since the aim is to objectively and accurately evaluate model performance, the train-test split approach is incomplete at this point and therefore the k-Fold cross-validation method is applied. In this section, the k-cross validation method used in this study is explained.

1. The resulting data is mixed. (optional)
2. It is separated according to the selected k group.

The following steps are applied for each group:

3. The selected group is used as the validation set.
4. All other groups (k-1 groups) are used as train sets.
5. The model is built using the Train set and evaluated with the validation set.
6. The evaluation score of the model is stored in a list.
7. The statistical summary of the evaluation scores is checked.

In our study, we chose k as 3. The Figure 3.29 shows the schematic representation of 3-fold cross validation.
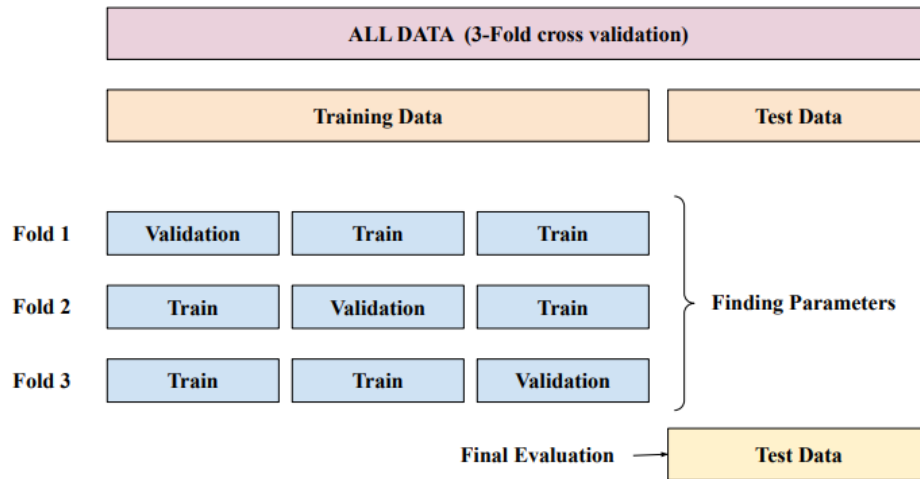
Figure 3.29: The schematic representation of 3-fold cross validation

A total of 1500 MSST images were acquired each. The discrimination rate of the training test is 1/3, that is, 1000 of them were used for data training phase and 500 for testing phase. After the test and training datasets were separated, the validation rate of the training data was chosen as 0.25, so 250 of 1000 data were used for validation. Calculations are made for precision, recall, specificity, F1 score and ROC-AUC values for visual and auditory data during the validation and testing phases.

Training and Validation Accuracy and Training and Validation Loss:

While performing the analysis, model is created. It is divided into traning, testing and validation phases. The Figure 3.30 shows the training and validation accuracy and loss graphics. Loss decreases and accuracy increases with each epoch. In this figure, while training ACC converged faster to 1, validation ACC converged more slowly and the training loss converged to 0 faster while the validation loss converged to 0 slowly.

Figure 3.30: An example of training and validation gaphics

Chapter 4

Result

In this study, emotion classification is performed with EEG signals obtained using visual and auditory stimuli. An experiment was designed using the IADS and IAPS datasets and experiment was conducted with 25 participants. and the participants' EEG signals were recorded using visual and auditory stimuli that stimulate and direct emotions. The obtained EEG signals have been preprocessed. Then segmentation has been performed according to the time intervals of the stimuli an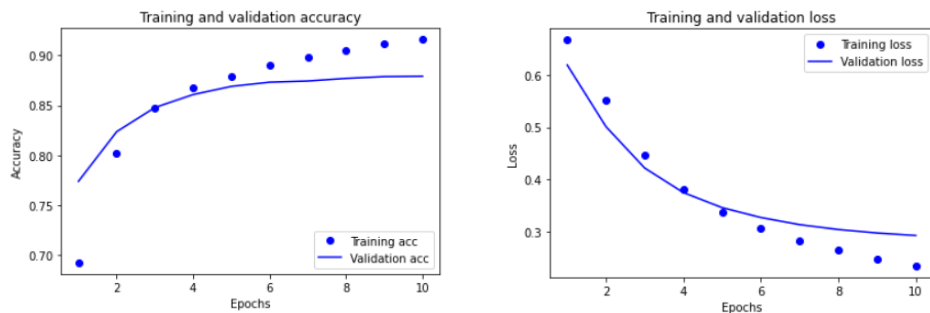d the proposed method was applied to this data. The feasibility of an enhanced TF presentation technique called Multivariate Synchronization Transform (MSST) in multi-channel electroencephalogram (EEG) signals for emotional classification was investigated. With the MSST method, 32 channels were processed simultaneously. Then the 2D images obtained were given as input to the deep learning architecture. In this study, we used AlexNet, VGG-19, GoogLeNet, ResNet-50 and Inception-v3 architectures, which are among the most known CNN architectures. We tried valence visual on 5 different deep networks, the best AlexNet came out and therefore other classification scenarios were executed over AlexNet. All results are obtained using the AlexNet deep network. Valence's performance on other networks is as follows (cross validation average of 3 folds):

- ResNet-50:   Test ACC: 61.25, F1 Score: 64.18
- Inception-v3: Test ACC: 66.12, F1 Score: 69.26
- VGGNet-19:   Test ACC: 65.72, F1 Score: 67.24
- GoogLenet: Test ACC: 60.59, F1 Score: 62.76
- AlexNet: Test ACC: 67.93, F1 Score: 73.02

All performance measurement were done from the results obtained with the AlexNet deep network. In this study, Adam Optimizer was used in the training phase of the

65

proposed architecture. Optimized to batch size 16 to achieve the lowest error rate. Also learning rates was obtained the best value at 0.0001. The epochs adjusted at 150 to more clearly discuss the results and question the validity of the model.

In this study, k-fold cross validation was applied for the proposed model. The k-fold cross-validation reorganizes the dataset to ensure that each layer correctly represents the entire dataset. In this study, the k value was chosen as 3. The REC, PRE, ACC, SPE, F1-S, ROC-AUC, MSE values were calculated for visual and auditory data during validation and testing stages. The 3 experiments as valence, arousal and dominance in visual data, and 3 experiments in visual data as valence, arousal and dominance are performed. Performance measurements were made for all 3 folds and the average of these 3 folds in valence, arousal and dominance. The Table 4.1 shows the classification results of all test phases for visual data. In the table, all measurement values values are given as %.

For Visual Arousal: By training the AlexNet architecture, for the arousal, the average ACC was obtained as 71.60%, average training ACC was obtained as 100% and validation ACC was obtained as 72%. The most successful result was obtained for the ACC as 73.60% in the 1st fold. Precision, recall, specify, F1 score and ROC-AUC values were 66.12%, 65.60%, 75.88%, 65.84% and 78.22% according to the average result of all 3 folds. The average result was 0.05% for training loss and 28.42% for mean squared entropy.

For Visual Valence: For the valence, the average ACC was obtained as 67.93%, average training ACC was obtained as 100% and validation ACC was obtained as 67.60%. The most successful result was obtained for the ACC as 68.60% in the 1st fold. Precision, recall, specify, F1 score and ROC-AUC values were 71.79%, 74.51%, 58.71%, 73.02%, and 72.47% according to the average result of all 3 folds. The average result was 0.04% for training loss and 32.06% for mean squared entropy.

For Visual Dominance: For the dominance, the average ACC was obtained as 65.40%, average training ACC was obtained as 93.78% and validation ACC was obtained as 64.93%. The most successful result was obtained for the ACC as 67.60% in the 1st fold. Precision, recall, specify, F1 score and ROC-AUC values were 70.10%, 71.21%, 57.29%, 70.44%, and 70.66% according to the average result of all

3 folds. The average result was 12.05% for training loss and 34.06% for mean squared entropy in the dominance metric. All performance calculations for all folds are shown in this table.

The Table 4.2 shows the classifcation results of all test phases for auditory data. In the table, all measurement values values are given as %.

For Auditory Arousal: By training the AlexNet architecture, for the arousal, the average ACC was obtained as 70.58%, average training ACC was obtained as 99.53% and validation ACC was obtained as 66.54%. The most successful result was obtained for the ACC as 72.67% in the 3rd fold. Precision, recall, specificy, F1 score and ROC-AUC values were 71.89%, 68.12%, 73.03%, 69.88%, and 78.11% according to the average result of all 3 folds. The average result was 2.77% for training loss and 29.42% for mean squared entropy.

For Auditory Valence: By training the AlexNet architecture, for the valence, the average ACC was obtained as 62.46%, average training ACC was obtained as 100% and validation ACC was obtained as 59.73%. The most successful result was obtained for the ACC as 65.60% in the 3rd fold. Precision, recall, specificy, F1 score and ROC-AUC values were 66.74%, 70.96%, 51.36%, 68.75%, and 66.71% according to the average result of all 3 folds. The average result was 0.04% for training loss and 37.53% for mean squared entropy.

For Auditory Dominance: By training the AlexNet architecture, for the dominance, the average ACC was obtained as 72.60%, average training ACC was obtained as 100% and validation ACC was obtained as 69.86%. The most successful result was obtained for the ACC as 73.60% in the 2nd fold. Precision, recall, specificy, F1 score and ROC-AUC values were 76.42%, 72.61%, 72.59, 74.44%, and 79.28% according to the average result of all 3 folds. The average result was 0.05% for training loss and 27.40% for mean squared entropy.

When we compare Table 4.1 with Table 4.2, the best average ACC value was obtained as 71.60% for arousal in visual data, 67.93% for valence in visual data, and 72.40% for dominance in auditory data. The best ROC-AUC value was obtained as 78.22% for arousal in visual data, 72.47% for valence in visual data and 79.28% for dominance in auditory data.

Table 4.1: Performance evaluation results of trained models for Visual data

		Folds	ACC	Training ACC	Validation ACC	PRE	REC	SPE	F1-S	AUC	Loss	MSE	Training time
VISUAL	Arousal	Fold 1	73.60	100.0	75.20	69.39	65.38	79.45	67.33	78.28	0.02	26.40	5309.88
		Fold 2	69.20	100.0	70.80	62.62	64.42	72.60	63.51	77.90	0.02	30.80	5252.52
		Fold 3	72.00	100.0	70.00	66.35	66.99	75.60	66.67	78.47	0.10	28.00	5161.73
		Average	**71.60**	**100.0**	**72.00**	**66.12**	**65.60**	**75.88**	**65.84**	**78.22**	**0.05**	**28.40**	**5241.38**
	Valence	Fold 1	68.60	100.0	71.60	74.10	70.79	65.55	72.41	73.36	0.03	31.40	5675.13
		Fold 2	67.60	100.0	62.80	71.81	73.29	59.62	72.54	73.53	0.03	32.40	5747.69
		Fold 3	67.60	100.0	68.40	69.46	79.45	50.96	74.12	70.52	0.06	32.40	5308.34
		Average	**67.93**	**100.0**	**67.60**	**71.79**	**74.51**	**58.71**	**73.02**	**72.47**	**0.04**	**32.06**	**5877.05**
	Dominance	Fold 1	67.60	94.00	63.60	68.80	81.10	48.80	74.45	72.14	13.50	32.40	5236.47
		Fold 2	64.00	94.27	69.60	70.29	66.44	60.58	68.31	68.80	11.07	36.00	5277.15
		Fold 3	64.60	93.07	61.60	71.22	66.10	62.50	68.56	71.06	11.59	35.40	5194.22
		Average	**65.40**	**93.78**	**64.93**	**70.10**	**71.21**	**57.29**	**70.44**	**70.66**	**12.05**	**34.60**	**5235.94**

Table 4.2: Performance evaluation results of trained models for Auditory data

	Folds	ACC	Training ACC	Validation ACC	PRE	REC	SPE	F1-S	AUC	Loss	MSE	Training time
AUDITORY												
Arousal	Fold 1	71.57	100.0	68.34	73.22	67.83	75.29	70.42	78.62	0.02	28.43	5445.94
	Fold 2	67.50	100.0	68.34	66.91	69.50	65.50	68.18	72.82	0.44	32.50	5302.01
	Fold 3	72.67	98.58	62.93	75.55	67.05	78.29	71.05	77.89	7.87	27.33	5363.74
	Average	**70.58**	**99.53**	**66.54**	**71.89**	**68.12**	**73.03**	**69.88**	**78.11**	**2.77**	**29.42**	**5370.56**
Valence	Fold 1	60.40	100.0	60.80	65.25	68.38	49.28	66.78	67.49	0.03	39.60	4995.97
	Fold 2	61.40	100.0	58.80	66.78	67.47	52.88	67.12	67.28	0.03	38.60	5017.25
	Fold 3	65.60	100.0	59.60	68.18	77.05	49.52	72.35	65.37	0.06	34.40	5137.47
	Average	**62.46**	**100.0**	**59.73**	**66.74**	**70.96**	**51.36**	**68.75**	**66.71**	**0.04**	**37.53**	**5050.23**
Dominance	Fold 1	72.80	100.0	63.60	77.69	70.91	75.11	74.14	81.18	0.04	27.20	5259.14
	Fold 2	73.60	100.0	73.20	76.19	75.64	71.11	75.91	79.43	0.06	26.40	5131.72
	Fold 3	71.40	100.0	72.80	75.38	71.27	71.56	73.27	77.23	0.04	28.60	5372.45
	Average	**72.60**	**100.0**	**69.86**	**76.42**	**72.61**	**72.59**	**74.44**	**79.28**	**0.05**	**27.40**	**5254.43**

The Figure 4.1 shows the training accuracy and validation accuracy graphics for training during fine-tuning the model in visual data for arousal, valence and dominance. When we examine this figure, it is seen that the training ACC for arousal and valance reaches 100% at the highest level in Fold-1, Fold-2 and Fold-3. They have reached the desired value. In dominance, on the other hand, it is seen that the training ACC value approaches 100% but not 100% in Fold-1, Fold-2 and Fold-3. It is seen that validation ACC reaches the highest value in arousal Fold-1 as 75.42%, in valence Fold-1 as 71.60% and in dominance Fold-2 as 69.60%.

The Figure 4.2 shows the training accuracy and validation accuracy graphics for training during fine-tuning the model in auditory data for arousal, valence and dominance. When we examine this figure, it is seen that the training ACC for arousal, valance and dominance reaches 100% at the highest level in Fold-1, Fold-2 and Fold-3. It has reached the desired value. It is seen that validation ACC reaches the highest value in arousal Fold-1 and Fold-2 as 68.34%, in valence Fold-1 as 60.80% and in dominance Fold-2 as 73.20%.

The confusion matrices obtained in Fold1, Fold2, Fold3 in testing phase and validation phase in arousal, in valence and in dominance as visual and auditory are given in Figure 4.3, Figure 4.4 and Figure 4.5, respectively. When we examine Figure 4.3, in the training phase of the visual arousal, confusion matrices of three folds appear. To prove the robustness of the model, the model is trained three times and different confusion matrices are obtained for each. Also, the validation confusion matrix obtained in one-fold of the validation phase of the arousal of the image appears here. Validation is the part that separates from training data. This is a confusion matrix for the moment of validation. In the visual part, confusion matrices of Fold-1, Fold-2, Fold-3 and validation phase are obtained for arousal, valence and dominance. The same procedures apply to the auditory part.
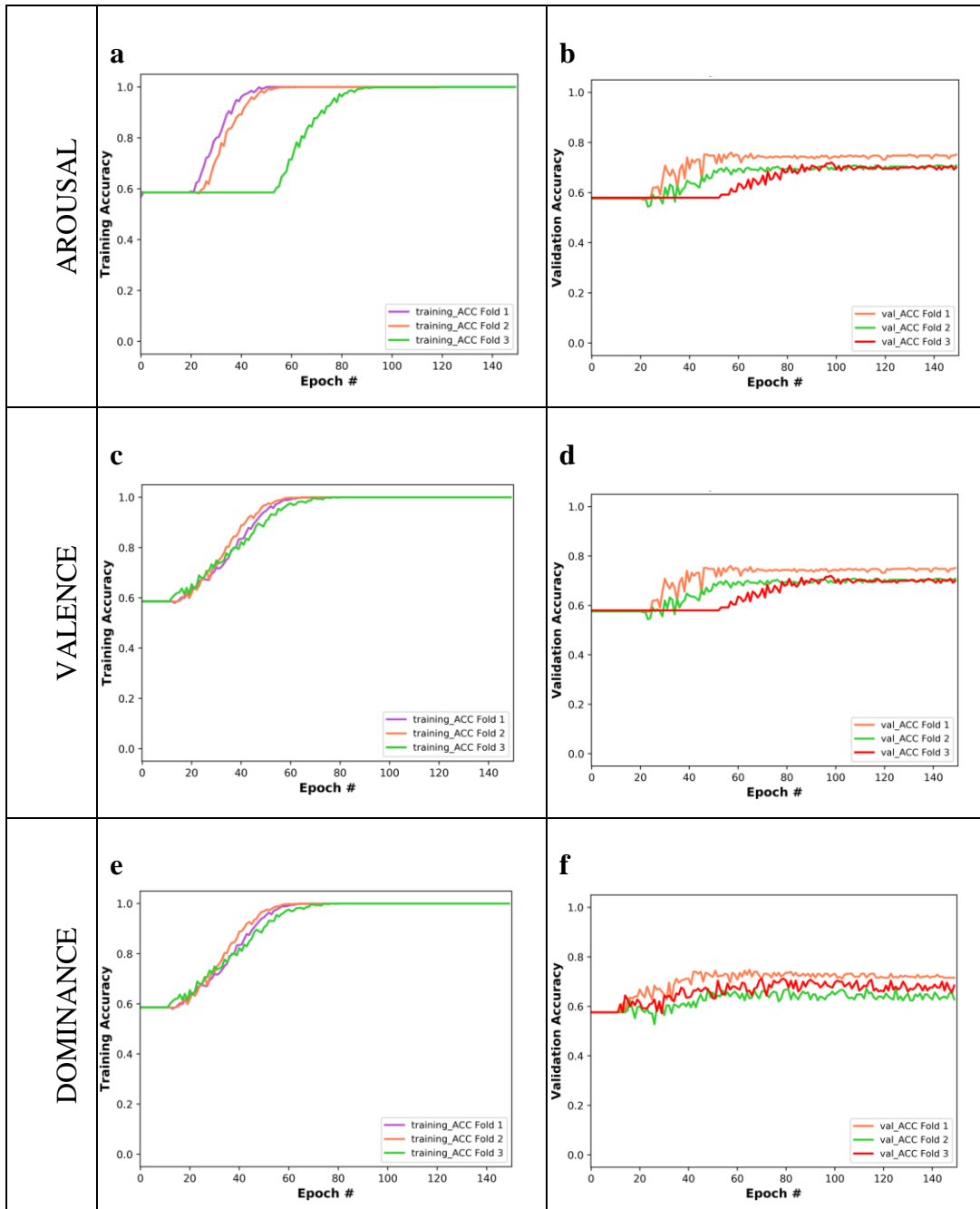
For Visual;

Figure 4.1: The training accuracy vs epochs for training during fine-tuning the model in visual data for **a** arousal **c** valence **e** dominance, and the validation accuracy vs epochs in visual data for **b** arousal **d** valence **f** dominance
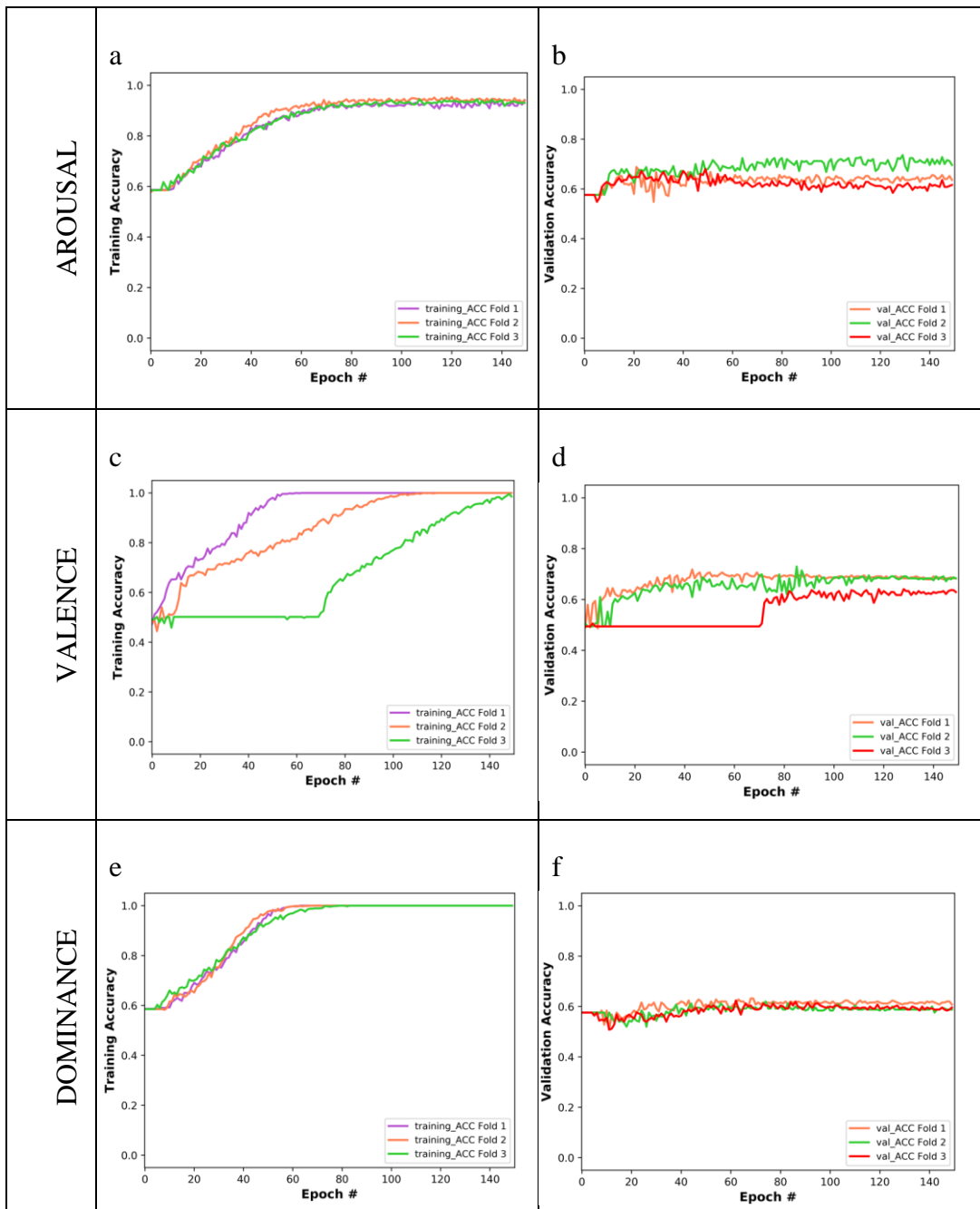
For Auditory;

Figure 4.2: The training accuracy vs epochs for training during fine-tuning the model in auditory data for **a** arousal **c** valence **e** dominance, and the validation accuracy vs epochs in visual data for **b** arousal **d** valence **f** dominance

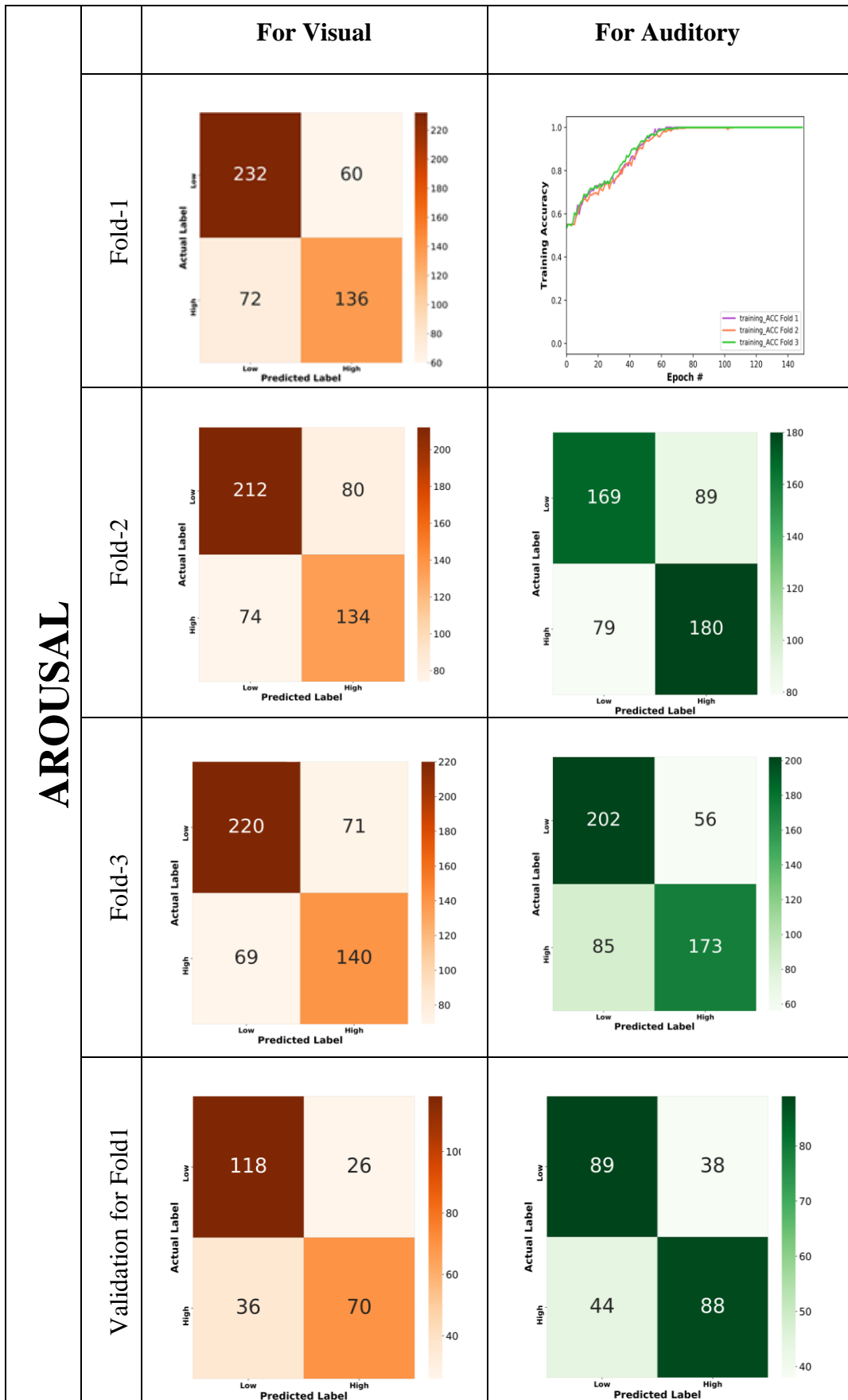|	For Visual	For Auditory
Fold-1	Low: 232 / 60, High: 72 / 136	Training Accuracy vs Epoch # (training_ACC Fold 1, Fold 2, Fold 3)
Fold-2	Low: 212 / 80, High: 74 / 134	Low: 169 / 89, High: 79 / 180
Fold-3	Low: 220 / 71, High: 69 / 140	Low: 202 / 56, High: 85 / 173
Validation for Fold1	Low: 118 / 26, High: 36 / 70	Low: 89 / 38, High: 44 / 88

AROUSAL

Figure 4.3: The confusion matrices obtained in Fold1, Fold2, Fold3 in testing phase and validation phase in arousal as visual and auditory
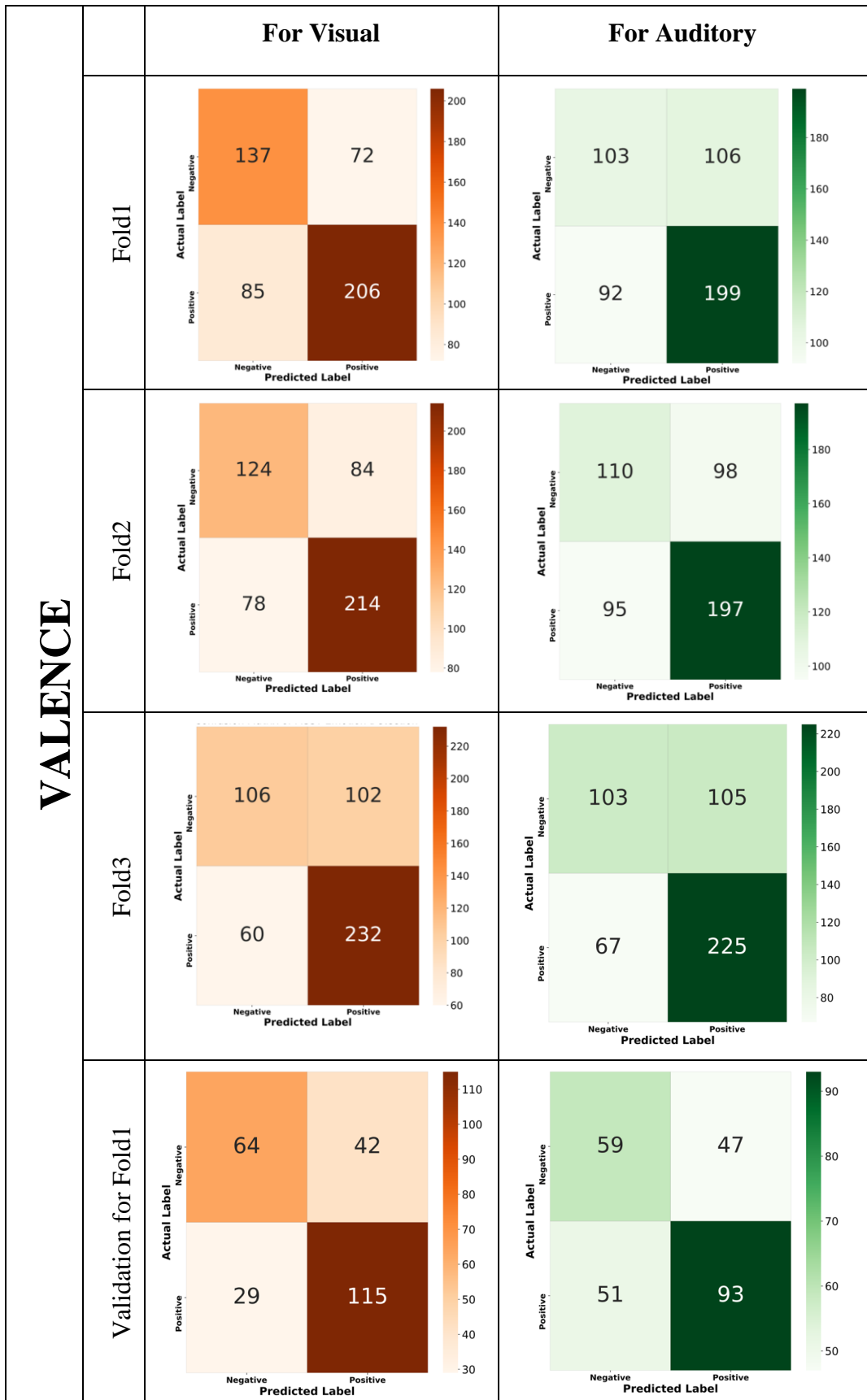
Figure 4.4: The confusion matrices obtained in Fold1, Fold2, Fold3 in testing phase and validation phase in valence as visual and auditory
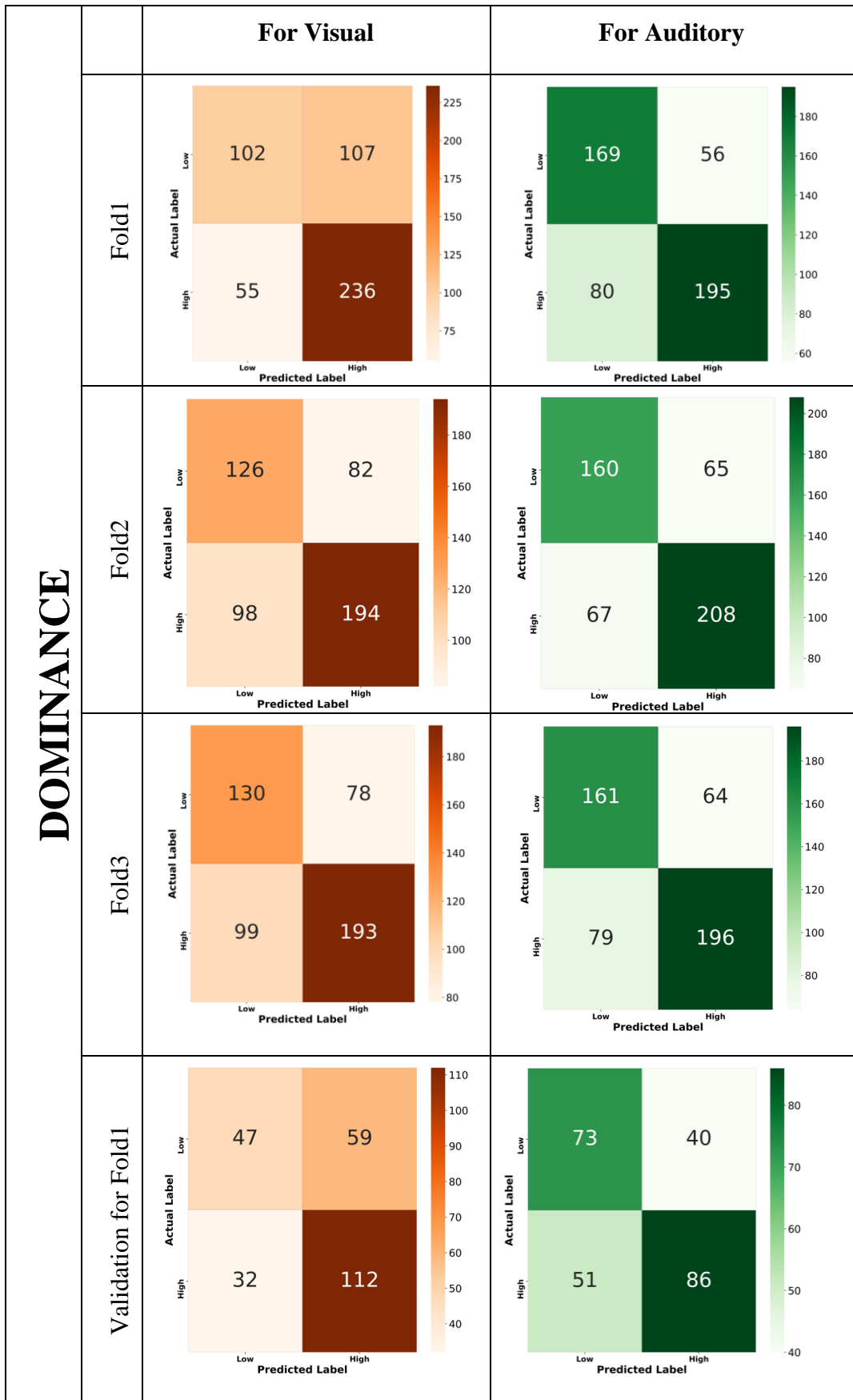
Figure 4.5: The confusion matrices obtained in Fold1, Fold2, Fold3 in testing phase and validation phase in dominance as visual and auditory
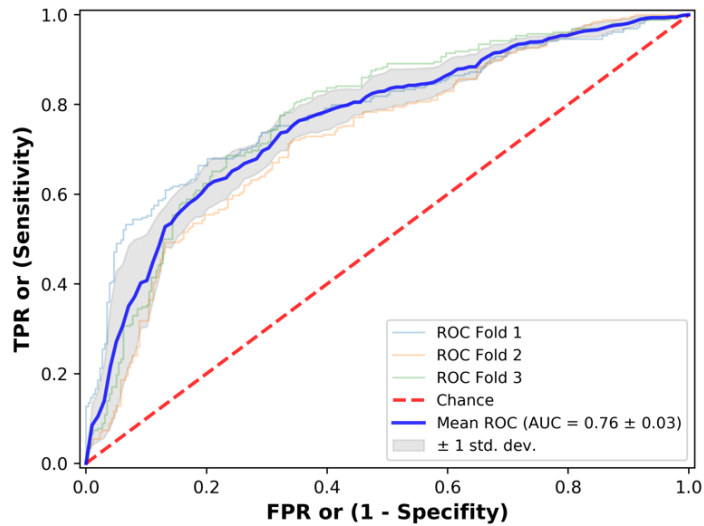
Figure 4.6: Testing ROC curve of folds for MSST emotion detection Arousal in auditory
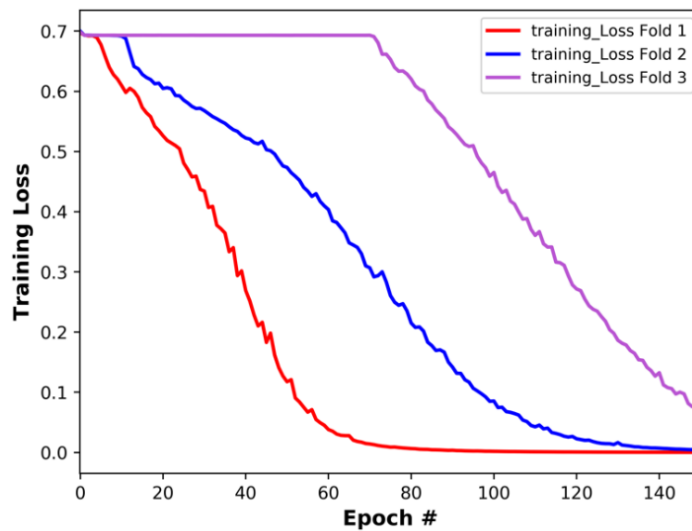
Figure 4.7: Model training Loss of MSST emotion detection Arousal in auditory

The Figure 4.6 shows the testing ROC curve of used all folds in this study. ROC curve is used summarize the classifier's performance over all possible values. It generates sensitivity, sensitivity, and specificity. As the area under the curve increases, the discrimination performance between classes increases. In this figure, Fold-1 and Fold-3 converged rapidly to 1, while Fold-2 converged more slowly and later converged at 1. However, each of the 3 folds reached 1 TPR value after a certain time. This is a desired result.

The Figure 4.7 shows the model training loss curve for folds. The system performs error control for its predictions and tries to minimize the error continuously. For this, it calculates the error and tries to converge to the expected result, that is, to reduce the error. While Fold-1 converged to 0 very well, Fold-2 converged towards the 120th epoch and Fold-3 converged towards the 140th epoch, although slightly delayed.

Chapter 5

Conclusion

Neurological problems can be detected with signals from the EEG device and data about psychiatric diseases can be obtained. EEG is used in neurology, neurosurgery, pediatrics, anesthesia and psychiatry. It is also used to diagnose epilepsy, to diagnose tumor, to investigate behavioral disorders, to analyze mood disorders and to help diagnose and treat such diseases. Emotion classification is very important in detecting various diseases. For this, we have proposed a method that has not been used so far. When we look at the results, it is seen that the method we propose is a robust method that can be an alternative to existing methods.

Emotion recognition has been performed using visual and auditory stimulus-based EEG signals with 25 participants using 32 EEG channels. In this thesis, it is aimed to classify the emotions based on visual and auditory stimulus with EEG signals. A novel and effective approach is proposed to classify the emotions using Multivariate SynchroSqueezing Transform. The results have been analyzed and compared. The results showed that the proposed method is successful in classifying emotions and can be used as a reference for further studies. The proposed method reduced the computational cost, and significantly improved the classification performance.

We have achieved a promising result from our study, various methods can be used to achieve better results with improvements. In the study, it is presented that system performance will increase with the supplementation of using MSST. It is planned to work on improving system performance in the future and we are planning to classify more complex emotion models in future studies.

References

[1]   Q. Gao, C. han Wang, Z. Wang, X. lin Song, E. zeng Dong, and Y. Song, "EEG based emotion recognition using fusion feature extraction method," Multimed. Tools Appl., vol. 79, no. 37–38, pp. 27057–27074, Oct. 2020, doi: 10.1007/S11042-020-09354-Y.

[2]   M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel EEG signals using K-nearest neighbor classification," in Technology and Health Care, Jul. 2018, vol. 26, no. S1, pp. S509–S519, doi: 10.3233/THC-174836.

[3]   L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: Features and classification models," Digit. Signal Process. A Rev. J., vol. 22, no. 6, pp. 1154–1160, 2012, doi: 10.1016/j.dsp.2012.05.007.

[4]   A. Azcarate, F. Hageloh, K. Van De Sande, and R. Valenti, "Automatic facial emotion recognition," Univ. van Amsterdam, no. June, 2005, Accessed: Sep. 18, 2021. [Online]. Available: https://staff.fnwi.uva.nl/r.valenti/projects/mmis/Automatic Facial Emotion Recognition.pdf.

[5]   M. K. Kim, M. Kim, E. Oh, and S. P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," Computational and Mathematical Methods in Medicine, vol. 2013. 2013, doi: 10.1155/2013/573734.

[6]   X. W. Wang, D. Nie, and B. L. Lu, "Emotional state classification from EEG data using machine learning approach," Neurocomputing, vol. 129, pp. 94–106, 2014, doi: 10.1016/j.neucom.2013.06.046.

[7]   R. Ranjan, R. Arya, S. L. Fernandes, E. Sravya, and V. Jain, "A fuzzy neural network approach for automatic K-complex detection in sleep EEG signal," Pattern Recognit. Lett., vol. 115, pp. 74–83, Nov. 2018, doi:

10.1016/J.PATREC.2018.01.001.

[8]   D. S. Naser and G. Saha, "Influence of music liking on EEG based emotion recognition," Biomed. Signal Process. Control, vol. 64, 2021, doi: 10.1016/j.bspc.2020.102251.

[9]   B. Xing et al., "Exploiting EEG Signals and Audiovisual Feature Fusion for Video Emotion Recognition," IEEE Access, vol. 7, pp. 59844–59861, 2019, doi: 10.1109/ACCESS.2019.2914872.

[10]   A. Bhardwaj, A. Gupta, P. Jain, A. Rani, and J. Yadav, "Classification of human emotions from EEG signals using SVM and LDA Classifiers," 2nd Int. Conf. Signal Process. Integr. Networks, SPIN 2015, pp. 180–185, Apr. 2015, doi: 10.1109/SPIN.2015.7095376.

[11]   Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," Appl. Soft Comput., vol. 100, Mar. 2021, doi: 10.1016/J.ASOC.2020.106954.

[12]   J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Classification of burst and suppression in the neonatal electroencephalogram," J. Neural Eng., vol. 5, no. 4, pp. 402–410, 2008, doi: 10.1088/1741-2560/5/4/005.

[13]   "Brain Facts: The Four Lobes." https://brainframe-kids.com/brain/facts-lobes.htm (accessed Oct. 30, 2021).

[14]   P. J. Morgane, J. R. Galler, and D. J. Mokler, "A review of systems and networks of the limbic forebrain/limbic midbrain," Prog. Neurobiol., vol. 75, no. 2, pp. 143–160, 2005, doi: 10.1016/j.pneurobio.2005.01.001.

[15]   G. J. Mogenson, D. L. Jones, and C. Y. Yim, "From motivation to action: Functional interface between the limbic system and the motor system," Prog. Neurobiol., vol. 14, no. 2–3, pp. 69–97, 1980, doi: 10.1016/0301-0082(80)90018-0.

[16]   P. Adjamian, D. A. Hall, A. R. Palmer, T. W. Allan, and D. R. M. Langers,

"Neuroanatomical abnormalities in chronic tinnitus in the human brain," Neuroscience and Biobehavioral Reviews, vol. 45. Elsevier Ltd, pp. 119–133, 2014, doi: 10.1016/j.neubiorev.2014.05.013.

[17]   N. Zink, M. Mückschel, and C. Beste, "Resting-state EEG Dynamics Reveals Differences in Network Organization and its Fluctuation between Frequency Bands," Neuroscience, vol. 453, pp. 43–56, 2021, doi: 10.1016/j.neuroscience.2020.11.037.

[18]   M. Teplan, "Fundamentals of EEG measurement," Meas. Sci. Rev., vol. 2, no. 2, pp. 1–11, 2002.

[19]   G. Brihadiswaran, D. Haputhanthri, S. Gunathilaka, D. Meedeniya, and S. Jayarathna, "EEG-based processing and classification methodologies for autism spectrum disorder: A review," J. Comput. Sci., vol. 15, no. 8, pp. 1161–1183, 2019, doi: 10.3844/jcssp.2019.1161.1183.

[20]   L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," Annu. Rev. Psychol., vol. 58, pp. 373–403, 2007, doi: 10.1146/annurev.psych.58.110405.085709.

[21]   P. Ekman, "Universals and Cultural Differences in Facial Expressions of Emotion BT - Nebraska Symposium on Motivation," Nebraska Symposium on Motivation, vol. 19. pp. 207–282, 1972, [Online]. Available: papers3://publication/uuid/FDC5E29A-0E28-4DDF-B1A4-F53FEE0B4F70.

[22]   J. A. Russell, "A circumplex model of affect," J. Pers. Soc. Psychol., vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[23]   B. H. Kim and S. Jo, "An Affective Situation Labeling System from Psychological Behaviors in Emotion Recognition," pp. 1–12, 2019, [Online]. Available: http://arxiv.org/abs/1911.01158.

[24]   Y. Liu and O. Sourina, "EEG-based subject-dependent emotion recognition algorithm using fractal dimension," Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern., vol. 2014-Janua, no. January, pp. 3166–3171, 2014, doi: 10.1109/smc.2014.6974415.

[25]   R. Plutchik and R. Plutchik, "Emotions and life : perspectives from psychology, biology, and evolution," p. 381, 2003.

[26]   S. Sheykhivand, Z. Mousavi, T. Y. Rezaii, and A. Farzamnia, "Recognizing Emotions Evoked by Music Using CNN-LSTM Networks on EEG Signals," IEEE Access, vol. 8, pp. 139332–139345, 2020, doi: 10.1109/ACCESS.2020.3011882.

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

[28]   I. S. Ahmad et al., "Deep Learning Based on CNN for Emotion Recognition Using EEG Signal," WSEAS Trans. SIGNAL Process., vol. 17, pp. 28–40, Apr. 2021, doi: 10.37394/232014.2021.17.4.

[29]   R. H. D. Haqque, "Emotion Recognition of EEG Signals Using Wavelet Filter and Convolutional Neural Networks Emotion Recognition of EEG Signals Using Wavelet Filter and Convolutional Neural Networks," EasyChair Prepr., p. 6623, 2021.

[30]   N. E. Elsayed, A. S. Tolba, M. Z. Rashad, T. Belal, and S. Sarhan, "A Deep Learning Approach for Brain Computer Interaction-Motor Execution EEG Signal Classification," IEEE Access, vol. 9, pp. 101513–101529, 2021, doi: 10.1109/ACCESS.2021.3097797.

[31]   J. Chen, D. Jiang, Y. Zhang, and P. Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset," Comput. Commun., vol. 154, pp. 58–65, Mar. 2020, doi: 10.1016/J.COMCOM.2020.02.051.

[32]   M. Miao, W. Hu, and W. Zhang, "A spatial-frequency-temporal 3D convolutional neural network for motor imagery EEG signal classification," Signal, Image Video Process., 2021, doi: 10.1007/s11760-021-01924-3.

[33]   V. M. Joshi and R. B. Ghongade, "EEG based emotion detection using fourth

order spectral moment and deep learning," Biomed. Signal Process. Control, vol. 68, no. February, p. 102755, 2021, doi: 10.1016/j.bspc.2021.102755.

[34]   K. Anjana, M. Ganesan, and R. Lavanya, "Emotional Classification of EEG Signal using Image Encoding and Deep Learning," Proc. 2021 IEEE 7th Int. Conf. Bio Signals, Images Instrumentation, ICBSII 2021, pp. 3–7, 2021, doi: 10.1109/ICBSII51839.2021.9445187.

[35]   A. Topic and M. Russo, "Emotion recognition based on EEG feature maps through deep learning network," Eng. Sci. Technol. an Int. J., no. xxxx, 2021, doi: 10.1016/j.jestch.2021.03.012.

[36]   M. R. Islam et al., "EEG Channel Correlation Based Model for Emotion Recognition," Comput. Biol. Med., vol. 136, no. August, p. 104757, 2021, doi: 10.1016/j.compbiomed.2021.104757.

[37]   P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Technical manual and affective ratings," NIMH Cent. Study Emot. Atten., pp. 39–58, 1997.

[38]   A. P. Soares, A. P. Pinheiro, A. Costa, C. S. Frade, M. Comesaña, and R. Pureza, "Affective auditory stimuli: Adaptation of the International Affective Digitized Sounds (IADS-2) for European Portuguese," Behav. Res. Methods, vol. 45, no. 4, pp. 1168–1181, 2013, doi: 10.3758/s13428-012-0310-1.

[39]   D. Xue, 10. Graphical user interface design using MATLAB. 2020.

[40]   M. Lesan, N. Ayub, and Q. U. Islam, "Emotional Evaluation of Homelike Residence Halls Using Self-Assessment Manikins Emotional Evaluation of Homelike Residence Halls Using Self-Assessment Manikins," no. July, 2020, doi: 10.22068/ijaup.30.1.20.

[41]   T. Ergin, M. A. Ozdemir, and A. Akan, "Emotion Recognition with Multi-Channel EEG Signals Using Visual Stimulus," in 2019 Medical Technologies Congress (TIPTEKNO), Oct. 2019, pp. 1–4, doi: 10.1109/TIPTEKNO.2019.8895242.

[42] M. A. Ozdemir, O. K. Cura, and A. Akan, "Epileptic EEG Classification by Using Time-Frequency Images for Deep Learning," Int. J. Neural Syst., vol. 31, no. 08, p. 2150026, Aug. 2021, doi: 10.1142/S012906572150026X.

[43] D. Degirmenci, M. Yalcin, M. A. Ozdemir, and A. Akan, "Synchrosqueezing Transform in Biomedical Applications: A mini review," in 2020 Medical Technologies Congress (TIPTEKNO), Nov. 2020, no. 3, pp. 1–5, doi: 10.1109/TIPTEKNO50054.2020.9299225.

[44] A. Ahrabian, D. Looney, L. Stanković, and D. P. Mandic, "Synchrosqueezing-based time-frequency analysis of multivariate data," Signal Processing, vol. 106, pp. 331–341, 2015, doi: 10.1016/j.sigpro.2014.08.010.

[45] N. Salankar, P. Mishra, and L. Garg, "Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot," Biomed. Signal Process. Control, vol. 65, no. February 2020, p. 102389, 2021, doi: 10.1016/j.bspc.2020.102389.

[46] L. Deng and D. Yu, "Deep learning: Methods and applications," Found. Trends Signal Process., vol. 7, no. 3–4, pp. 197–387, 2013, doi: 10.1561/2000000039.

[47] L. Chu, R. Qiu, H. Liu, Z. Ling, T. Zhang, and J. Wang, "Individual Recognition in Schizophrenia using Deep Learning Methods with Random Forest and Voting Classifiers: Insights from Resting State EEG Streams," no. September, 2017, [Online]. Available: http://arxiv.org/abs/1707.03467.

[48] M. A. Ozdemir, M. Degirmenci, E. Izci, and A. Akan, "EEG-based emotion recognition with deep convolutional neural networks," Biomed. Tech., vol. 66, no. 1, pp. 43–57, 2021, doi: 10.1515/bmt-2019-0306.

[49] L. Ma, J. W. Minett, T. Blu, and W. S. Y. Wang, "Resting State EEG-based biometrics for individual identification using convolutional neural networks," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2015-Novem, pp. 2848–2851, 2015, doi: 10.1109/EMBC.2015.7318985.

[50] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," Remote Sens., vol.

11, no. 5, pp. 1–25, 2019, doi: 10.3390/rs11050523.

[51] B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2012.

[52] N. Kumari, S. Anwar, and V. Bhattacharjee, "Convolutional Neural Network-Based Visually Evoked EEG Classification Model on MindBigData," in Proceedings of Research and Applications in Artificial Intelligence, Springer, Singapore, 2021, pp. 233–241.

[53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.

[54] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[56] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, "Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification," Front. Neurol., vol. 11, no. May, pp. 1–11, 2020, doi: 10.3389/fneur.2020.00375.

[57] A. Shalbaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," Phys. Eng. Sci. Med., vol. 43, no. 4, pp. 1229–1239, 2020, doi: 10.1007/s13246-020-00925-9.

[58] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15, 2015.

[59] M. Degirmenci, M. A. Ozdemir, E. Izci, and A. Akan, "Arrhythmic Heartbeat Classification Using 2D Convolutional Neural Networks," Irbm, vol. 1, pp. 1–12, 2021, doi: 10.1016/j.irbm.2021.04.002.

[60] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity.," Emerg. (Tehran, Iran), vol. 3, no. 2, pp. 48–9, 2015, doi: 10.22037/emergency.v3i2.8154.

Curriculum Vitae

Name Surname   : Tuğba Ergin

E-mail (1)   : tugbaergin1996@gmail.com

E-mail (2)   : tugba_ergin96@hotmail.com

Education:

2014–2019   İzmir Kâtip Çelebi University, Dept. of Biomedical Engineering

2019–2021   İzmir Kâtip Çelebi University, Dept. of Biomedical Technologies

Work Experience:

2020 June - 2021 July   Han Medikal A.Ş

2021 July - Contiuned   LM Wind Power Blades Turkey Sanayi ve Ticaret A.Ş

Publications:

1. Emotion detection using EEG signals based on Multivariate Synchrosqueezing Transform and Deep Learning, 2021 Medical Technologies Congress (TIPTEKNO), November 2021

2. Emotion Recognition with Multi-Channel EEG Signals Using Visual Stimulus, 2019 Medical Technologies Congress (TIPTEKNO), October 2019, DOI:10.1109/TIPTEKNO.2019.8895242

Republic of Turkey

İzmir Kâtip Çelebi University

Graduate School of Natural and Applied Sciences

Multi-Channel EEG-Based Emotion Classification Using Deep Neural Networks

Department of Biomedical Technologies

Master's Thesis

Tuğba Ergin

ORCID 0000-0002-1040-4848

Thesis Advisor: Assist. Prof. Dr. Onan Güren

October 2021

ERGiN

Multi-Channel EEG-Based Emotion
Classification Using Deep Neural Networks

MASTER'S THESIS

2021