B. YEŞİLKAYA IZMIR KATIP CELEBI UNIVERSITY 2020

IZMIR KATIP CELEBI UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ESTIMATION OF EMOTIONAL SITUATION USING EEG SIGNALS AND MACHINE LEARNING METHODS

M.Sc. Thesis

Bartu YEŞİLKAYA

Department of Biomedical Technologies

JUNE 2020

IZMIR KATIP CELEBI UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ESTIMATION OF EMOTIONAL SITUATION USING EEG SIGNALS AND MACHINE LEARNING METHODS

M.Sc. Thesis

Bartu YEŞİLKAYA
Y170204005
ORCID NO: 0000-0003-2235-4240

Department of Biomedical Technologies

Thesis Advisor: Prof. Dr. Aydın AKAN

JUNE 2020

İZMİR KATİP ÇELEBİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EEG SİNYALLERİ VE MAKİNE ÖĞRENME YÖNTEMLERİNİ KULLANILARAK DUYGUSAL DURUM KESTİRİMİ

YÜKSEK LİSANS TEZİ

Bartu YEŞİLKAYA
Y170204005

Biyomedikal Teknolojiler Ana Bilim Dalı

Tez Danışmanı: Prof. Dr. Aydın AKAN

HAZİRAN 2020

Bartu YEŞİLKAYA, a **M.Sc.** student of **IKCU Graduate School Of Natural And Applied Sciences**, successfully defended the thesis entitled "**ESTIMATION OF EMOTIONAL SITUATION USING EEG SIGNALS AND MACHINE LEARNING METHODS**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :

Prof. Dr. Aydın AKAN
İzmir University of Economics
....

Jury Members :

Assoc. Prof. Olcay AKAY
Dokuz Eylül University
....

Asst. Prof. Onan GÜREN
İzmir Katip Çelebi University
....

Date of Submission : **01.06.2020**
Date of Defense : **24.06.2020**

To my lovely family,

FOREWORD

First and foremost, I would like to thank my advisor, Prof. Dr. Aydın Akan for his support, guidance and patience throughout my thesis. I would like to thank my co-adviser Asst. Prof. Dr. Onan Güren for her guidance.

This thesis was supported by Izmir Katip Celebi University, Coordination Office of Scientific Research Projects, Project number is 2019-TYL-FEBE-0060.

Additionally, I would like to thank Res. Asst. Mazlum Unay for his help and patience during experiments. I would also like to thank my lab colleagues Leyla Nur Turhal, Mehmet Tugay Bahar, Res. Asst. Mehmet Uğur Soydemir, Res. Asst. Burak Ceylan for their support and contributions.

I would like to thank my family and my fiancée Miss Hazal Su Bıçakcı for their support, help, understanding and encouragement. I am grateful them for always being there for me through all my life.

June 2020 Bartu YEŞİLKAYA

TABLE OF CONTENT

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABBREVIATIONS

EEG	**:** Electroencephalogram
EMD	**:** Empirical Mode Decomposition
IMF	**:** Intrinsic Mode Function
SAM	**:** Self Assesment Manikin
IAPS	**:** International Affective Picture System
PSD	**:** Power Spectrum Density
SVM	**:** Support Vector Machine
BCI	**:** Brain Computer Interface
HCI	**:** Human-Computer Interaction
ECG	**:** Heart-electrocardiogram
EMG	**:** Muscle-electromyogram
EOG	**:** Electrooculography
EDA	**:** Electrodermal Activity
GSR	**:** Galvanic Skin Resistance
2D	**:** 2-Dimensional
HVHA	**:** High Valence-High Arousal
LVLA	**:** Low Valence-Low Arousal
LVHA	**:** Low Valence-High Arousal
HVLA	**:** High Valence-Low Arousal
TPR – TNR	**:** True Positive Rate-True Negative Rate
FPR – FNR	**:** False Positive Rate-False Negative Rate
FFT	: Fast Fourier Transform
EDR	**:** Electrodermal Response
SC	**:** Skin Conductance
MEMD	**:** Multivariate Empirical Mode Decomposition

DEAP	**:** Database for Emotion Analysis using Physiological Signals
ANN	**:** Artificial Neural Network
KNN	**:** K-Nearest Neighbors Classification
DWT	**:** Discrete Wavelet Transform
TF	**:** Time-Frequency
MSST	**:** Multivariate Synchrosqueezing Transform
ICA	**:** Independent Component Analysis
3D	**:** 3-Dimensional
CNN	**:** Convolutional Neural Networks
STFT	**:** Short Time Fourier Transform
HHT	**:** Hilbert Huang Transform
SD	**:** Standart Deviation
RMS	**:** Root Mean Square
TP – TN	**:** True Positive - True Negative
FP – FN	**:** False Positive - False Negative
AUC	: Area under the ROC curve
ROC	: Receiver Operating Characteristic
ReLU	: Rectified Linear Activation Unit
Tanh	: Hyperbolic Tangent
MLP	: Multilayer Perceptron

ESTIMATION OF EMOTIONAL SITUATION USING EEG SIGNALS AND MACHINE LEARNING METHODS

ABSTRACT

Emotion estimation is an effective analysis method used to increase the interaction between humans and machines. Electroencephalogram (EEG) based emotion prediction studies based on brain signals become very attractive since they provide successful results of emotion analysis. In this study, new methods for emotion prediction are presented in accordance with dimensional emotion modeling. Multichannel EEG signals are recorded while the subjects viewed pictures from the International Affective Image System (IAPS) data set. Signal preconditioning and artefact elimination was performed by applying necessary filters on the recorded data. Several features are extracted and the signals are classified using classification methods such as support vector machines and K-nearest neighbor. To improve the classification performance, we propose a second method where EEG signals are further analyzed by Multivariate Empirical Mode Decomposition (MEMD) and similar features are extracted from the intrinsic mode functions (IMFs) of the MEMD and classified using machine learning methods. As a third method, Deep Learning (DL) approach is proposed for classification of emotional labels. Time-frequency (TF) representations of the IMFs extracted using MEMD method are calculated by the Short-time Fourier Transform (STFT) and spectrogram. These spectrograms considered as TF images are applied to a Convolutional Neural Network (CNN) to classify the emotional labels. Performance results of the proposed methods suggest that utilizing an advanced signal processing method such as MEMD and using DL approach for classification provides encouraging results and may be used in future human-machine interaction studies.

Keywords —Emotion Analysis, Multivariate Empirical Mode Decomposition, Support Vector Machine, Spectrogram, Convolutional Neural Network

EEG SİNYALLERİ VE MAKİNE ÖĞRENME YÖNTEMLERİNİ KULLANILARAK DUYGUSAL DURUM KESTİRİMİ

ÖZET

Duygu tahmini insanlar ve makineler arasındaki etkileşimi arttırmak amaçlı kullanılan etkili bir analiz yöntemidir. Duygu analizi deneylerinde başarılı sonuçlar verdiği için günümüzde beyin sinyallerine dayalı, elektroensefalogram (EEG) tabanlı duygu tahmini araştırmaları ilgi çekmektedir. Bu çalışmada boyutsal duygu modellemesi doğrultusunda duygu tahmini analizi için yeni yöntemler önerilmektedir. Uluslararası Afektif Görüntü Sisteminden (IAPS) alınan görseller ile oluşturulmuş veri seti kullanılarak çok kanallı EEG sinyalleri kayıt edilmiştir. Bu veriler üzerinde gerekli olan filtreler uygulanarak veri temizleme ve gürültü giderme ön işlemleri gerçekleştirilmiştir. Daha sonra EEG sinyallerinden çeşitli öznitelikler elde edilmiş ve Destek Vektör Makineleri (DVM) ve K-en yakın komşu gibi sınıflandırma yöntemleri kullanılarak sınıflandırılmıştır. Başarı oranını arttırmak için EEG sinyalleri Çok Değişkenli Görgül Kip Ayrışım (ÇDGKA) yöntemi ile analiz edilerek elde edilen Özgün Kip Fonksiyonlarından (ÖKF) benzer öznitelikler hesaplanmış ve sınıflandırılmıştır. Duygu durum kestirimi için üçüncü bir yöntem olarak derin öğrenme tabanlı bir yaklaşım önerilmiştir. ÇDGKA ile elde edilen ÖKF'lerin zaman-frekans (ZF) gösterimleri Kısa Zamanlı Fourier Dönüşümü (KZFD) ve Spektrogram ile elde edilmiştir. Hesaplanan spektrogram matrisleri ZF imgesi olarak değerlendirilmiş, evrişimsel sinir ağını eğitmek ve duygu durum kestirimi için giriş olarak kullanılmıştır. Önerilen yöntemlerin başarım sonuçları, ÇDGKA gibi ileri sinyal analiz yöntemi ve derin öğrenme yaklaşımı kullanılarak duygu durum kestiriminde başarılı sonuçlar elde edildiğini ve gelecek insan-makine etkileşim sistemlerinde kullanılabileceğini göstermektedir.

Anahtar Kelimeler — Duygu Analizi, Destek vektör makineleri, Grup Görgül Kip Ayrışımı, Spektrogram, Evrişimsel Sinir Ağları

1. INTRODUCTION

Demonstration and perception of emotions are a part of communication between people and also affect people's perceptions and reactions to events. During mutual communication, how to say as much as what is said, facial expressions and body movements play an effective role in the success of communication. Emotions can be expressed through many different channels, and the emotional states of individuals can be determined by analyzing various features. Speech and facial expressions are the most beneficial information channels in determining emotion [1,2]. However, such channels can be easily manipulated by the person and thus can provide misleading information about the true emotional state of the person. The information to be obtained from physiological channels can be considered as a more reliable source for determining the emotional state of the person, since it is the result of more natural and more difficult to control reactions. In this study, emotion recognition study was carried out by using physiological signals measured by EEG.

1.1 Physiological Signals

Physiological signals are signals detected from the living body through electrodes or transducers. While signals such as electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), electroneurogram (ENG) are biological signals, signals such as blood pressure, blood flow rate, respiratory volume, heart sounds, temperature and skin resistance are not electrical origin. Electric origin signals are formed as a result of chemical events occurring in cells. The passage of Na +, K +, Ca + and Cl- ions through the cell membranes creates potential differences, causing such signals to appear.

1.1.1 Action potential

Neurons consist of axons, soma and dentrites. Dentrites transmit the information that other neurons get from their axons to the cell body, whereas axons carry the information from the cell body to other neurons are shown as in Figure 1.1. In a resting neuron, Na +, Cl- concentration outside the cell and K + concentration in the cell are high. When the neuron is stimulated, the permeability of the cell membrane to Na + ions increases and Na + ions pass into the cell by diffusion. Intracellular sodium ion concentration increases rapidly and a potential difference occurs between intracellular and extracellular. If the intensity of the excitation exceeds the threshold voltage of the cell, there is a potential that rises to about +30 mV. This rise is called depolarization.

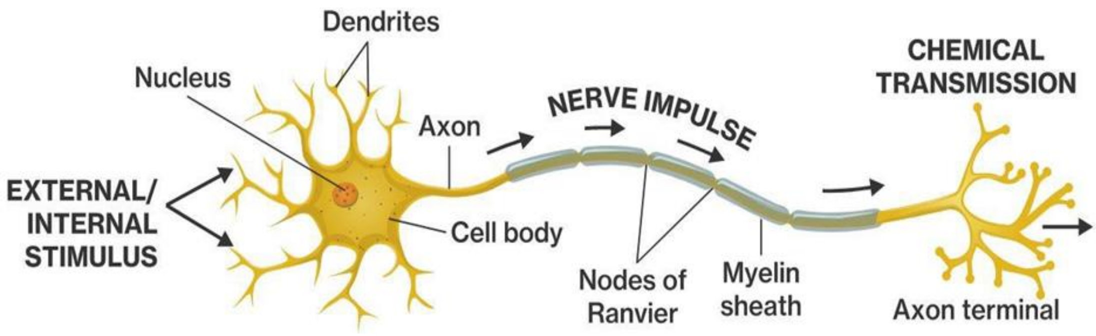


Figure 1.1 The structure of biologic nerve cell [3].

After depolarization, while Na + channels suddenly close, K + channels open. When the K + channels are opened, the cell membrane begins to return to its resting potential, which is called repolarization. The potential difference decreases from the resting potential to a negative value, to -90 millivolts. This is called hyperpolarization. In case of hyperpolarization, Na + - K + active pumps start working with energy and Na + ions are pumped out of the cell and K + ions into the cell to bring the potential difference to the resting level of -70 millivolts.

1.2 Brain Concept and Nervous System

The system that reacts against the actions of living things and controls all body activities of sight, smell, taste etc. is called the brain. As shown in Figure 1.2, the brain consists of four parts, the brain stem, the cerebellum, the midbrain and the membrane.

Brain stem: Responsible for autonomous processes and reflexes. It plays an important role in maintaining sleep patterns by controlling the central nervous system. It is responsible for organizing vital activities such as blood pressure and heartbeat.

Cerebellum: It plays a key role in creating the balance of the body. It regulates the coordination in the body, allowing people to move and stand.

Midbrain: Regulates body temperature. Provides control of instinctive activities and emotions

Cerebrum: Responsible for memory. Performs control of motor functions. Emotions are also covered in this section.

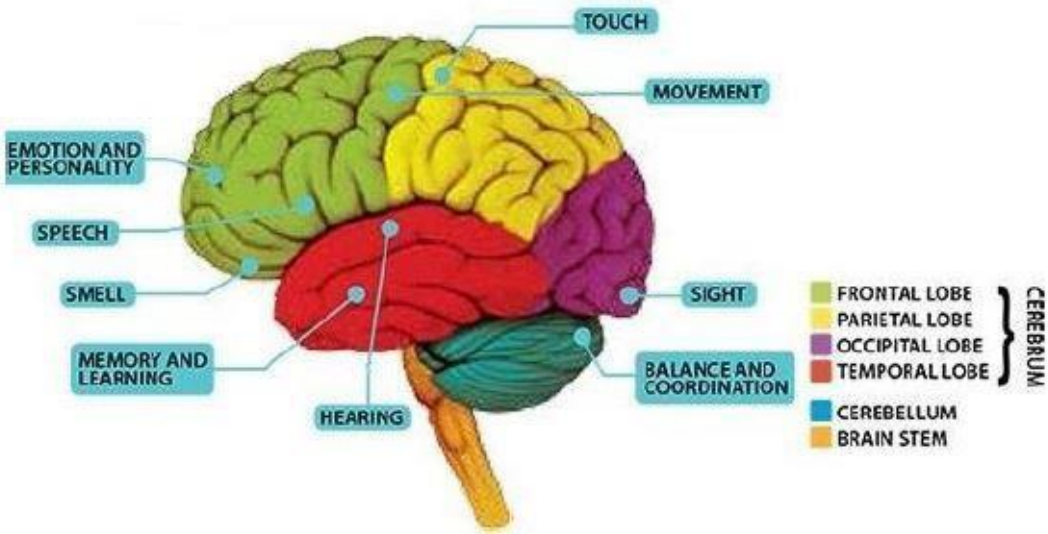


Figure 1.2 Brain lobes [4].

The brain is divided into two separate parts from front to back and each is called hemisphere. Each hemisphere consists of four separate lobes and these lobes are referred to as the temple lobe, the backbone lobe, the side peak lobe and the frontal lobe. The function of each lobe in the body differs. While the frontal lobe located in the front part of the brain controls behavioral actions such as problem solving and social activities, the side lobe is more responsible for activities such as feeling, perceiving and touching. The arterial lobe, the lobe located at the back of the brain, performs tasks related to the sense of vision. In case of damage, it causes delusions and visual impairments in people. The temple lobe, the last lobe of the brain, is responsible for sound and smell. They are located at the level of the ear of the brain and perform actions such as speech, hearing.

1.3. Electroencephalography (EEG)

When the history of brain signals is examined, it was seen that the first study carried out in this field was done in 1887 by Richard Caton. Caton conducted experiments on various animals using a galvonometer to measure electrical activity in the brain [5]. Likewise, in 1890, Adolf Beck used various stimuli on some animals and observed that fluctuations in the brain disappeared. In 1912, after the study of Pravdich-Neminsky on dogs, signals were obtained from the brain [6]. Following these developments, electrodes were tested on individuals injured during the First World War and the first EEG recording marks were obtained [7].

EEG is a method that does not require intervention to measure and evaluate the electrical activity in the brain through electrodes placed on the scalp. The interaction of the cells with each other causes the electrical activity in the brain to be continuous and causes the signal to be variable (non-stationary). In the past, collecting EEG data, difficult to obtain and making with large machines have affected the technology positively in this direction. Today, EEG data can now be obtained via portable devices via Wi-Fi and Bluetooth. This has led to increased studies with EEG and the development of a large number of BCI applications. It is used not only in the field of health but also in academic studies and provides important information. With EEG, studies such as emotion analysis and prediction, sleep disorders, epilepsy, noise classification, electrode channel selection are carried out.

1.3.1. Electroencephalography signals and subbands of eeg signals

According to frequency EEG oscillations can be divided into 5 categories. These are briefly described as below.

Table 1.1 The five frequency bands of EEG signal.

FREQUENCY BAND NAME	FREQUENCY BANDWIDTH	MIND STATUS
DELTA (Δ)	0.5 Hz – 4 Hz	Moment of deep sleep
THETA (Θ)	4 Hz – 8 Hz	Deep hypnosis
ALPHA (Α)	8 Hz – 14 Hz	Mild Hypnosis
BETA (Β)	14 Hz – 33 Hz	Alertness
GAMA (Ɣ)	33 Hz – 100 Hz	Overactive

Delta Waves: These are waves formed as a result of slow-acting brain activities with a frequency below 4 Hz. Occurs in deep sleep in adults [8].

Theta Waves: Its frequency range is between 4 and 8 Hz, and is seen in normal babies and children. In adults, however, it is in drowsiness, but it is also rarely seen in awake states. They can occur in situations such as emotional tension. Their amplitude is less than 100 μV (p - p) [8].

Alpha Waves: Brain waves with a frequency range of 8 to 14 Hz. Alpha band rhythms occur in normal adults at rest and in slightly sleepy situations. Their amplitude is generally below 50 μV. Alpha waves can be clearly observed in the occipital region. The formation of these waves is prevented when the eyes are open or deep in thought [8].

Beta Waves: They are waves with a frequency range of 14 to 30 Hz. Beta waves often dominate the front of the brain. They are waves with higher frequency and lower amplitude than alpha waves. They become more evident in the moments of problem solving and recall. Also, activation increases in beta waves when various stimulants such as coffee and energy drinks are taken [8].

Gamma Waves: They are waves with a frequency above 30 Hz. Waves in this frequency range are often used outside clinical trials. So in many applications these waves are filtered from EEG recordings [8].

1.3.2. Measurement of eeg signals

In studies based on brain signals, EEG data is obtained through electrodes placed on the scalp. Electrodes are placed on the head surface according to the 10-20 system defined by the International EEG Federation Union. In this system, the electrodes are placed in four main areas of the head, the nose, the back of the head and two ears, with a distance of 10-20-20-20-20-20-10 cm between them. The image of the system used in Figure 1.3 is given from the top and side.

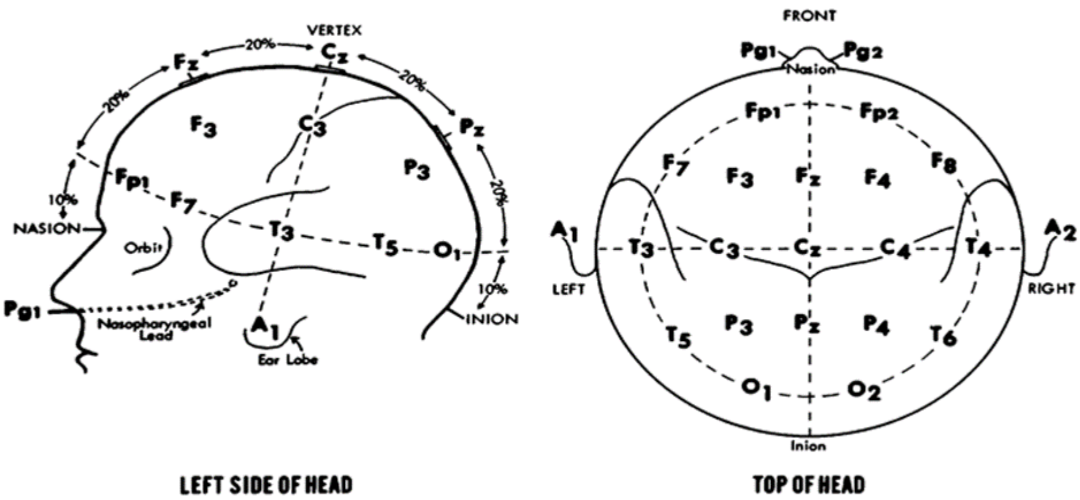


Figure 1.3 Electrode placements according to international 10/20 system.

Reference electrodes are needed when acquiring brain signals with EEG. The purpose of the reference electrode is to reduce the noise and errors that occur during the measurement. In the 10-20 electrode system, reference electrodes are placed in the ear. As seen in Figure 1.3, the parts that each electrode corresponds to on the head surface are expressed with letters and numbers. While the left hemisphere of the brain is defined by odd numbers, the right hemisphere is indicated by even numbers. The middle part is expressed by the letter z. Fp represents the forehead region, F front, C middle, T temple, P rear upper and O rear region.

1.4. Emotional Phenomena

Emotion phenomena can vary from person to person. 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.

1.4.1 Classification of emotions

According to the data obtained in the experimental studies, it was revealed that there is a connection between the mental activities of the person and the EEG frequencies. As a result of these studies, it is understood that the human brain produces different frequencies according to the emotional state and mind activities. Emotion can be briefly called the response created to the external factor. It is observed in the literature that many studies have been carried out on emotion estimations. Examples of these are emotional analysis studies aimed at increasing human-computer interface interactions. The emotional analysis study was first carried out in the 1990s on facial expressions and voice toning. In this study, in which EEG signals were not used, there were many deviations on the data. To eliminate these deviations, the researchers applied physiological measurements. Emotion analysis studies started to increase with the increase of human-machine interaction by using EEG signals and records. Multiple emotion models are used in emotion detection studies. These can be exemplified as discrete emotional model and dimensional emotional model. The discrete emotion model structure contains six different emotions in itself. These are fear, sadness, hate, surprise, nervousness and happiness. In the dimensional emotional model, the valence-arousal plane defined by Russell is used.

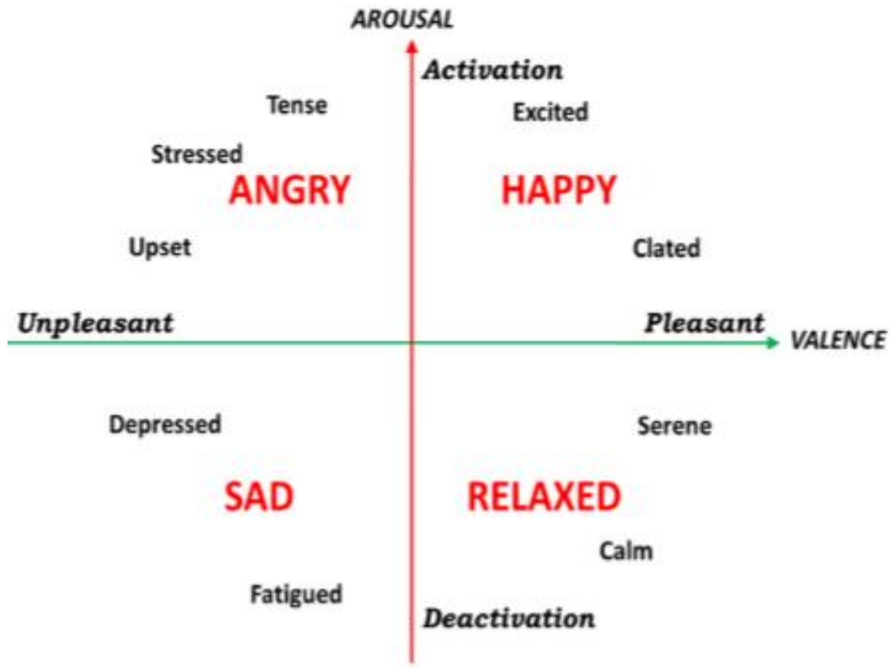


Figure 1.4 The circumflex model of emotion by James Russell [9].

In this model structure, emotions in arousal coordinates are defined from calmness to excitement, while emotions are expressed from negative to positive in the valence coordinate system. Valence represents the love or hatred that individuals show towards a particular situation or event. On the other hand, arousal indicates the physiological state of individuals and represents the passive or active state shown against a stimulus.Two-dimensional arousal-valence coordinate in Figure 1.4 the plane is shown.

Although the discrete emotional model structure is easy to implement and understand, in some languages there is no translation of certain emotions [10]. Therefore, in most studies, a 2D or 3D arousal-valence plane is preferred [11,12]. Also, the universality of the arousal-valence model makes the model more attractive. In this model, a certain emotion is expressed not based on the discrete but on the coordinate system. For example, the sense of happiness is in the HAHV (High Arousal High Valence) area according to the plane. Likewise, his sad expression belongs to the LALV (Low Arousal Low Valence) region.

1.5 Literature Review

Based on the literature review, there have been a lot of studies about emotion recognition from EEG signals by using visual and auditory stimulus and other stimulus, emotion classification using stimulated EEG signals applying empiricial mode decomposition algorithm for EEG signal classification.

Ahmet Mert and Aydın Akan [13], explored the advanced properties of empirical mode decomposition (EMD) and multivariate empirical mode decomposition (MEMD) for emotion recognition. The MEMD-based feature extraction method is proposed and this poroposed method is applied to DEAP emotional EEG data set. The authors take EEG signals from 32 channels EEG device. From these 32 channels, 18 channels most informative commonly used for emotion analysis are selected. Multichannel IMFs issued by MEMD are analyzed using features such as spectral power asymmetry, power spectral density, Hjorth parameters, entropy, band power ratios, correlation and coherence. These features were classified using the K-nearest neighbor (K-NN) and artificial neural network (ANN) classifiers. The results show that the proposed method with K-NN has accuracy rates of 51,01% and 67%, while ANN is 75% and 72,87 respectively for arousal and valence states. As a result, nonlinear and nonstationary decomposition of MEMD with ANN provide more appropriate analysis to detect emotion form nonstationary EEG signals.

Yi-Hung Liu et al. [14] in their study, EEG-based emotion prediction was aimed. International Affective Picture System (IAPS) was used to stimulate the emotions. Data is obtained from subjects through visual stimuli and these data were subjected to classification in the arousal-valence plane. The method they proposed consists of three parts. In the first part, the EEG data obtained was decomposed to its lower frequencies using the 5th order Butterworth filter and the spectrum power feature were extracted from each frequency. The size of the feature vector obtained in the second part is reduced by using two separate (using different parameters) Kernel Fisher Discrimination Analysis method. In the last part, the classification process with the reduction of the features, arousal and valence distinction was made using the K-NN classifier and then compared with other methods. As a result of the classification process, 75.53% success was obtained for valence and 79.19% success for arousal.

Raja Majid Mehmood and others [15] obtained EEG data using visual stimuli and aimed to classify positive / negative emotions. After the EEG signals are obtained with the portable Emotiv Epoc device, they are free from noise by independent component analysis. The steps for using stimuli are summarized as follows. In the first stage, the subjects focused on the plus (+) sign displayed on the screen for 4 seconds, and then looked at the image reflected on the screen for 1.5 seconds. In the last part, black screen is shown for 0.5 seconds. This procedure was completed by showing 180 images for 21 subjects. Feature space was obtained using Hjorth parameters and feature selection was made using genetic search and particle herd optimization. In the last part of the study, the features selected with particle herd optimization and genetic search were classified with support vector machines and approximately 48.85% and 57.42% success were obtained, respectively.

Mostafa Mohammadpour and others [16] in their proposed method, it was aimed to classify six different emotions by using neural networks, support vector machines, K-NN and Bayes and compare their results. In the study using the dimensional model, the signals were obtained visually and separated into wavelets by the discrete wavelet transform technique. The feature vector obtained by using statistical methods is reduced by principal component analysis. The average performance at the end of the study was 55.58%, 51.82%, %49.82, %50.7 using neural networks, support vector machines, K-NN and bayes, respectively.

In the system that Bharasaka Krisnandhika et al. [17] developed, emotion recognition was performed using radial basis function networks. After the original signal was divided into 5 level discrete wavelet transform, it was obtained by the relative wavelet energy method and the classification process was made. In the classification process, the parameters of artificial neural networks are designed so that the input layer consisted of 192 neurons, the hidden layer consisted of 45 neurons, and the output layer consisted of 2 neurons. In addition, the effect of the number of people on emotion prediction was investigated and it was observed that the classification success increased as the number of people increased.

Henry Candra and others [18] in the study, the separation of the excitation and valence plane was made using support vector machines. The pre-treated EEG signals are decomposed into db5 wavelet and alpha, beta, theta, gamma and delta frequencies. A feature vector was obtained from the separated frequencies and emotion separation was made using support vector machines. When the performance rates were compared, it was seen that the valence classification was higher than the arousal plane. In addition, it is stated that the feelings of valence are effective in the F3, F4, Fz, FC1, FC2, FC5, FC6, C3, C4, CP1, CP2, CP5, P3, P8, Pz and O1 EEG channels. It is stated that excitation feelings are effective in Fp2, AF3, F3, F7, F8, PC5, FC6, C3, T8, P3, P4, P7, Pz, PO4, O2 and Oz channels.

Ahmet Mert, Aydın Akan [19], analyze the feasibility of using time–frequency (TF) representation of EEG signals for emotional state recognition. Multivariate synchrosqueezing transform (MSST) that is time-frequency analyzing method is used for feature extraction. Independent component analysis (ICA) and proposed feature selection are applied. The TF-domain reduction performance of ICA and non-negative matrix factorization (NMF) is compared to observe the accuracy levels of high/low arousal, and high/low valence emotional recognition. As classifier, artificial neural network (ANN), K-nearest neighbor (K-NN), support vector machine (SVM) are used and the best result is obtained by ANN. By using ANN, the proposed feature extraction shows the accuracy rates of 82,11% and 82,03% for arousal and valence, respectively. TF representation is used successfully for feature extraction.

1.6 Contributions of the Thesis

In this thesis, three new methods for emotion prediction are presented in accordance with dimensional emotion modeling. **Multichannel EEG signals are recorded while the subjects viewed pictures from the International Affective Image System (IAPS) data set using Brain Product Acti CHamp EEG system.** Signal preconditioning and artefact elimination were performed by applying necessary filters on the recorded data. Then the following approaches are presented for emotion recognition:

I. Multichannel EEG signals are processed to extract features and the signals are classified using machine learning methods.

II. We propose a second method where EEG signals are further analyzed by Multivariate Empirical Mode Decomposition (MEMD) and similar features are extracted from the intrinsic mode functions (IMFs) of the MEMD and classified using machine learning methods.

III. In order to eliminate the feature extraction step, we propose a third method and apply Deep Learning (DL) approach for the classification of emotional labels. Time-frequency (TF) representations of the IMFs extracted using MEMD method are calculated by the Short-time Fourier Transform (STFT) and spectrogram. These spectrograms considered as TF images are applied to a Convolutional Neural Network (CNN) to classify the emotional labels.

The first method has achieved a maximum success rate of 69%. The second method has achieved a maximum success rate of 81%. The third method has achieved a maximum success rate of 92%. Thus, the last method was found to be the most successful. Success rates of similar studies using Deap dataset are given in Table 1.2.

Table 1.2 Similar studies.

Authors	**Emotion Parameter**	**Method**	**Classification Success**
Zhang and others, 2016	Valence, Arousal	Power (FFT)	%58.75
Kumar and others, 2016	High, Low Valence, Arousal	Bispectral Analysis	%64.84, Arousal %61.17, Valence
Atkinson and Campos, 2016	Valence, Arousal	Statistical Features Band Power of Different Frequencies Hjorth Parameters Fractal Size	%60.72 %62.47
Zhuang and others, 2017	Valence, Arousal	EMD & Normalized Energy	Valence: %69.10 Arousal: %71.9
Liu and others, 2017	Valence, Arousal	DWT, Differential Asymmetry, PSD, Connectivity Time domain features	High- Low Arousal: %74.3 Valence: %77.2
Tripathi and others, 2017	Valence, Arousal	DANN, CANN	DANN: %58.44, CANN: %66.79
Al-Nafjan and others, 2017	Valence, Arousal	Frontal Asymmetry PSD	%82
Mert and Akan, 2018	Valence, Arousal	MEMD & Power Ratio, PSD, Entropy Hjorth Parameters, Correlation	%75.00±7.48 % 72.87±4.68

2. MATERIALS-METHODS

2.1 Data Acquisition

2.1.1 Brain product acticap

In this study, Brain product Acticap EEG device was used to collect data from subjects. The device has 32 EEG channels and 8 extra AUX channels for the beginner level. EEG channels can be increased with modules added to the tool so that they can be 64, 96, 128, 160 channels, respectively. The lithium-ion battery is portable and has a long battery life. It allows to do 15 experiments at full filling. The battery and device are shown in Figure 2.1.

Figure 2.1 Brain Product Acticap EEG device**.**

2.1.2 Imotion

Information slides with SAM surveys were prepared thanks to the program. It helped us determine how long each slide would stay on the screen. Since the brain product was able to establish a software connection with the Acticap device, Imotion 8.0 program was used while recording EEG. The program allowed us to get EEG recordings as excel file after the experiment. The program interface is shown in Figure 2.2.

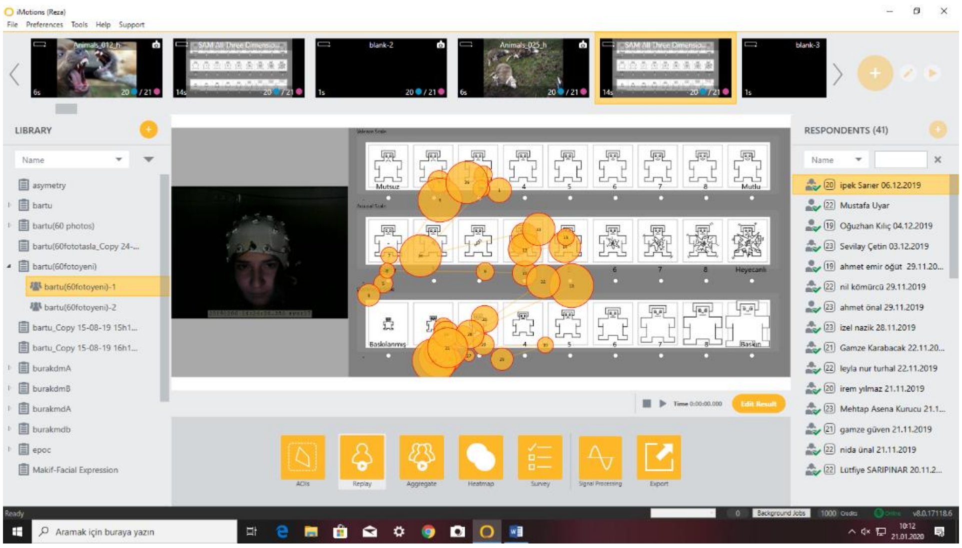


Figure 2.2 Imotion interface.

The program offers an interface to check the connection with the electrodes during the experiment. In this interface, impedance values are displayed for each electrode. The impedance values gradually decrease by means of EEG gel from 999 kilo ohm. The interface shows red color for electrodes above 45 kilo ohms, yellow color for values between 45 kilo ohms and 30 kilo ohms, and green for values below 30 kilo ohms. As shown in Figure 2.3., it appears as red for high impedance electrodes, yellow for medium impedance electrodes, and green for ready to record electrodes. Before starting the experiment, after making sure that the impedance value of all electrodes were less than 30 kilo ohms, the experiments were started. Electrode impedance values before the experiment starts are shown in Figure 2.3.

Figure 2.3 Impedance values of EEG electrodes.

2.1.3 IAPS database

In this study, in order to use the photos of the International Affective Image System (IAPS) data set, the forms were signed by contacting and the data were accessed over the Internet. A total of 1356 photographs were taken from the database, including 5 categories (landscape, human, animal, object, and face). There are a different number of photos in the categories. There are 250 photographs in the human category, 221 in the animal category, 372 in the face category, 185 in the landscape category, and 328 in the object category. As a result of previous trials, these images were classified according to arousal and valence values on 3 groups as female, male, and total (female-male). Considering the arousal-valence values, resolutions and physical parameters of the photographs, 12 photos were selected from the animal, object, and face categories, 13 photos from the human category and 11 photos from the landscape category. A total of 60 photographs were determined for the experiment. Selected photographs are exemplified in Figure 2.4.

Figure 2.4 IAPS visual examples.

Table 2.1 Details of stimuli used in visual experiment.

	IAPS NAME	VALA-NCE	AROU-SAL	VALA-NCE	AROU-SAL
HIGH VALENCE HIGH AROUSAL (HA-HV)	Animal_106_h	6,00	4,81	7,80	6,05
	Animal_175_h	7,72	5,43	6,89	5,95
	Animal_179_h	7,36	5,44	7,68	5,32
	Faces_234_h	7,65	6,03	6,95	5,32
	Faces_321_h	7,04	6,38	7,71	6,00
	Faces_347_h	7.31	5,84	6,59	5,61
	Landscapes_008_h	7,50	4,07	7,25	4,90
	Landscapes_023_h	6,38	4,33	7,05	4,71
	Objects_037_h	6,79	4,93	7,78	6,04
	Objects_069_h	7,07	5,45	6,40	5,50
	Objects_291_h	7,16	6,72	6,91	6,86
	People_175_h	7,11	5,77	7,35	6,65
	People_189_h	7,39	5,84	6,70	6,23
	People_193_h	7,21	6,09	7,43	5,86
	People_196_h	7,65	5,52	7,52	6,16
HIGH VALENCE LOW AROUSAL (HA-LV)	Animal_095_h	7,08	2,43	7,16	3,12
	Animal_140_h	7,50	2,52	6,72	3,17
	Animal_184_h	8,00	2,57	7,91	2,25
	Faces_052_h	6,79	2,97	6,67	3,33
	Faces_061_h	7,17	3,87	7,05	3,05
	Faces_089_h	7,90	3,24	7,78	3,06
	Landscapes_104_h	6,94	2,68	7,60	3,00
	Landscapes_165_h	8,26	2,16	8,13	3,40
	Landscapes_176_h	7,54	2,29	7,48	2,95
	Objects_190_h	7,22	3,63	7,18	3,57
	Objects_326_h	7,60	2,93	7,91	3,91
	Objects_327_h	7,89	3,19	7,42	3,97
	People_110_h	8,29	2,53	7,83	2,92
	People_112_h	7,34	3,14	7,28	3,11
	People_154_h	7,50	2,81	7,72	3,00
LOW VALENCE HIGH AROUSAL (LA-HV)	Animal_012_h	3,30	6,64	4,05	6,95
	Animal_025_h	2,37	6,73	3,05	6,50
	Animal_048_h	3,03	6,73	2,50	6,41
	Faces_016_h	2,81	6,89	3,07	7,17
	Faces_174_h	2,43	7,39	2,81	6,33
	Faces_293_h	2,52	7,15	2,35	7,00
	Landscapes_005_h	2,66	6,66	3,94	6,67
	Landscapes_007_h	3,04	6,04	2,80	6,70
	Landscapes_022_h	2,85	6,88	2,64	6,82
	Objects_001_h	2,54	7,25	2,91	7,04
	Objects_003_h	2,46	6,97	2,30	6,54
	Objects_283_h	3,60	6,00	3,96	5,91
	People_017_h	2,60	7,40	3,53	7,50
	People_038_h	1,71	7,86	1,76	8,38
	People_127_h	2,21	7,61	2,09	7,63

Table 2.1 (continuation) Details of stimuli used in visual experiment.

	IAPS NAME	VALANCE	AROUSAL	VALANCE	AROUSAL
LOW VALENCE LOW AROUSAL (LV-LA)	Animal_053_h	3,13	6,20	4,05	5,00
	Animal_054_h	2,56	6,56	2,80	6,47
	Animal_110_h	4,13	5,88	4,61	5,28
	Faces_011_h	3,10	5,67	3,47	5,00
	Faces_155_h	4,08	5,07	3,84	5,30
	Faces_157_h	4,28	5,03	4,23	4,05
	Landscapes_011_h	3,38	5,81	3,70	5,40
	Landscapes_028_h	4,33	5,25	3,33	5,13
	Landscapes_156_h	4,16	5,42	4,25	5,25
	Objects_122_h	4,33	5,15	3,77	4,83
	Objects_144_h	3,13	5,72	2,81	5,63
	Objects_236_h	3,84	5,22	4,42	4,86
	People_080_h	3,61	5,22	4,68	5,00
	People_137_h	3,78	5,94	4,00	5,16
	People_141_h	4,11	5,44	4,21	5,65

IAPS shared an excel file containing the names, properties, light values of the photographs and the survey results of 204 participants (119 women, 85 men; mean age = 23.9 years, SD = 3.4) created using these photos [20]. In the survey results, the valence and arousal values of the photographs have average values for women, men and all participants. Using this excel file, we decided which photos would be used in the experiment as stimuli. First, we removed vertical photos from the list, leaving only horizontal photos. We first sorted the photos according to their valence value. Then we grouped the first 150 as high valence and the last 150 as low valence. Then we sorted these 150 photo groups according to their arousal values. We grouped the first 75 as high arousal and the last 75 as low arousal. Thus, we obtained 4 groups of photographs, among which we eliminated blood, brutality and nausea. Finally, a total of 60 photos were selected for the experiment, with 15 photos per group. Valence and arousal values of the selected photos are shown in Table 2.1. These groups are given to the classifier as output. Thus, supervised classification was made.

2.1.4 SAM scale

After each photo was shown to the users, Self Assessment Manikin (SAM) questionnaire was applied to enable users to evaluate their emotions. An image of the SAM survey is given in Figure 2.5. In SAM form, there are 3 scales rated from 1 to 9. These are arousal, valence, and dominance. The users were asked to evaluate these

3 scales according to the emotional state they were in after each image. In short, the valence scale expresses unhappiness when it approaches 1 and happiness when it approaches 9. The arousal scale expresses calmness when it approaches 1 and excitement when it approaches 9. Dominance refers to emotion intensity on the person. When it approaches towards 1, that means that the person feels intense and the emotion is dominant over the person. When it approaches towards 9, it means that the intensity of the emotion that the person is in is not enough and that the person is dominant.

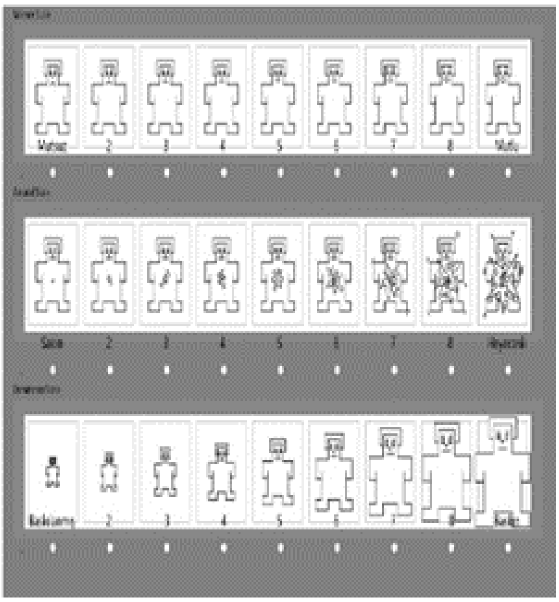


Figure 2.5 Self Assessment Manikin (SAM) survey.

2.1.5 Exprimental setup

The experimental group consists of 40 people, 20 women and 20 men, selected from the age range of 18 - 25, and consists of volunteers who are studying at university. The experimental setup consisting of photographs to be shown to individuals and to be evaluated after seeing the related photographs was created with Imotion 8.0 program. Using the 32-channel Brain Vision BrainAmp EEG device, the electrodes were placed at 32 different points of the skull and the received signals were recorded. The individual was asked to evaluate the visuals presented in a soundproofed and dark room with the supervision of a person according to 3 evaluation scales (arousal, valence, dominance). The test environment consists of two compartments and the screen showing the photographs was transferred in parallel to another screen in order for the supervisor to follow the test process. In this way, the subject can focus on the experiment without distraction. In Figure 2.6 visual of the experimental environment is given. In addition, the supervisor had instant information about the assessment form filled by the subject.

Figure 2.6 Experimental environment.

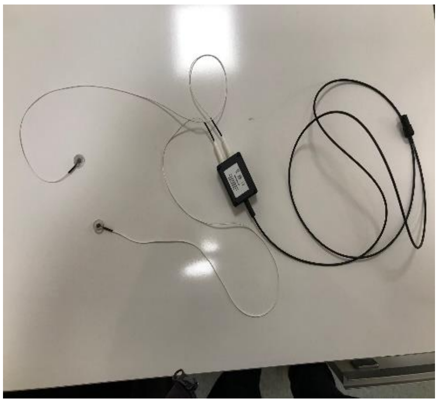
Test procedure:

- Electrode gels to be used were filled into syringes.
- Connections of the EEG device and eye tracking system were made to the computer.
- We had 2 different EEG caps. The cap suitable for the subject's head structure and head size was chosen.
- The cap was firmly placed on the subject's head.
- The hair in the electrode areas on the sheath was opened to allow the electrodes to contact the skin.
- 32 electrodes were placed according to the Acticap system.
- In order to increase the conductivity of the electrodes, gel was applied to the scalp of the subject through the cavities on the surface of the electrodes.
- The impedance level of the electrodes was observed thanks to the Imotion 8.0 program.
- Between the electrode surface and the scalp, gel was squeezed through the cavities in the sheath to increase the impedance values to less than 30 kilo ohms.
- After this process was applied for 32 channels, a registration page was created by entering the subject's name, gender, and age information.

- GSR (Galvanic Skin Response) electrodes (2 pieces) were connected between the left index and the first and second nodes of the left middle finger. A separate gel was also applied to these electrodes.
- The distance of the subject from the monitor was captured by the Tobii eye tracking system.
- Finally, the subject's electronic items were taken. The materials that could distribute attention were eliminated. The light and the door of the room was closed.
- In order for the experimental procedure to work properly, a moderator inside examined the operation of the subject from a different screen.

(A)

(B)



(C)

(D)

Figure 2.7 (A) GSR gel, **(B)** EEG electrolyte-gel for active electrodes,

(C) GSR (Galvanic Skin Response) electrodes, **(D)** Tobii eye tracker.

Each stage of each photograph shown to the individual is completed with: 1 second rest period, 6 seconds photo display, and 14 seconds for subsequent 3 scale evaluation. In addition, a total of 6 slides were added at the start of the experiment for informational purposes, and each of these slides was displayed for 20 seconds. During the experiment, eye tracking was performed with Tobiipro Eye Tracking System and the experiment was recorded with a camera. In Figure 2.8, the planning of the experiment period is visualized.

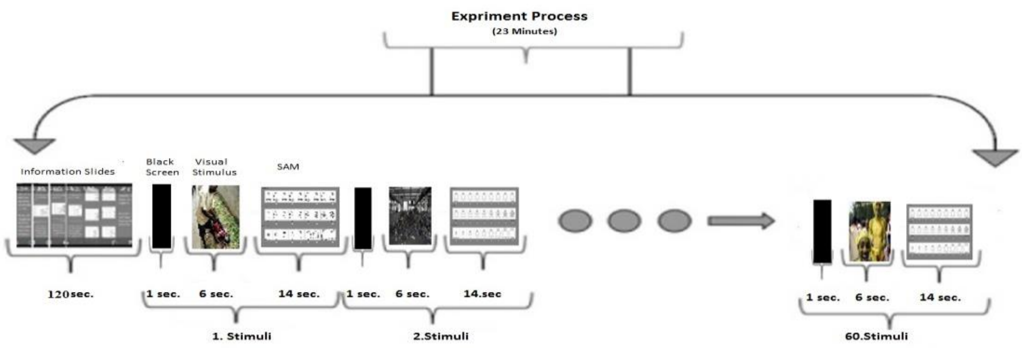
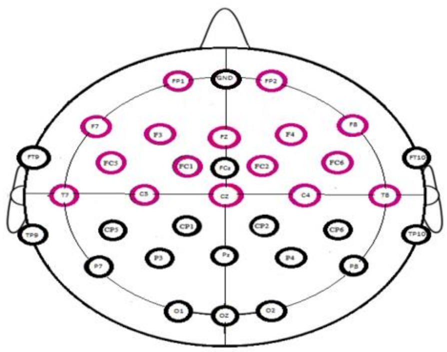


Figure 2.8 Visual scheme of the experimental procedure.

2.2 Channel Selection

Data was received from 32 channels in total using the Brain Vision BrainAmp EEG device. However, in this study, 16 channels, FP1, FP2, F7, F3, FC1, FC2, FC5, FC6, C3, C4, T7, T8, F4, F8, Fz, Cz and earth electrode were studied. In the literature, we have made this selection by seeing that the channels used in frontal lobe of the brain have achieved more successful results [19]. The distribution of the used channels and their names are shown in Figure 2.9 and Table 2.2.



Used Channels   Unused Channels

Figure 2.9 Distribution of selected channels.

Table 2.2 EEG channel numbers and location of electrodes.

Number of Electrodes	Name of Channel
1	Fp1
2	Fz
3	F3
4	F7
5	FT9
6	FC5
7	FC1
8	C3
9	T7
10	TP9
11	CP5
12	CP1
13	Pz
14	P3
15	P7
16	O1
17	Oz
18	O2
19	P4
20	P8
21	TP10
22	CP6
23	CP2
24	Cz
25	C4
26	T8
27	FT10
28	FC6
29	FC2
30	F4
31	F8
32	Fp2

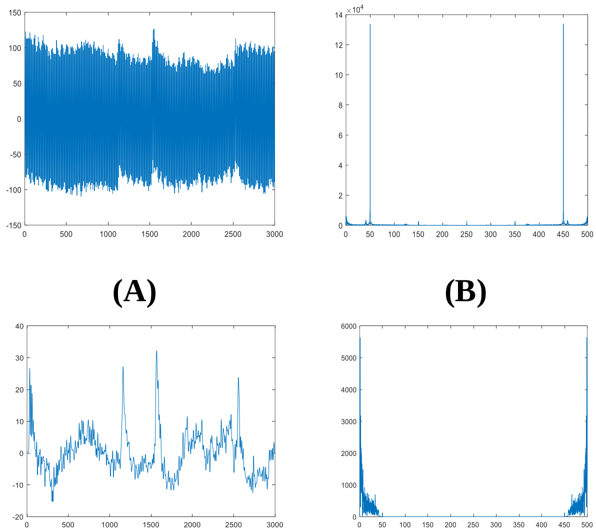
2.3 Preprocessing

The 32 channel Brain Vision BrainAmp EEG device used has a sampling frequency of 500 Hz. However, this frequency is quite large for the signal frequency range examined in EEG. It is also classified as 5 main spectral bands for EEG signal examination. These include alpha (α), beta (β), theta (θ), delta (Δ), and gamma (ɣ).

Filtering was performed to make the recorded EEG data suitable to lower frequency ranges. MATLAB interface was used in the filtering process. The data recorded in the Imotion 8.0 interface are taken as excel files. The EEG data recorded during the duration of the experiment were extracted to process appropriate ranges. The sorting process was performed by removing the EEG data from the data of each subject, first in the display period of the information slides, and then in the duration of the black screen shown before each stimulus, and during the SAM surveys.

The signals were filtered to remove unwanted frequency components. MATLAB program was used in the filtering process. After the data were transferred to the MATLAB program, they were stored with the names of the stimuli. Firstly, the average value of the data was subtracted from it so that the DC component was discarded from the signal. Then, stimulus data were passed through filters one by one. It was first passed through a low-pass filter. Stop band of the low-pass filter is set to suppress frequency components of 45 Hz and beyond. Pass band of the low-pass filter is adjusted to pass the frequency components of 35 Hz and below. Then, the signals were passed through the high-pass filter. Pass band of the high-pass filter is adjusted to pass the frequency components of 0.1 Hz and below. Stop band of the high-pass filter is adjusted to suppress frequency components of 0.4 Hz and beyond. Thus the transition band frequency of high-pass filter was between 0.1 – 0.4 Hz. As a result, frequencies between 0.1 Hz and 45 Hz were taken.

In Figure 2.10(a), an eeg signal is first drawn by the number of samples. In Figure 2.10(b), components of the eeg signal in the frequency domain are plotted according to the frequency scale. In Figure 2.10(c) an eeg signal is drawn with the sample number after filtering. In Figure 2.10(d), components of the filtered eeg signal in the frequency domain are plotted according to the frequency scale.



(A) (B)

(C) (D)

Figure 2.10 (A) Sample of unfiltered data, **(B)** FFT of unfiltered data, **(C)** Sample of filtered data, **(D)** FFT of filtered data.

2.4 Signal Processing

Finally, we applied the MEMD (Multivariate empirical mode decomposition) algorithm to open the IMFs without inserting the filtered signals into the classifier. In this algorithm, we used 16 channels that we have previously selected. With MEMD, we achieved 10 oscillations for 16 channel EEG recording from each stimulus. We created the feature vectors of these oscillations by combining some of these oscillations one by one and some of them in 2 or 3 combinations. How much success rate we get from which combination is explained in detail in the Results section. In addition, the classification results made with the feature vectors created without using the MEMD algorithm are given.

2.4.1 Empirical mode decomposition

Empirical mode decomposition is a transformation technique widely used in signal processing. It is an effective method for processing non-linear and nonstationary signals. The method expresses the time series signals in the form of the sum of a finite number of internal mode functions. Internal mode functions contain different frequency elements of the original signal. As a result of the empirical moderation, the signal diverged as internal mode functions and the remainder [21].When calculating the internal mode functions, the local limit values in the signal and the zero crossing number of the signal should be the same or the difference between them should be 1. At the same time, at any time t, the average value of the upper and lower envelopes formed by combining local maxima and local minima should be zero. The internal signal functions and the first signal can be obtained without any information loss and error when the remainder is collected [21,22].

The conditions given above are obtained empirically, so the band gap and number of IMFs to be obtained cannot be predicted. The average of the local maximum and local minimum points expressed in the second condition indicates the local average calculated pointwise throughout the envelope.

IMFs are functions that show different frequency bands and time scales of the original signal. IMFs are found by calculating the local averages of the signal, so that the signal can be analyzed locally.To effectively apply the EMD algorithm to a sample $x(t)$ signal, the so-called elimination process must be performed. The steps to be followed are as follows:

1. All local maximum points of $x(t)$ signal; M_k, (k = 1,2,3, ...) and all local minimum points M_i, (i = 1,2,3, ...) are determined.
2. By applying cubic spline interpolation, the local maximum and local minimum points are combined and the upper and lower envelopes of the signal are detected e_{max} (t) and e_{min} (t).
3. The average of the upper and lower envelopes that found is calculated $m(t) = (e_{max}(t) + e_{min}(t))/2$.
4. As a result, the function m (t) is subtracted from the signal itself. We obtained $h1(t) = x(t) - m1(t)$. If the found $h1(t)$ does not meet the conditions for being an IMF, the elimination process continues until the IMF conditions are met on $h1(t)$. Thus $h1(t)$ is considered as new data.

The remaining $m(t)$ signal in the last iteration process is called "residual signal". Usually, when the power of the residual signal obtained at the end of K iteration is below a certain threshold, the iterations stop, in this case $K \simeq \log_{10} 2N$ IMF, which is the N signal length. EMD algorithm is given in Figure 2.11.

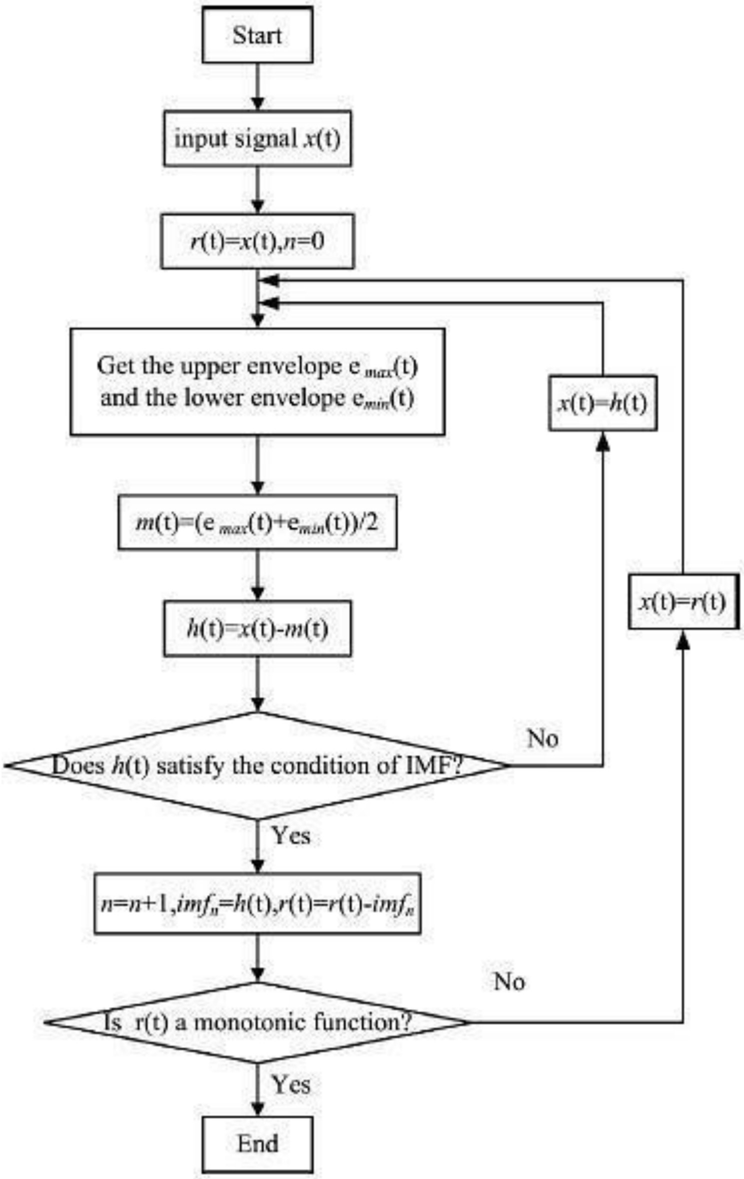


Figure 2.11 Flowchart of EMD algorithm.

The elimination process is applied to find IMFs and does not guarantee that it will find it in one go. The waveform that occurs after a single screening process can be asymmetrical, as a result of which the local average of the upper and lower envelopes can be calculated incorrectly. For this reason, the sieving process continues until the asymmetry is corrected and the IMF conditions are met. Another situation that may arise during the sieving process is the end effect resulting from curve fitting. The cubic curve fitting method used to obtain envelopes can have large oscillations when the end point is reached. As a result of the end effect, the low frequency components contained in the data suffer deterioration.

Although the sieving process contains disruptive effects in itself, it can accurately extract the time-scales of the amplitude and frequency modulated oscillation modes of the signal.

In order to ensure symmetry throughout the sieving process, uneven amplitudes are softened. But this can eliminate physically significant fluctuations. Therefore, screening should be done carefully. In the event that the sieving process is too long, the amplitude diversity hosted by the IMF decreases and approaches a fixed form. Attention should be paid to the stopping criterion to avoid physically insignificant IMFs. Standard deviation is used as a criterion for this purpose and must be calculated to avoid such negativities. If the normalized square difference S between the two consecutive sieving processes are less than the predetermined 0.3 value, the sieving is stopped.

2.4.2 Multivariate empirical mode decomposition

The multivariable EMD (MEMD), the standard EMD developed by Rehman and Mandic [23], is the extension of the two-variable and three-variable EMD. Standard EMD calculates the local average using the average of the up and down envelopes. However, the local average of a n-dimensional signal cannot be calculated directly and multidimensional envelopes are produced by reflecting the signal in variable spaces along different directions, then averaging these reflections to obtain the local average. For the set of direction vectors used to reflect the input of the multivariate signal, Low Discrepancy Hammersley Sequences were used to make semi uniform dots on high dimensional spheres [23].

However, it also requires attention to arrange a suitable direction vectors to project a signal determined in n-dimensional space. Local average estimation along multiple directions in an n-dimensional space can be seen as a calculation of the integral of all envelopes, and the accuracy of the calculation depends on the consistent selection of direction vectors, especially for a limited amount. Since direction vectors in n-dimensional spaces can be symbolized in the same way as points on the respective unit spheres, the issue of choosing an appropriate arrangement of direction vectors can be addressed as having a uniform sampling scheme on n spheres. When an appropriate arrangement of the direction vectors is obtained on the N sphere, the reflections of the signal are determined throughout that arrangement. The endpoints of the reflected signals compute interpolation to provide the requested multidimensional envelopes of the signal.

EMD is obtained by operating each EEG channel separately. This creates a problem called uniqueness problem. This situation reveals that the EEG signals and the random structure of noise due to the random nature of the IMFs obtained from different EEG channels have different numbers and different statistical properties. Thus, it is reflected according to different discriminations obtained from signals with similar statistics, and occurs at similar frequencies against different IMFs in univariate EMD. The multivariate extensions of the EMD have the advantage of processing multi-channel signals more conveniently upon processing each channel with a single EMD. The first IMFs of EEG channels in MEMD have similar frequency fluctuations, bandwidths or autocorrelation features. However, MEMD still shows some mode mixing sensitivity. Details of MEMD are summarized in the Algorithm below.

Algorithm Multivariate extension of EMD [23].

1. Choose a suitable pointset for sampling on an (n−1) sphere.

2. Calculate a projection, denoted by $\mathrm{p}\theta_k(\mathrm{t})\}_{\mathrm{t}=1}^{\mathrm{T}}$, of the input signal $\{\mathrm{v}(\mathrm{t})\}_{\mathrm{t}=1}^{\mathrm{T}}$ along the direction vector $\mathrm{x}\theta_k$, for all k (the whole set of direction vectors), giving $\mathrm{p}\theta_k(\mathrm{t})\}_{\mathrm{k}=1}^{\mathrm{K}}$ as the set of projections.

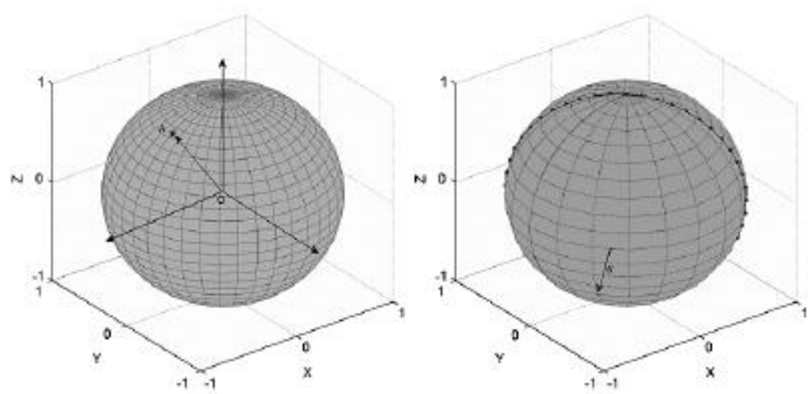
3. Find the time instants $\{t_i^{\theta_k}\}$corresponding to the maxima of the set of projected signals $\mathrm{p}\theta_k(\mathrm{t})\}_{\mathrm{k}=1}^{\mathrm{K}}$

4. Interpolate $[t_i^{\theta_k}$,$\mathrm{v}(t_i^{\theta_k})]$ to obtain multivariate envelope curves $e^{\theta k}(\mathrm{t})\}_{\mathrm{t}=1}^{\mathrm{T}}$.

5. For a set of K direction vectors, the mean $m(t)$ of the envelope curves is calculated as $\mathrm{m(t)} = \frac{1}{K}\sum_{k=1}^{K} e^{\theta k}(t)$.

6. Extract the ‘detail’ $d(t)$ using $d(t) = x(t) - m(t)$. If the ‘detail’ $d(t)$ fulfills the stoppage criterion for a multivariate IMF, apply the above procedure to $x(t) - d(t)$, otherwise apply it to $d(t)$.

Direction vectors are found through selected points on the sphere. Each point on the sphere becomes the end point of the vectors, the beginning of which is located in the center of the sphere. In Figure 2.12 (a), there is a sphere located in the 3D plane. The direction vector OA is shown on the sphere. More points are needed to find different direction vectors, so in Figure 2.12 (b), equidistant points are formed along the top of the sphere. The signal is rotated along the axis of rotation in the xy plane, and there are projections of the signal along the direction vectors and transferred to the z axis. The sample rotation axis u is shown in Figure 2.12(b).



(a) (b)

Figure 2.12 (a) O centered sphere in three-dimensional plane (b) rotation axis u.

There are different ways to create a large number of points to cover the sphere completely. In Figure 2.13 (a), there is a point distribution created using angular unit sampling method, which is a simple and useful method. These points in the 3-dimensional plane can be expressed as follows: in the n-dimensional sphere located in the (n + 1) dimensional plane $R = \sum_{j=1}^{n+1}(xi - Ci)^2$; In this equation, R represents the diameter of the n-dimensional sphere, and points on the x sphere represent the center of the sphere.

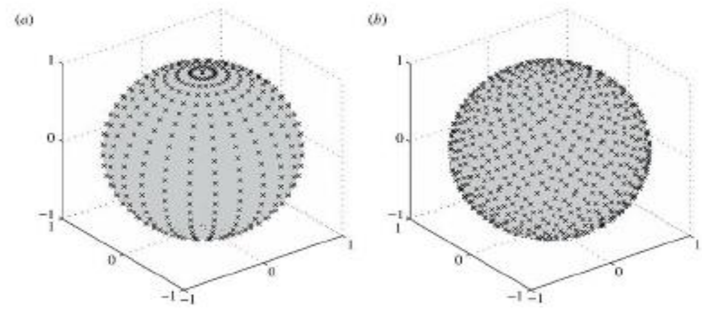


Figure 2.13 Point distributions to form direction vectors.

(a) Unifrom angular sampling method (b) low inconsistency Hammersley method.

Figure 2.13 (a) shows the points created by this method. Uniform distribution is impaired as the distribution of the points concentrates on the poles of the sphere. In order to obtain a more uniform distribution, the so-called 'low inconsistency point clusters' method was proposed by Rehman and Mandic [23]. Hammersley series was used to create low inconsistency points and a more uniform point distribution was obtained than uniform angular sampling method. Figure 2.13 (b) shows the points obtained on the two-dimensional sphere with the low inconsistency Hammersley sequence.

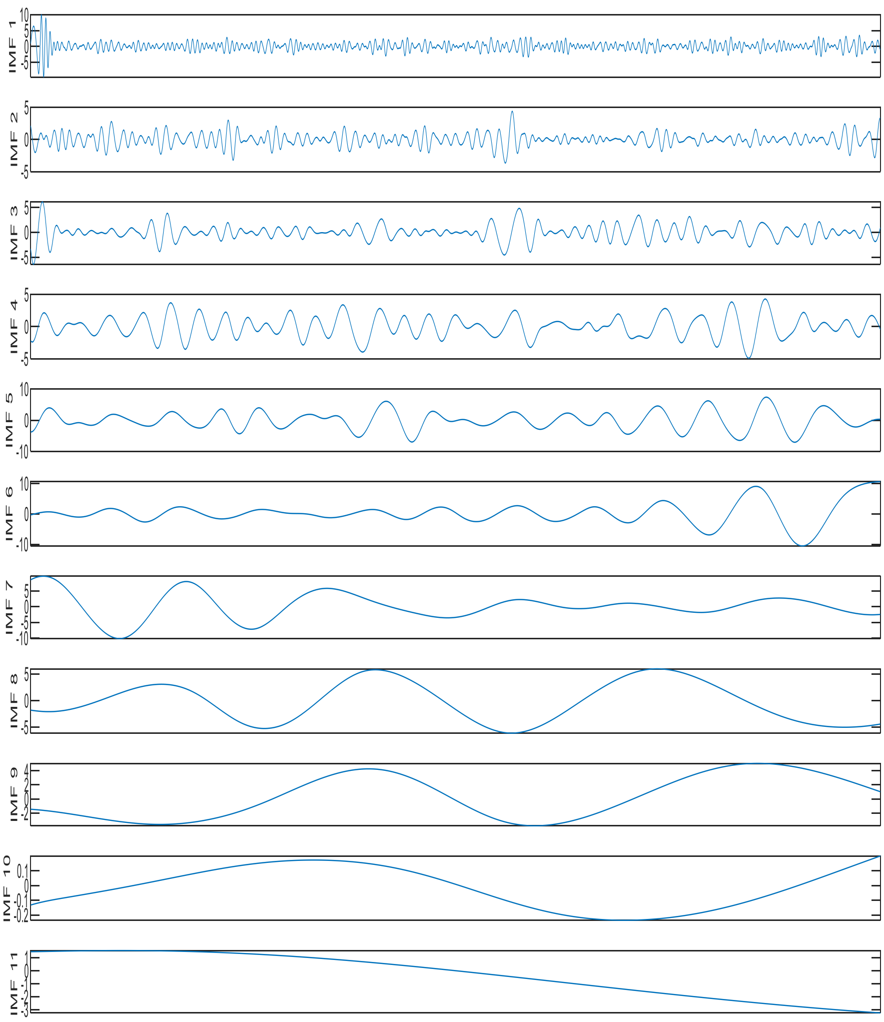


Figure 2.14 IMFs of one channel.

Figure 2.14 shows occured IMFs at the end of the EMD implementation. The first time series is the EEG signal. The decomposition consists of 11 IMF. EMD divides the signals into narrow-band components with decreasing frequency. For this reason, the first IMFs carry the high frequency components of the original signal and the frequencies of the IMFs decrease as their order increases. EMD implementation has been done to the segmented EEG signals. The IMFs are obtained from decomposition of multicomponent signal with the helping of EMD algorithm [24]. More than one IMF has been obtained after EMD implementation. A number of methods were used to find the significance of the IMF used in the emotion recognition.

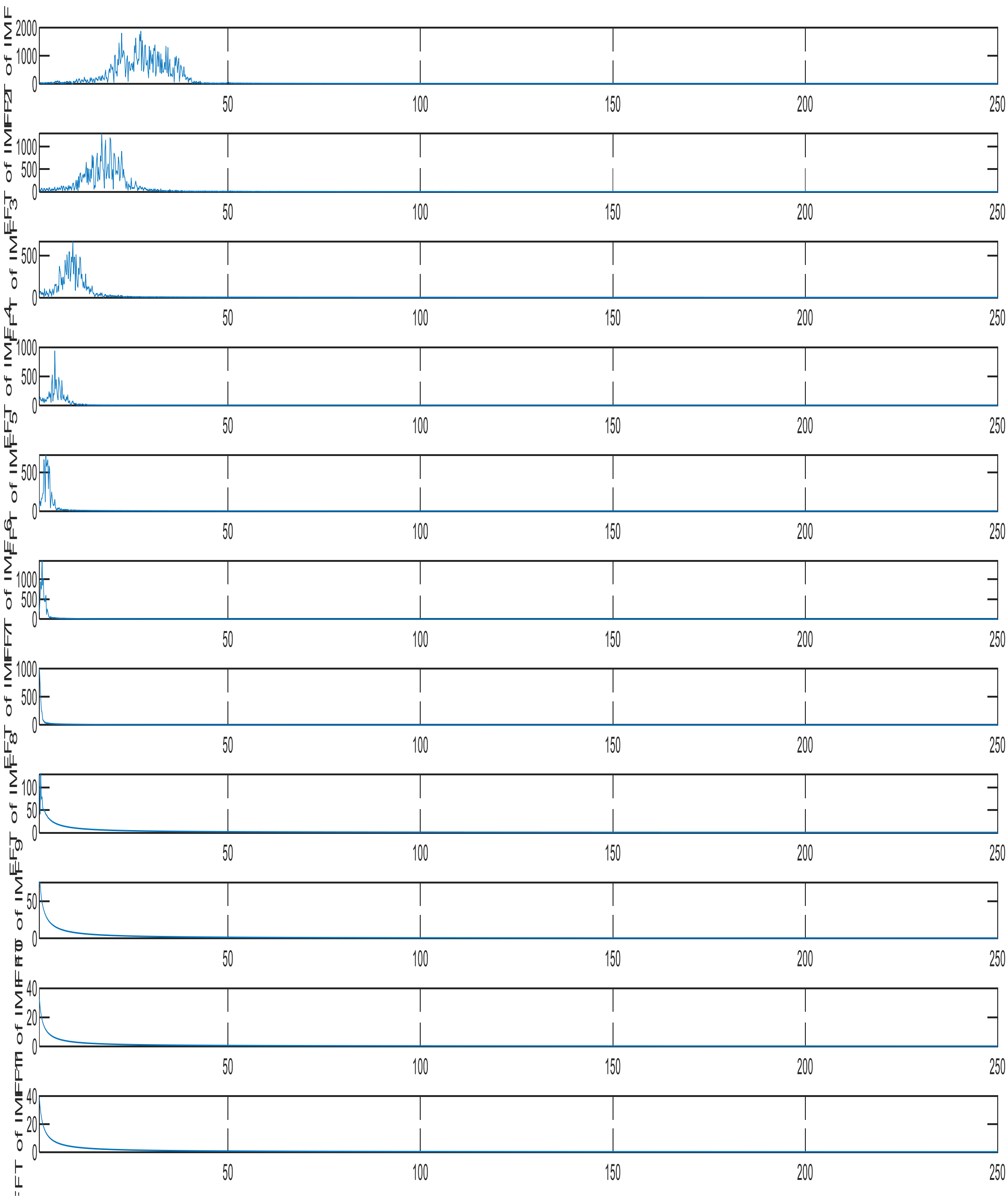


Figure 2.15 FFT of all IMFs between 0 and 250 Hz.

When we look at the FFTs of the IMFs one by one, we see in Figure 2.15 that the high frequency components are removed from the original signal from the first IMF to the last IMF.There are the lowest frequency components in the last IMF. Since our signal is sampled with 500 Hz, we can see all components up to 250 Hz in frequency domain. As we mentioned before, we have filtered the portion of our EEG signal up to the first 50 Hz. For this reason, we see that frequencies greater than 50 Hz are excluded from the signal since the first IMF. For better interpretation, the FFT of the first 50 Hz section is shown in Figure 2.16.

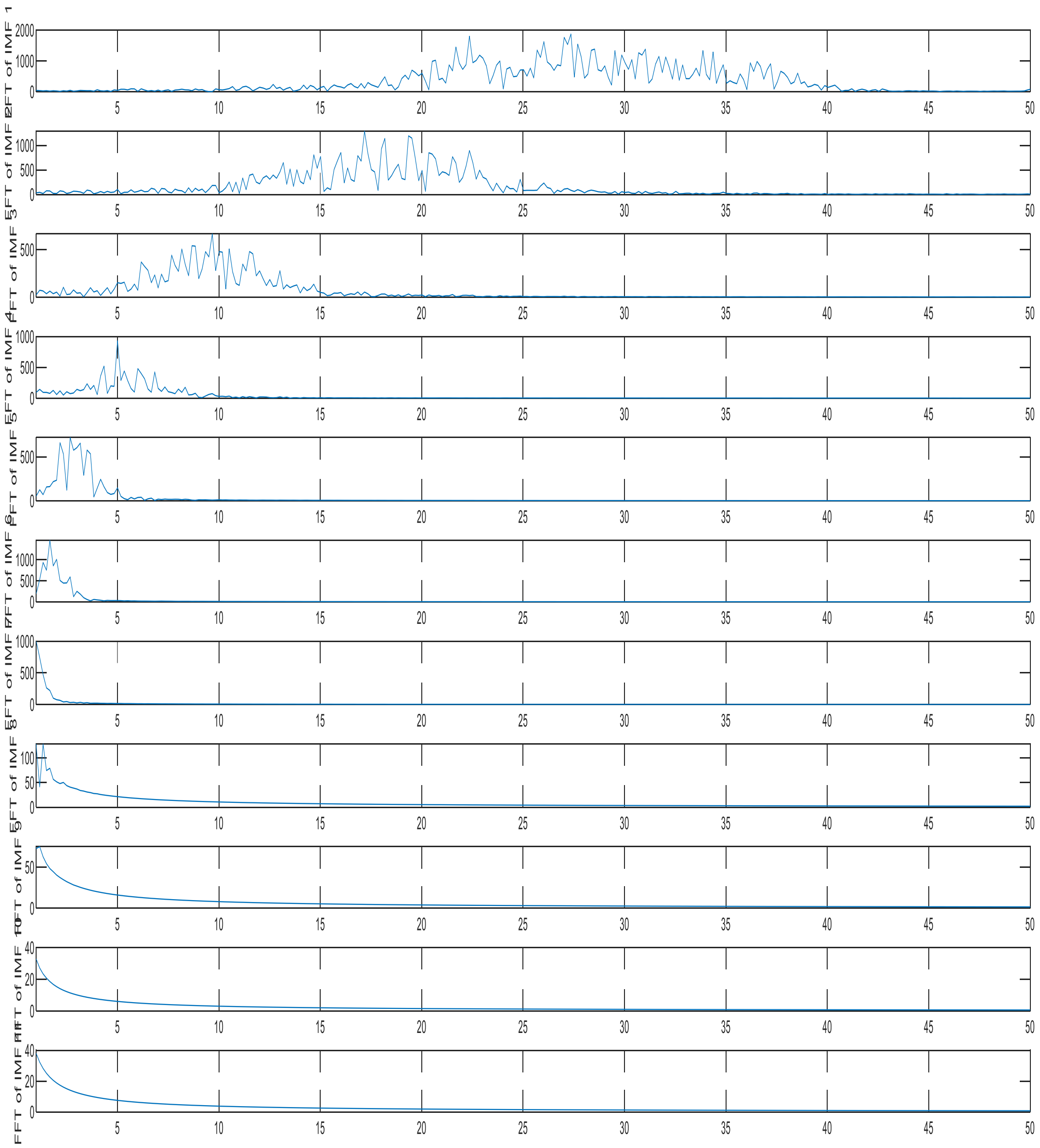


Figure 2.16 FFT of all IMFs between 0 and 50 Hz.

2.4.3 Short time Fourier transform and spectrogram

The traditional Fourier transform is insufficient in the frequency domain analysis of linear time varying signals and non-stationary random processes. Therefore, time information is lost. In other words, at what moments the frequency content of the signal occurs, it cannot be directly understood from the signal spectrum. To overcome this problem, Dennis Gabor (1946) put forward the idea of performing frequency domain analysis of each part by splitting the signal into small pieces by windowing method [25]. According to this method known as Short Time Fourier Transform (STFT), the time-varying signal examined is divided into small time intervals and the traditional Fourier transform is applied over the sample values remaining within this time interval. Thus, the frequency content of the signal is obtained depending on the time, which corresponds to the nonparametric time-frequency analysis of the signal. As stated earlier, the time-frequency analysis provides information about the change of frequency content of time-varying signals over time, and this information corresponds to a map or image expressed as a spectrogram reflecting the change of signal energy in the two-dimensional time-frequency space.

The analysis equation for the STFT of a $x(n)$ signal that changes over time is defined in Equation (2.1).

$$X(k, no) = \sum_{n=0}^{Np-1} x(n + no)p(n)e^{-j2\pi kn/Np} \tag{2.1}$$

Here, no indicates the scrolling moment of the window, $p(n)$ indicates the window function, and Np indicates the window size in terms of the number of samples.

Both time and frequency resolution must be at an acceptable level in order to make an accurate assessment of signal behavior from the time-frequency representation obtained with STFT. In other words, there is a balance between time and frequency resolution in STFT [25]. The width of the window function, which is the core of this transformation method, plays a decisive role in the time-frequency analysis of the signal. The width of the window to be used to determine the moment or moments of the changes occurring in the signal frequency content should be narrow enough and wide enough to reflect the frequency content as well as possible. In other words, if it is desired to obtain the time information correctly, there is an obligation to use a narrow window. However, using a narrow window increases the time resolution and causes the frequency resolution to decrease. If it is desired to reflect the frequency content as well as possible, then a large window should be used. Using wide windows causes frequency resolution to increase while time resolution decreases. In this case, it is not possible to determine exactly when the signal frequency content changes. The reason for this is expressed by the principle of uncertainty between time and frequency [25]. What is meant by uncertainty is that the conversion total calculated with Equation (2.1) covers only the window, not the entire signal. In other words, it is processed regionally without any frequency information coming from the general of the signal.

Window type is as important as window width in STFT. There are five types of window functions that are frequently used in applications for time-frequency analysis. Mathematical expressions of the width of main lobe and the value of approximation error and the value of the side lobes in dB for each window are listed in Table 2.3.

Table 2.3 Window types and properties.

Window Type	**Peak Sidelobe Amplitude (Relative,dB)**	**Approximate Width of Main Lobe**	**Peak Approximation Error, 20log(δ) (dB)**
Rectangular	-13	$\frac{4\pi}{M+1}$	-21
Bartlett	-25	$\frac{8\pi}{M}$	-25
Hann	-31	$\frac{8\pi}{M}$	-44
Hamming	-41	$\frac{8\pi}{M}$	-53
Blackman	-57	$\frac{12\pi}{M}$	-74

As can be seen from the values in Table 2.3, the frequency resolution of the rectangular window is higher than the other windows. On the other hand, the frequency resolution appears to be inversely proportional to the window length M. In other words, if the number of samples is low, it will not make sense to use a rectangular window since the frequency resolution will be low. In this case, the use of windows other than the rectangular window is used to prevent the time resolution being low. Because compared to other windows, the side lobes of the rectangular window have a big effect. This causes the sharpness of the starting and ending moments of the signal frequency content in the time-frequency plane to disappear.

The Fourier analysis gives the energy / power separation of each frequency component of the signal, but there is no time information, it does not give information about the time period of the frequency component [25, 26]. The main difference between STFT and Fourier transform is that STFT should be calculated in a certain time period. STFT is the process of applying Fourier transforms to each segment by dividing the $x(t)$ signal into windows $\big(w(t)\big)$ of certain sizes.The mathematical equation of the split signal in the continuous time domain is shown in Equation (2.2) [27].

$$x(t_c,t)=x(t).w(t-t_c) \quad (2.2)$$

In Equation (2.1), t_c is the center value of the symmetric window function. The Fourier transform applied to the split signal is given in Equation (2.3).

$$X(t_c,\omega_c)=\int_{-\infty}^{\infty} x(tc,t)e^{-j\omega t}dt \quad (2.3)$$

In Equation (2.3), w_c is the center frequency of the window. Energy intensity known as spectrogram defined as $|x(tc,\omega_c)|^2$. This equation calculates the energy of $x(t)$ around the $x(t_c,\omega_c)$. Equation (2.4) and Equation (2.5) are used for signal windowing and STFT operations respectively for the discrete time domain.

$$x(m,n)=x(n)w(n-m) \quad (2.4)$$

$$X(m,l)=\sum_{n=-\infty}^{n=\infty} x(n)w(n-m)e^{-j2\pi ln/m} \quad (2.5)$$

We used spectrogram images obtained by using the STFT method while obtaining an image from our EEG signal to be an input to deep learning. Spectrogram is found by taking absolute value of the STFT. This makes it possible to consider both time and frequency components when finding features. It carries the time and frequency information of the signal at the same time. Thus, we obtain pictures that contain information in time and frequency. We can give these pictures as an input to the deep learning classifier.

Once we realized that the fourth IMF was working best, we could now use this information when classifying with deep learning. With the help of spectrogram, we converted the 4th IMFs of all channels EEG recording that we previously found into time-frequency axis. Thus, we obtained a picture showing the time-frequency representation. While getting this picture, we sampled the signal with Hamming window. We set our window width to 32 samples. Number of overlapping was set as 31 samples. We tried to achieve maximum resolution in time by keeping our window size small in time domain. We chose the number of fft samples as 1280 samples. Thus, we tried to get a high resolution picture in the frequency domain. Since the matrix we obtained with the spectogram contains imaginary numbers, we first took its absolute value. Then; we rescaled it. We converted our newly formed matrix to unsigned 8 bits digits. Finally, we colorized our matrix using jet colarmap. Thus, we obtained a colored spectrogram picture free from scale and axis names. It is known in the literature that this procedure is performed for scalogram. We saved the images in 224x224x3 matrices according to the tag values in ".jpg" format to make them suitable for Googlenet. By sending these pictures to Googlenet architecture, we trained the system. As a result of the training, we achieved a success rate of up to 90%.

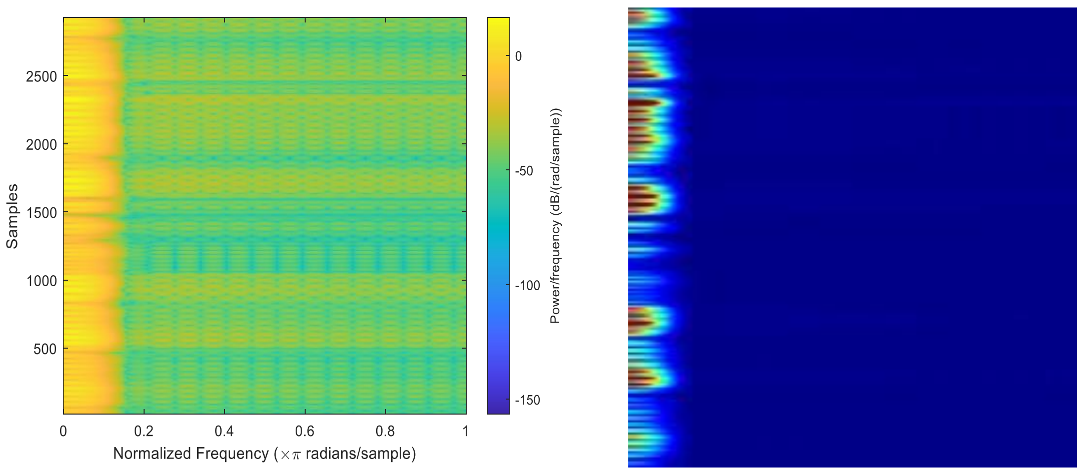


Figure 2.17 Sample of spectrogram from LVHA pictures for EEG channel 1.

The pictures were divided into two groups: 20 percent validation and 80 percent training. Spectogram and its image form are shown in Figure 2.17, 2.18, 2.19, 2.20.

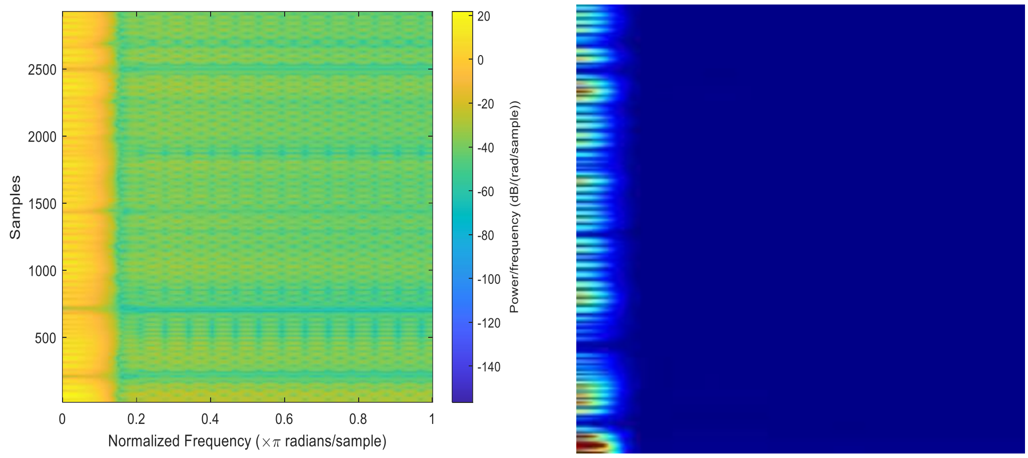


Figure 2.18 Sample of spectrogram from HVHA pictures for EEG channel 1.

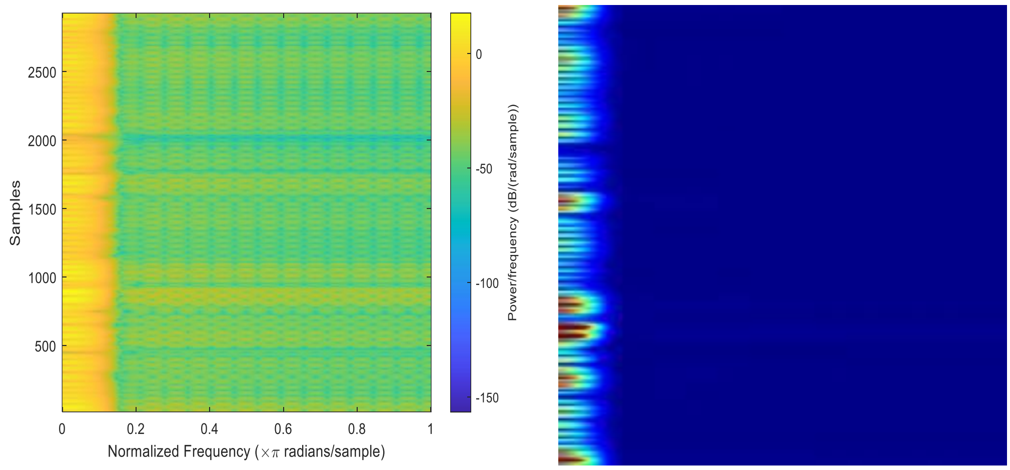


Figure 2.19 Sample of spectrogram from LVLA pictures for EEG channel 1.

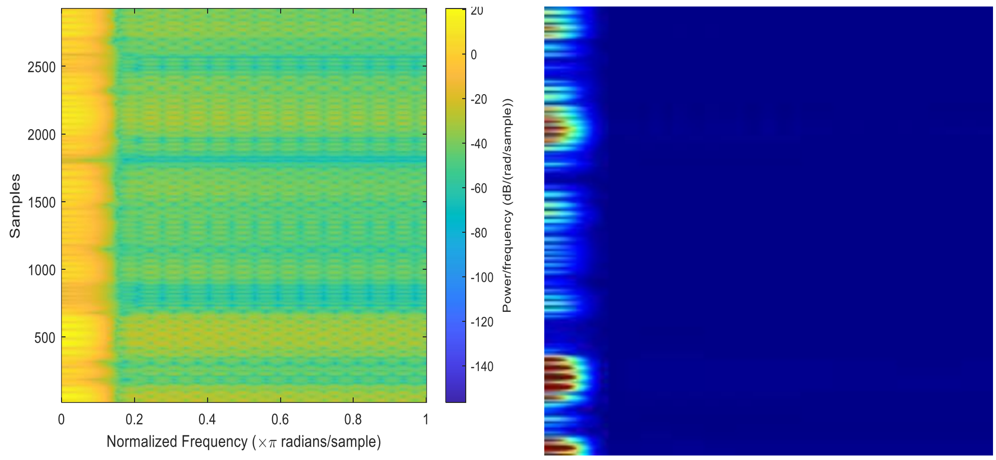


Figure 2.20 Sample of spectrogram from HVLA pictures for EEG channel 1.

2.5 Feature Extraction

The way to process EEG data obtained using the Brainvision BrainAmp EEG device was made as a scheme in Figure 2.21.

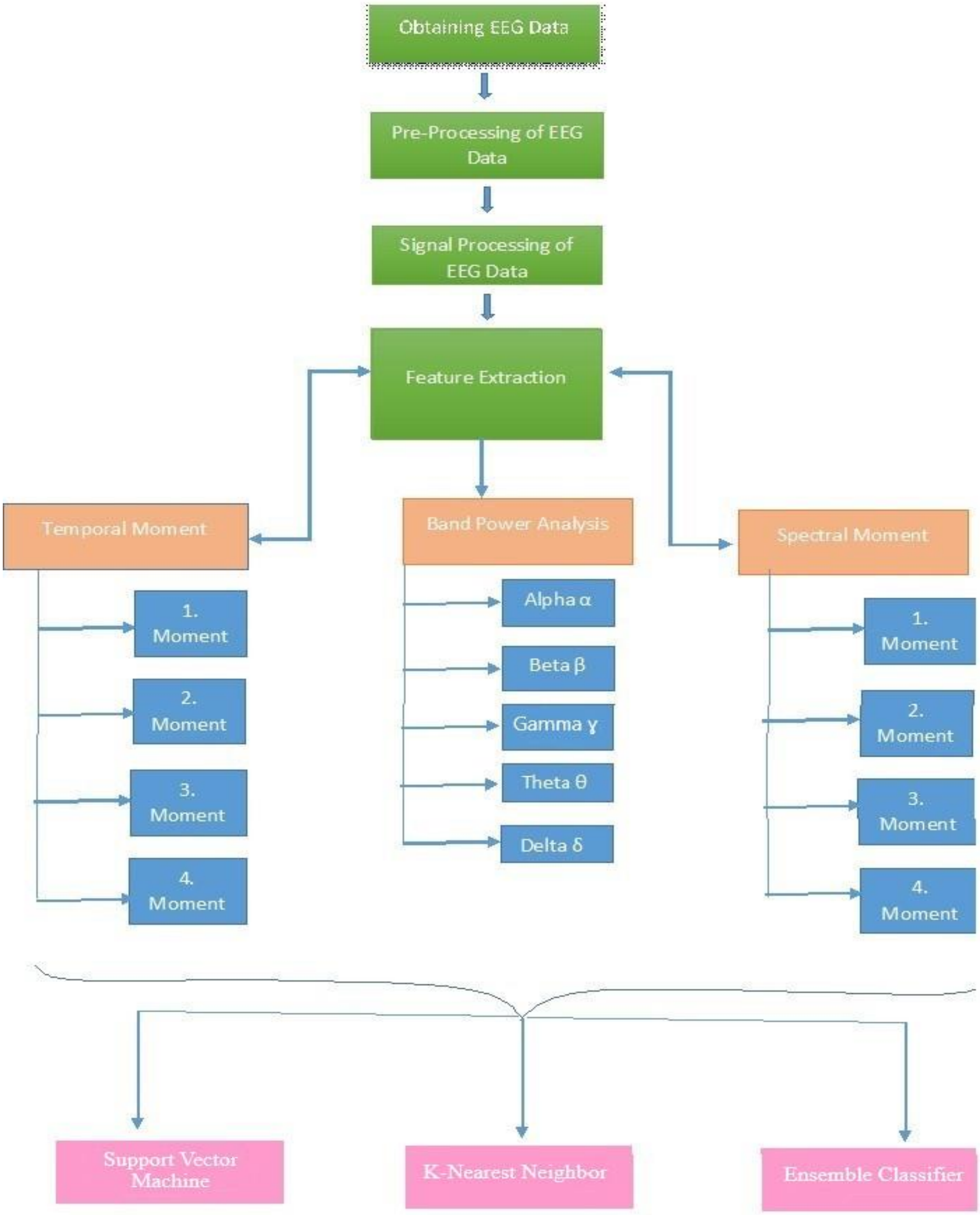


Figure 2.21 Flowchart of the project.

In order to create the feature vector of the filtered data, band power analysis, temporal and spectral moments are used. In general, temporal moment can refer to any property that changes over time [28]. However, within the scope of signal processing, the Fourier transform of the signal is taken to pass to the frequency domain and the spectral moments of the data are obtained [28, 29]. Considering that the signal coming from a single channel is expressed as '$\boldsymbol{x(n)}$', the temporal and spectral moments of the signal are obtained as shown in Equation (2.6) and Equation (2.7) [25,28].

$$\langle \mathrm{n}^{\mathrm{i}} \rangle = \sum_{\mathrm{n}=0}^{\mathrm{N}-1} \mathrm{n}^{\mathrm{i}} |\mathrm{x(n)}|^2 \tag{2.6}$$

$$\langle \omega_{\mathrm{k}}{}^{\mathrm{j}} \rangle = \sum_{\mathrm{k}=0}^{\mathrm{N}-1} \omega_{\mathrm{k}}{}^{\mathrm{j}} \frac{1}{N} |\mathrm{X(k)}|^2 \tag{2.7}$$

$$\omega_{\mathrm{k}} = \frac{2\pi}{\mathrm{N}} \mathrm{k} \tag{2.8}$$

The first 4 moments are obtained by replacing the i and j values used in the equation, respectively (i, j) = (1,2,3,4). In addition to these moments obtained, it is decomposed into EEG subbands. Power analysis of EEG signals separated according to the subband frequency range was performed. After these operations, a total of 8 moments, temporal and spectral, were obtained. In the band power analysis process, the power analysis of alpha, beta, gamma, theta, and delta bands, which are EEG subbands, were obtained and a total of 5 features were obtained. A total of 13 features were obtained with these two methods and an feature vector was created. This created feature vector is used in classification methods.

All feature vectors are normalized by taking their logarithms. A separate normalization formula is used for temporal and spectral moments. The logarithmic normalization method in Equation (2.10) is used for temporal moments, and the logarithmic normalization method in Equation (2.11) is used for spectral moments. For the normalization of band powers, only their logarithms were taken. Thus, all feature vectors are drawn to the same dynamic range value [30]. Features vary from 0 to 10. In this way, the importance of each features in the classification has been brought to the same level. EEG data from each channel in the equations is expressed as $\boldsymbol{x(n)}$. The Fourier transform of the $\boldsymbol{x(n)}$ signal is equal to $\boldsymbol{X(k)}$, shown in Equation 2.9. Time density is obtained by squaring the absolute value of $\boldsymbol{x(n)}$. In other words, the energy of the signal was obtained in time. Then, it is multiplied by the length of the data to find the 1st time moment and all the elements are summed under the total symbol. For the calculation of the subsequent moments, the square of the length, the cube of the length, and the 4th force of the length are used as multipliers. Thus, the first four temporal moments are obtained. These values obtained from each channel are written into an feature vector.

While spectral moments are found, first Fourier transform of EEG data is performed, $\boldsymbol{X}(\boldsymbol{k})$ is found and the square of the absolute value of $\boldsymbol{X}(\boldsymbol{k})$ is calculated. Then this value is divided by the length of the data. Thus, frequency density is obtained. This time, the power of the signal in the frequency domain has been found. Then, the values found are multiplied by the angular frequency given in the formula in Equation (2.8) to find the frequency moment and all its elements are summed under the total symbol. The square, cube, and 4th force of the angular frequency are used as multipliers to obtain subsequent frequency moments. Thus, the first four spectral moments are obtained. These values obtained from each channel are written into another feature vector.

The calculation of band powers comes from the power formula we use to find the frequency moments. The power of the signal, whose calculation is shown in Equation (2.12), is divided into bands. The power of each band was calculated by finding the number of samples in the band range. The signal collected with a sampling frequency of 500 Hz consists of approximately 3000 samples. It is assumed that the signal is represented by 6 samples of 1 Hz length. Therefore, the sum of the first 24 samples of the fast Fourier transform up to 4 Hz for the delta band are taken. While finding the power of theta band, values of fft samples from 25 to 48 were collected to be taken between 4-8 Hz. While finding the power of the alpha band, the values of the samples of fft from 49 to 78 were collected in order to be taken between 8 - 13 Hz. While finding the power of the beta band, the values of fft samples from 79 to 132 were collected to be taken between 13 - 22 Hz. While finding the power of the gamma band, the values of the samples of fft from 133 to 300 were collected to be taken between 22 - 50 Hz. As mentioned earlier in the preprocessing section, we eliminated frequencies greater than 50 Hz through a low pass filter. These operations were made for each channel. Feature vectors are created from 16 sample lengths of delta, theta, alpha, beta, gamma band powers obtained from a single stimulus. Considering that we have 60 stimuli, each feature length for a subject will be 960 samples. We created separate feature vectors for 20 female and 20 male subjects. Thus, the length of each feature vector is determined as 19200 samples. There are also features in 13 line vectors. Finally, our feature matrix consists of 13 rows and 19200 columns.

Before inserting it into the classifier, we add our output vector consisting of a line with label values to our feature matrix. Since we will first classify high arousal and low arousal, we will have 2 label values. All stimulus information coming from high arousal is given a '1' label and all stimulus information coming from low arousal is given a '2' label value. Similarly, we will have 2 label values when high valence and low valence classification will be made. It is written to the vector as the label value '1' for high valence and '2' for low valence. This vector is selected as output in the classifier.

$$\mathbf{X}(\mathbf{k}) = \sum_{-\infty}^{\infty} \mathbf{x}(\mathbf{n}) \boldsymbol{e}^{-\boldsymbol{j\omega t}} \tag{2.9}$$

$$\overline{\langle \mathbf{n}^{\mathbf{i}} \rangle} = \mathbf{log}\left(\frac{\langle \mathbf{n}^{\mathbf{i}} \rangle}{\boldsymbol{i}!}\right) \tag{2.10}$$

$$\overline{\langle \boldsymbol{\omega}_{\boldsymbol{k}}^{\boldsymbol{J}} \rangle} = \mathbf{log}\left(\frac{\langle \boldsymbol{\omega}_{\boldsymbol{k}}^{\boldsymbol{j}} \rangle}{\boldsymbol{j}!}\right) \tag{2.11}$$

$$\mathbf{S}(\mathbf{k}) = \sum \frac{|\mathbf{X}(\mathbf{k})|^2}{\mathbf{N}} \tag{2.12}$$

We have used skewness and kurtosis as time-domain based feature extraction.The Skewness is a measure of the asymmetry of the data in the sample mean [31]. If the skewness is negative, the data spread to the left. If the skewness is positive, the data is spread to the right. The skewness of an excellent symmetric distribution is zero.The kurtosis is a measure of the slope of the distribution [31]. The kurtosis of normal distribution is 3.

The Hjorth parameter is one of the ways to show the statistical property of a signal in a time period and has different types of parameters. To obtain the features in the study, activity, mobility and complexity were used from Hjorth parameters [32]. Also Hjorth parameters have low calculation cost.

Activity is defined as the variance of the time function, it can indicate the surface of power spectrum in frequency domain (σx).

Mobility parameter is defined as the ratio of the standard deviation of the first derivative of the x signal to the standard deviation of the signal, as expressed by Equation (2.13).

$$M = \frac{\sigma x\prime}{\sigma x} \tag{2.13}$$

Complexity is defined as the estimation of the band's width and it indicates how the shape of a signal is similar to a pure sine wave as expressed by Equation (2.14) [33].

$$c = \frac{Mx\prime}{Mx} = \frac{\sigma x\prime\prime/\sigma x\prime}{\sigma x\prime/\sigma x} \tag{2.14}$$

2.6 Classification

The EEG feature vector allows observation of what results from an emotional state. Generally, feature vector is used to resolve a classical possible emotional state. Numerous classification methods have been used in emotion prediction including K-nearest neighbor (K-NN), Support Vector Machine (SVM), Discriminant Analysis.

On the other hand, machine learning algorithms give humanity abilities so that machines can gain new knowledge from their own experience. This ability is called learning. Algorithms can improve their forecast performance with more examples. Machine learning algorithms are widely used in daily life, especially for medical purposes such as tumor detection, security purposes such as face recognition, computational finance, image processing such as motion detection, predictive maintenance for automotive, aviation, production, natural language processing.

Supervised machine learning algorithm uses the class known data as input and output for learning and creating a classification model. Then the model is used to estimate unknown data. Classification techniques assume the class of data points. For example, they determine if an e-mail is spam, and they detect neuro-phsiological cases using EEG data. Unsupervised learning reveals models that are covered without a tag response or intrinsic properties and includes only input values. It is often used for gene sequence analysis and market research [34]. The K-NN clustering method used first in the scope of the thesis is a non-hierarchical clustering system, at the same time, the ensemble classifier and support vector machine classifiers are used and performance parameters are examined [35].

2.6.1 K-nearest neighbor (K-NN)

K-NN learning method is nonparametric regression method. Sample to be analyzed is determined by looking at the class of the k neighbors in the nearest learning cluster.

As the example given in Figure 2.22, if the closest neighbor of the selected and classified sample for k = 3 is examined and the new sample should be included in the '∆' class where the '∆'samples are in the majority.

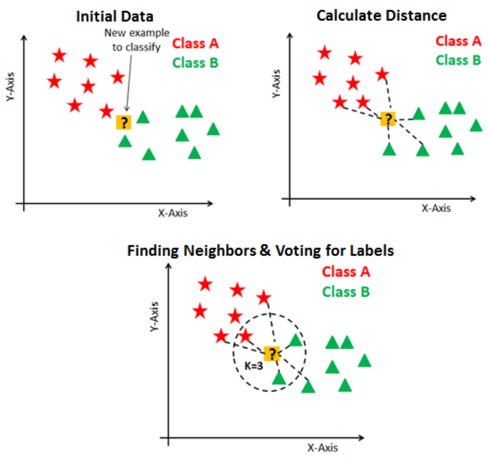


Figure 2.22 K-NN classifier [36].

Various distance functions are used when *N(x, y)* classification in N dimensional space. Euclid distance function is as stated in Equation (2.15). Minkowski distance function, another distance function frequently used in the studies in the literature, is as in Equation (2.17). In the case of Equation (2.17) inequality where *p=1*, the derived Manhattan distance function is in the form of Equation (2.16);

$$Euclidean\text{: } \sqrt{\sum_{i=1}^{k}(xi-yi)^2} \tag{2.15}$$

$$Manhattan\text{: } \sum_{i=1}^{k}|xi-yi| \tag{2.16}$$

$$Minkowski\text{: } \left(\sum_{i=1}^{k}(|xi-yi|)^q\right)^{1/q} \tag{2.17}$$

2.6.2 Support vector machine (SVM)

The main advantage of SVM is to solve the classification problem by transforming it into an optimization problem. In this way, the number of calculation processes will be reduced and a faster solution can be obtained compared to other techniques [37]. SVM is a member of linear binary class classifiers. Classification of objects in data sets is mainly based on tagging objects as -1 (first class) or +1 (other class). The labeling process varies depending on the specification of the research [38]. Another important function of SVM is to create an optimal hyperplane (linear decision boundary) that can distinguish differently labeled data points and maximize the distance between the support vectors [39,40].

The mathematical explanation of SVM can be summarized as follows. According to Equation (2.18); each entry point can be displayed as "xi" and the labels can be expressed as "f(x)", "w" represents the normal and weight vector of the hyper plane, and "b" represents the trend and constant value.

$$\mathbf{f(x) = wx+b} \tag{2.18}$$

The geometric drawing of the linear SVM model is shown in Figure 2.23 for classification of two classes and two dimensions.

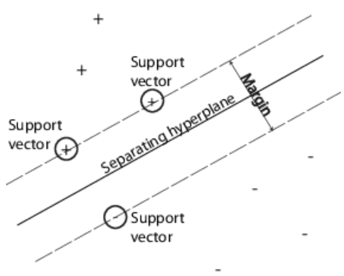


Figure 2.23 Geometric drawing of SVM.

The two parallel lines are called the boundary plane. The dark-colored plane passing through the middle of the boundary planes and separating both planes equally is expressed as a hyper plane [41,42]. SVM uses core functions such as polynomial and Gauss to create nonlinear adaptive data at higher dimensions where a linear decision boundary can be found.

2.6.3 Ensemble classifier

Ensemble models are widely used in many areas because they are considered to be more stable and, more importantly, better than individual classifiers [43]. It is also known to reduce model bias and variance [44,45].

Ensemble classifiers collect predictions of more than one basic model. Many empirical and theoretical evidence has shown that the combination of models increases predicted accuracy [46,47]. Ensemble classifier create basic models independently or in a dependent way. Retrained basic models are added to avoid mistakes from existing ensembles.

2.6.4 Artificial neural networks (ANN)

Artificial Neural Networks (ANN) is one of the important approaches used in machine learning. ANN consists of three layers (input layer, hidden layer, and output layer); each layer consists of adaptive processing units that are interconnected and called neurons. A neuron is a general calculation unit that receives n inputs and produces a single output. The parameter that distinguishes the output of neurons is their connection weights. Each neuron in a layer is connected to all neurons in a top layer with different weights. The input layer multiplies the incoming data with weights and transmits it to the hidden layer [48]. A transfer function is used to output from these multiplication results collected in the hidden layer. An example *(n:2)* ANN architecture is shown in Figure 2.24.

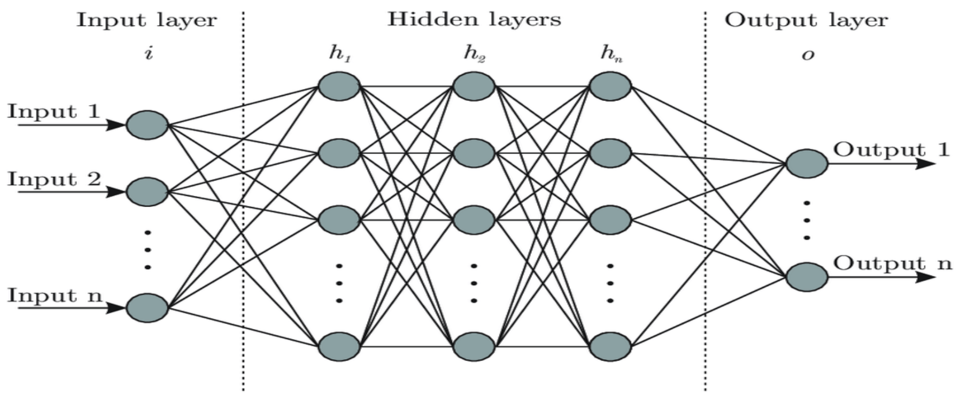


Figure 2.24 Artificial neural network architecture.

Each neuron in the ANN in the entry and latent layer is connected to the neurons of the next layer which is the classical neural network structure. In the ANN architecture, the number of hidden layers can be increased due to need. Calculating the number of parameters required to train a network containing L layers and N neurons in each layer can be a difficult problem. Working with a large stack of parameters can result in an untrained network, although not practical. Deep learning networks have been proposed to tackle such problems.

2.6.5 Deep learning networks

The main reason behind successful results in computerized pattern recognition, especially in recent years, is the adaptation of the statistical processing of visual data to the human brain. The deep learning networks developed from this adaptation regain the popularity lost by artificial intelligence and enable intensive studies on it. In this context, deep learning algorithms and CNN developed in parallel provide an active research area. The most important difference between classical machine learning and deep learning; the performance scale increases as the scale of data increases in deep learning. Deep learning considers concepts as nested hierarchy. It enables each concept to be defined in relation to simpler concepts. In this type of learning, it allows computational models consisting of multiple processing layers to learn the representations of data with multiple summarization levels. Due to this approach and performance of the method it appears to be used in many areas such as speech recognition, image recognition, and object detection [49,50]. Deep learning networks use the back propagation algorithm to learn the features they create within themselves. Techniques that are widely used as representatives of neural networks and probabilistic graph models, defined by the name of deep learning as a whole due to their similarities in education and architecture, provide modeling of more abstract and complex relationships instead of extracting linear features of the data.

2.6.6 Deep learning architectures

Today, deep learning architectures in data mining are used for different purposes in many fields. The most important reason for this is that these architectures can automatically obtain features from the related algorithms and data. Deep learning architectures can use large amounts of supervised or unsupervised data to extract the feature of complex demonstrations. In doing so, these architectures mimic the human brain's ability to observe, analyze, learn and make decisions to solve complex problems [51].

On the other hand, models based on shallow learning architectures such as decision trees, SVM may be inadequate while trying to extract useful information from complex structures. In contrast, deep learning architectures have the ability to produce relationships independent of the environmental change (neighborhood connections) in the data. When performing these functions, complex representations of the data are obtained. The purpose of artificial intelligence is to obtain unchanging features independent of human knowledge. With the deep learning architectures developed for this purpose, they ensure that the characteristics corresponding to the data are obtained in a lower dimensional form.

Deep learning networks are based on sequential layer architecture. In these architectures, data (such as pixels in the image) begins processing from the first layer, and the output of each layer is applied as input to the next layer. A pattern (eg image) that will be processed in deep learning layers is converted into pixel values form in the first place. These pixel values are given to the network as input and come to the first layer. In the first layer, the properties of the edge structures of the particular direction and location of the image are learned. In the second layer, the architecture that learns certain edge structures detects patterns by combining them and transmits them to the third layer. In the third layer, larger patterns corresponding to similar parts of the pattern are combined. In the later layers of architecture, the pattern is perceived as a combination of the patterns obtained [52]. In these layers in deep network architectures, nonlinear transformation methods are used to extract the distinguishing factors in the data.

2.6.7 Convolutional neural network

Convolutional Neural Networks (CNN) has been used by researchers in image and video processing, image classification, object detection and segmentation, which are the largest areas of deep learning research, especially in recent years. The fact that they do not need manually designed filters to extract features from the image makes the evolutionary neural networks useful in image processing, and is also used on other signals containing sequential and interrelated data outside the image [53]. Evolutionary neural networks, such as Artificial Neural Networks (ANN), also have neurons, weights, loss function and other parameters. CNNs consist of three main layers called the convolution layer, the pooling layer and the fully bonded layer [54].

The image, which is the input data of the CNN architecture, first processes in the convolution layer to create a feature map; in the next step, the size of these feature maps becomes smaller in the pooling layer. The fully attached layer at the end is responsible for image classification. CNN can be viewed from a functional perspective as a structure that aims to reduce the cost of calculating the features automatically detected by the model and to increase the classification functionality. The GoogLeNet architecture, which won the ImageNet competition held in 2014, has 22 layers as seen in Figure 2.25. The error rate for pattern recognition was 5.7%. In GoogLeNet, in order to overcome the cost of too much computing and memory, it used parallel interconnected modules [55].

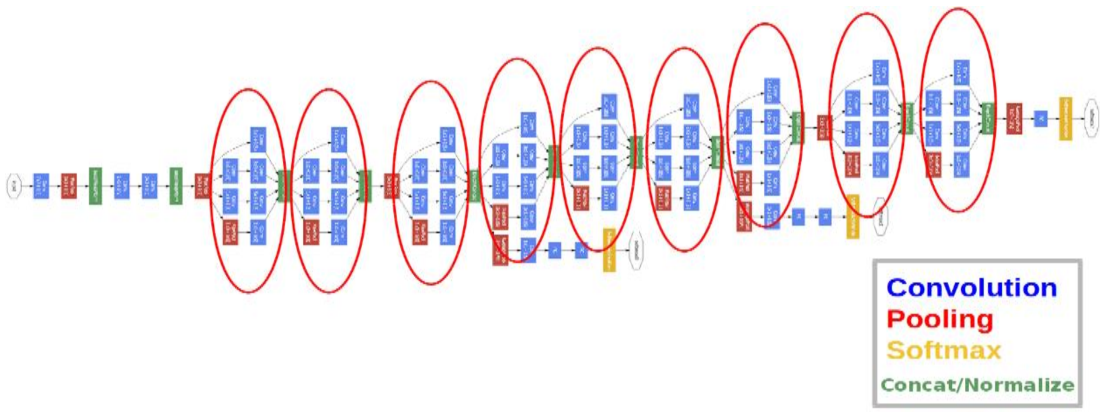


Figure 2.25 GoogLeNet structure [56].

We are installing Googlenet architecture which is a convolutional neural network consisting of 144 layers to the system as seen in Figure 2.26. We throw the dropout layer in the last layer and replace it with the softmax layer.

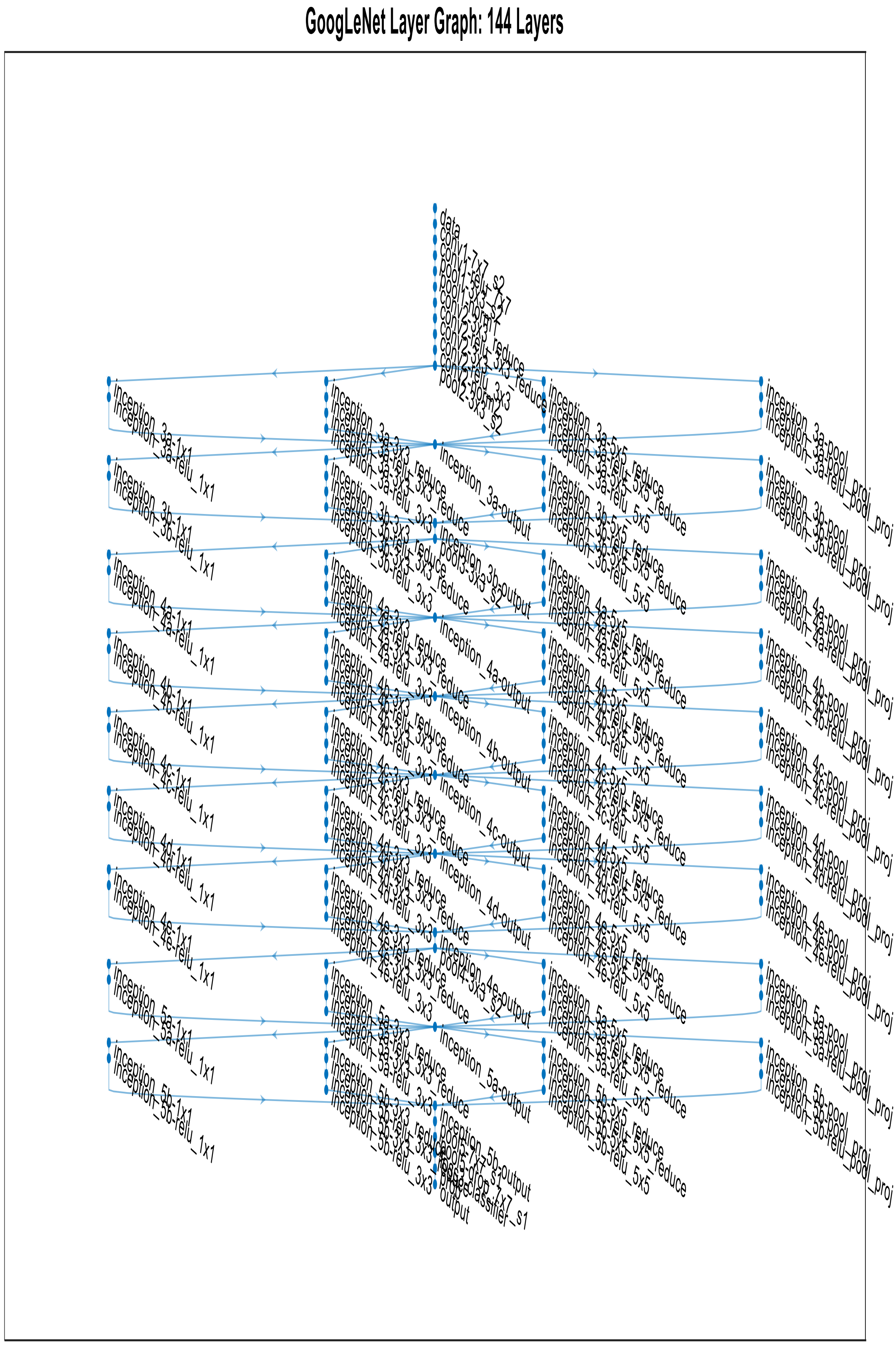


Figure 2.26 Googlenet architecture.

2.6.7.1 Convolution layer

This layer can be defined as a collection of feature extractor filters that slide over images. Each filter represents a specific matrix that performs the convolution process on the input image. Each filter optimizes the different features simultaneously, extracting them from the input image of the layer and creating the feature maps. The stride parameter determines the amount of shifting of the filter matrix above the input image. The output of the convolution layer is called feature map.

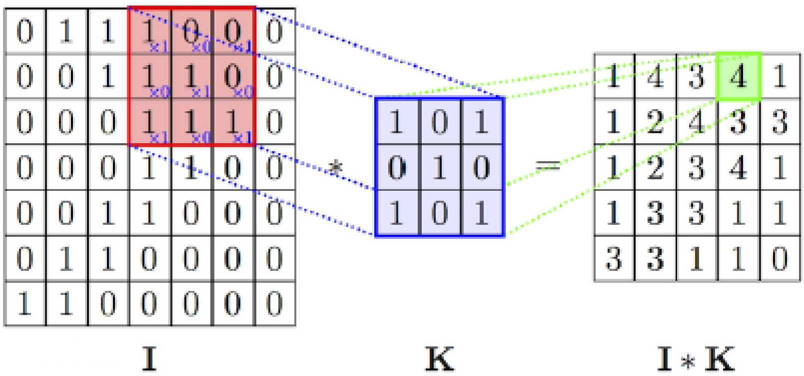


Figure 2.27 Filtering process.

There are filters on this layer, the filter is slid over the matrix and multiplied one-on-one, and the results are collected and replaced in the new matrix. Here each filter corresponds to a neuron. The values in the filter are weights. Pads consisting of spaces can also be added to the outer wall if desired to prevent data loss in the edge cells. The number of scroll cells during filter shift is expressed as Stride. The following calculation is used for the output matrix size.

ConvoldedWith= ((MatrixWidth-FilterWidth+2*Pad)/StrideW) +1

ConvoldedHeight= ((MatrixHeight-FilterHeight+2*Pad)/StrideH) +1

In the example shown in Figure 2.27, Input: 7, Filter: 3, Pad: 0, Stride: 1. When the formula is applied, ((7-3 + 2 * 0) / 1) +1 = 5-pixel output result is obtained.

2.6.7.2 Pooling layer

It is a common approach in deep learning architectures to add pooling layers between convolution layers. The pooling layer prevents overfitting by reducing parameters and computing load in the network structure. The purpose of this layer is to reduce the size of the feature maps to deal with the complexity of the image. The most common of the pooling layer approaches is the maximum pooling layer. Maximum pooling can be defined as a window that moves across a two-dimensional input field. The number with the greatest value in the window in question is taken as the output of the layer. The maximum and average pooling layer is shown in Figure 2.28.

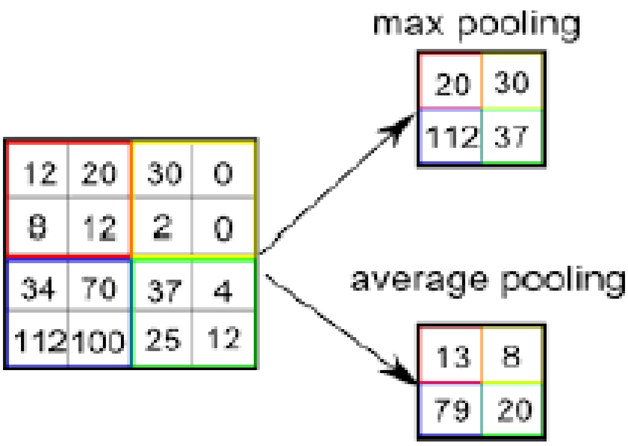


Figure 2.28 A sample application for maximum and average pooling.

2.6.7.3 Activation layer

The output is converted to a non-linear value after weights of the neurons are calculated using matrix multiplication in the hidden layer. As deep learning methods are also used in solving nonlinear problems, activation function usage is appropriate. There are several functions as Sigmoid, Tanh, and ReLU which are shown in Figure 2.29.

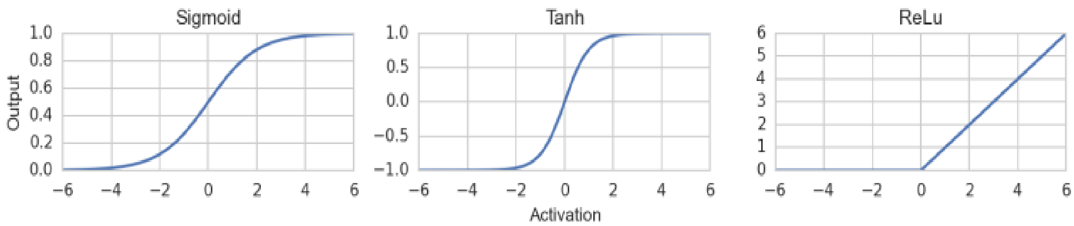


Figure 2.29 Graphical representation of Sigmoid, Tanh and ReLU functions.

In fact, the ReLU (Rectified Linear Activation Unit) activation function is designed for deep learning models. It enhances the distinctive feature of the layer output, thus allowing it to be generalized or adapted with various data. The use of ReLU as an activation function of CNN convolution layers is very common in the literature [54]. Although the tanh function is widely used as an activation function in machine learning models, training times are reduced several times using ReLU [57]. Equation (2.19) shows the mathematical expansion of the ReLU function.

$$f(x) = \{x \quad if\ x > 0, \quad 0 \quad otherwise\} \tag{2.19}$$

2.6.7.4 Fully connected layer

The last layer in the convolutional neural network structure is the fully connected layer. Neurons in the fully connected layer, as in Artificial Neural Networks, are fully connected to all activations in the previous layer. The task of this layer in the structure is to classify the extracted features. In fact, this layer is equal to the MLP (Multilayer Perceptron) classifier, which uses the Softmax activation function in the output layer. The difference between fully bonded and convolution layers is that the neurons in the convolution layer only bind to a local region on the input data, and most of the neurons in this layer share parameters. However, since the neurons in both layers calculate the dot product, their functional forms are the same. Since all neurons in this layer are interconnected, they are named as fully connected layers.

Here, instead of the ReLU activation function, which is more preferred in the previous layers, more tanh or sigmoid functions are preferred.

The Dilution Method that is called Dropout is used in this layer, by eliminating the information in the selected nodes, the system is prevented from over-fitting and training performance is increased. Usually, forgetting is done among the nodes below certain threshold values, but random forgetting method is also used regardless of the threshold value [58]. Dropout method is shown in Figure 2.30.

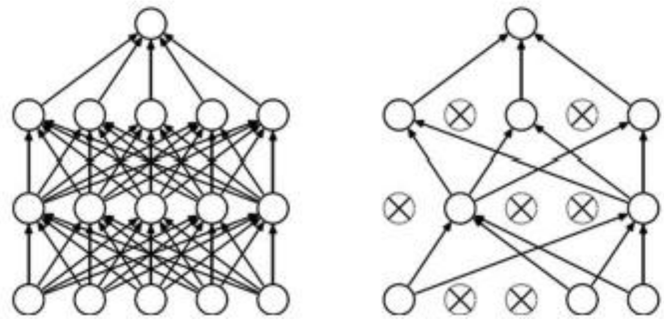


Figure 2.30 DropOut representation.

2.6.7.5 Softmax

Softmax classifier is the generalization of binary logistic regression classifier to multiple classes. In short, Softmax calculates the probability distribution of a statistical event 'n' according to different events. In general, this function calculates the probabilities of each target class on all possible target classes. The probabilities calculated later assist in determining the target class for the given inputs [59]. Equation (2.20) calculates the exponential (force of the number e) value of the given input value and the sum of its exponential values in all inputs. As a result, the ratio of these two values is the output of the softmax function. Here p is a vector of each input data indicated by the i indexed element. Also, i is the index of the output unit, i.e. i = 1, 2, ... K.

$$f(p)_i = \frac{e^{p_i}}{\sum_{k=1}^{k} e^{p_k}} \tag{2.20}$$

2.6.8 Performance management

The confusion matrix is a matrix consisting of rows of prediction classes and columns of real classes (y) as seen in Figure 2.31. Values such as accuracy, true positive rate (TPR), false positive rate (FPR), false negative rate (FNR), true negative rate (TNR) can be calculated on this chart. It is used to see the big picture when the class distribution is not equal.

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Figure 2.31 Confusion matrix.

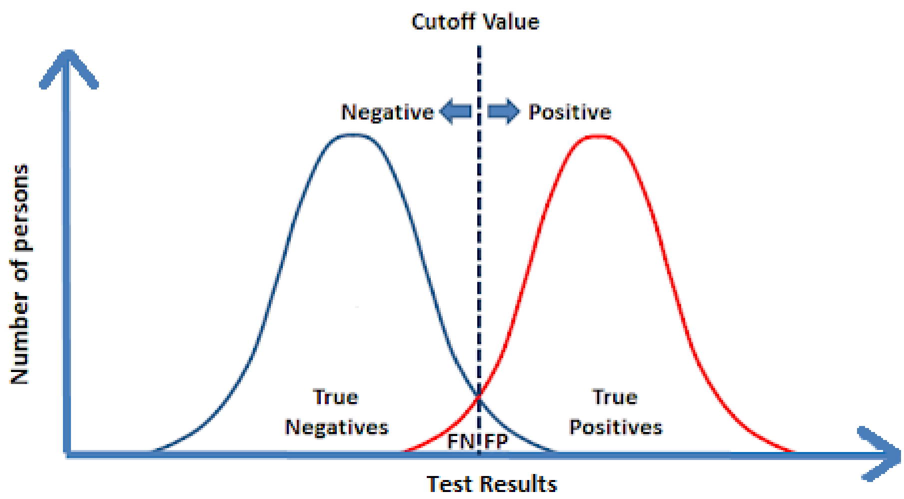


Figure 2.32 Metric expressions according to class distributions.

• The situation where (y = 1) & (prediction = 1) represents true positive (TP),
• The situation where (y = 0) & (prediction = 1) represents false positive (FP),
• The situation where (y = 1) & (prediction = 0) represents false negative (FN),
• The situation where (y = 0) & (prediction = 0) represents true negative (TN) as seen in Figure 2.32.

The Receiver Operating Characteristic (ROC) curve is a curve calculated for different threshold values, which are the x-axis FPR, the y-axis TPR and used in class separation. The area under the ROC curve (AUC), on the other hand, is a metric to show how far classes are separated from each other as seen in Figure 2.33. They are used in classification problems.

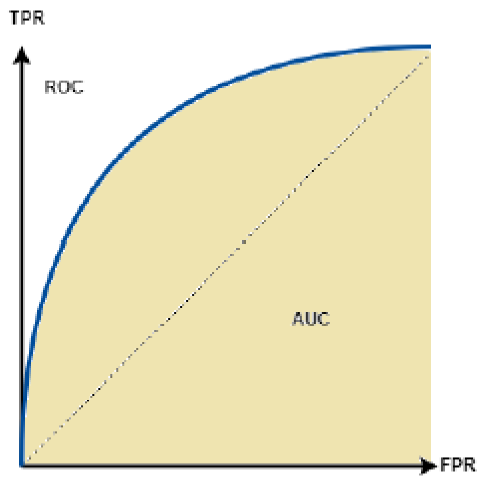
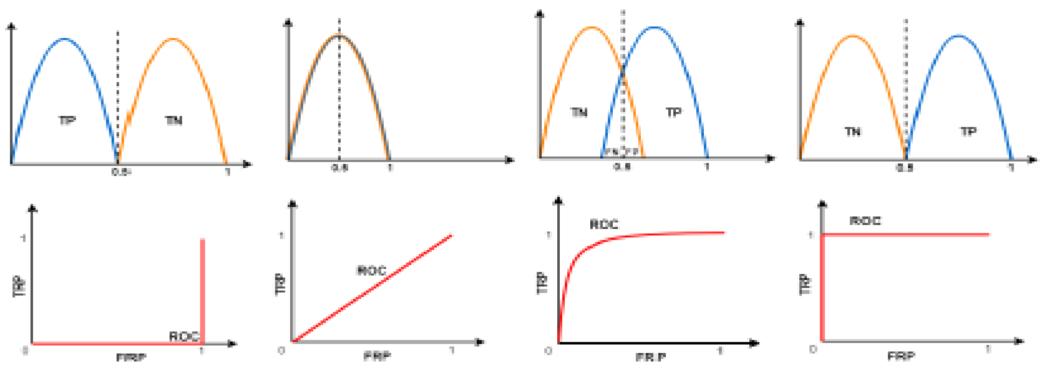


Figure 2.33 ROC curve and AUC.

$$PRECISION\!:TPR = \frac{TP}{TP+FN} \tag{2.20}$$

$$FPR = 1 - \frac{TN}{TN+FP} \tag{2.21}$$

AUC score of 1 indicates that classes are completely separated, that is, all 1's are 1 and all 0's are estimated as 0. The fact that 0 is also an indication that all classes are completely parsed, but all 1's are estimated to be 0 and all 0's are estimated to be 1. If it is 0.5, it is the worst case and it is an indication that the model cannot make any class distinctions as seen in Figure 2.34.



(a) AUC=0 (b) AUC=0.5 (c) 0.5<AUC<1 d) AUC=1

Figure 2.34 AUC scores according to the separation of distributions.

3. RESULTS

3.1 Classification Before Using MEMD

Classification results were obtained by using SVM, KNN and ensemble classifier methods in Matlab classification learner toolbox. Firstly, before using MEMD algorithm, classification was done using all but existing features except skewness, kurtosis and hjorth parameters. Classification results are given in Table 3.1.

Table 3.1 Classification results.

Gender	Emotion Parameters	SVM (Fine Gaussian)	KNN (Weighted)	Ensemble (Subspace KNN)
WOMEN	Arousal (High/Low)	%67,2	**%69,1**	%68,5
	Valence (High/Low)	%67,3	%69,2	**%69,4**
MEN	Arousal (High/Low)	%64,9	**%65,1**	%64,4
	Valence (High/Low)	**%66,1**	%65,9	%64,6

When classification is made separately for female and male subjects, the highest success rate in the arousal scale was obtained by the KNN classification method. When we look at the valence scale, the highest success rate was achieved with the ensemble classifier method for female subjects and the SVM classification method for male subjects. A higher success rate was obtained from female subjects than male subjects. It was observed that the KNN classifier was more successful in the arousal scale. The highest success rate was obtained in the Valence scale for female subjects.

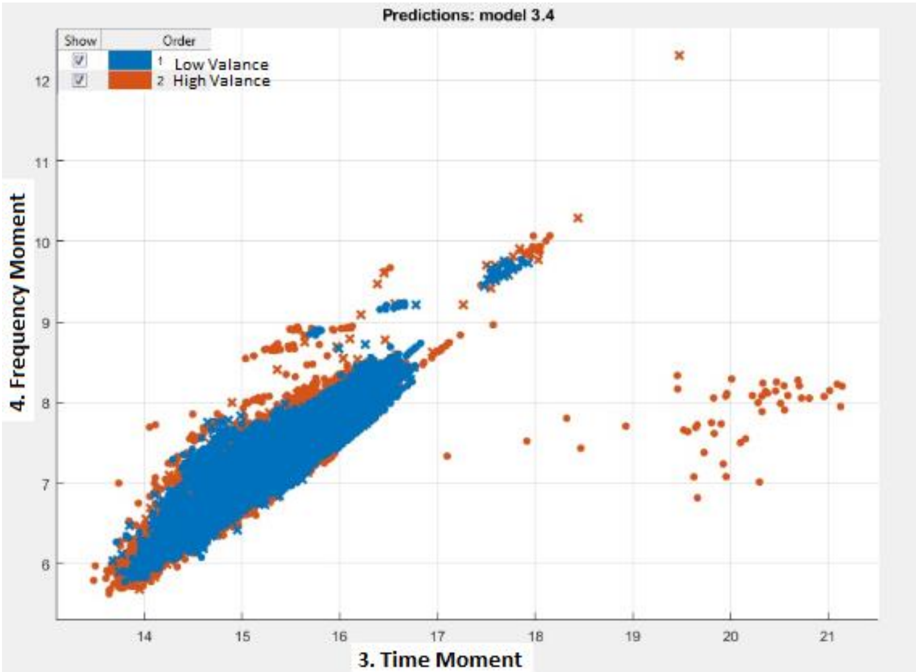


Figure 3.1 Scatter plot for valence classification.

One of the scatter plots made for Valence classification is shown in Figure 3.1. High and low valence values overlap in the scatter drawing of the Fourth frequency moment according to the third time moment. Therefore, the success of classification is not very high.

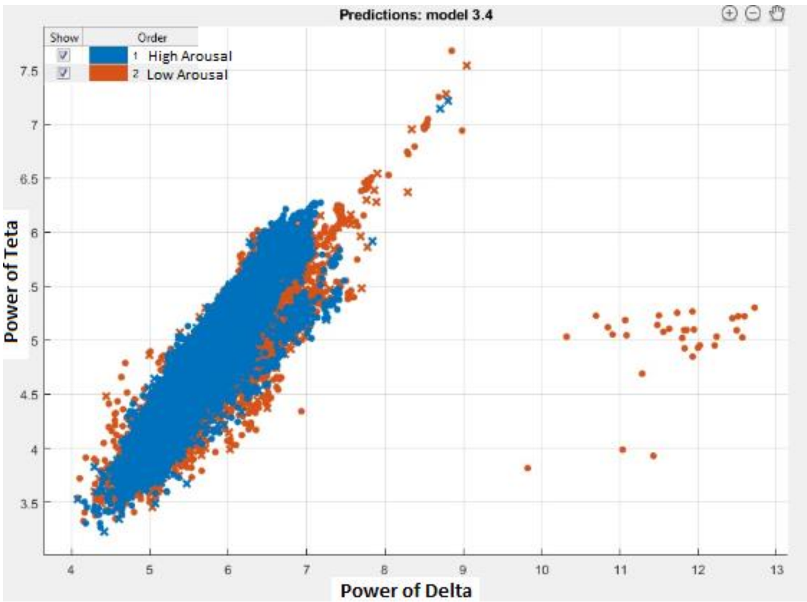


Figure 3.2 Scatter plot for arousal classification.

One of the scatter plots made for arousal classification is shown in Figure 3.2. High and low arousal values overlap in the scatter drawing of theta band power according to Delta band power. Therefore, the success of classification is not very high.

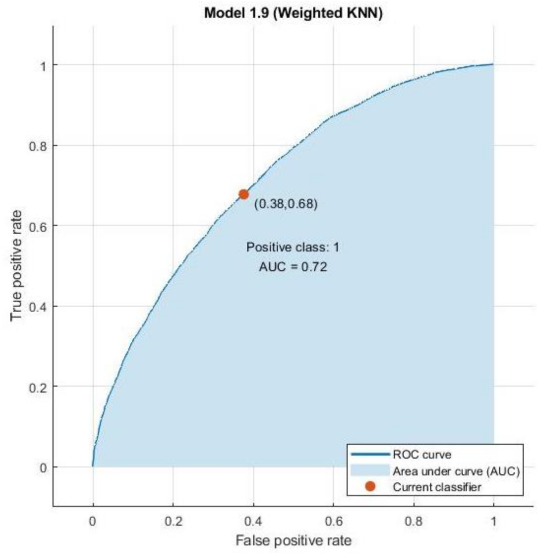


Figure 3.3 ROC curve of arousal for men.

The ROC curve obtained as a result of the classification made on the stimulation scale of male subjects is as shown in Figure 3.3.

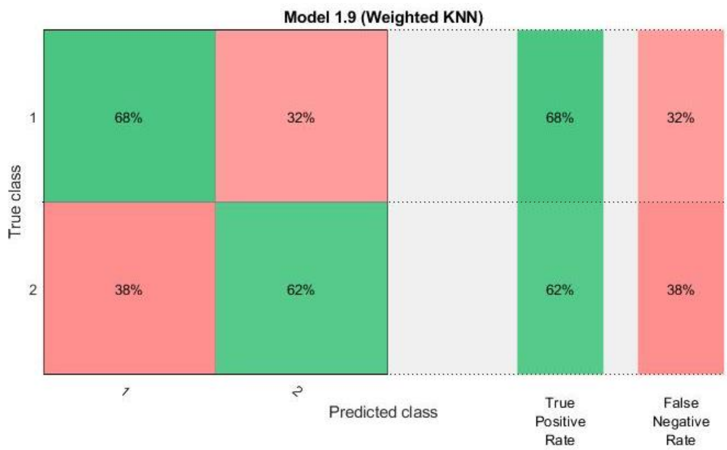


Figure 3.4 KNN classification of arousal for men.

The complexity matrix of arousal with using KNN for 20 men is given in Figure 3.4. The highest success in women and men subjects was obtained by classification of the stimulation parameter with K-NN classifier.

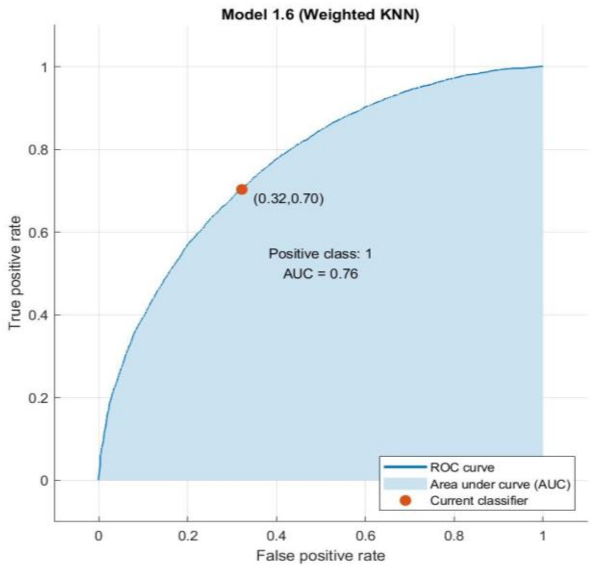


Figure 3.5 ROC curve of arousal for women.

The ROC curve obtained as a result of the classification made on the arousal scale of female subjects is as shown in Figure 3.5. The complexity matrix of arousal with using KNN for 20 women is given in Figure 3.6.

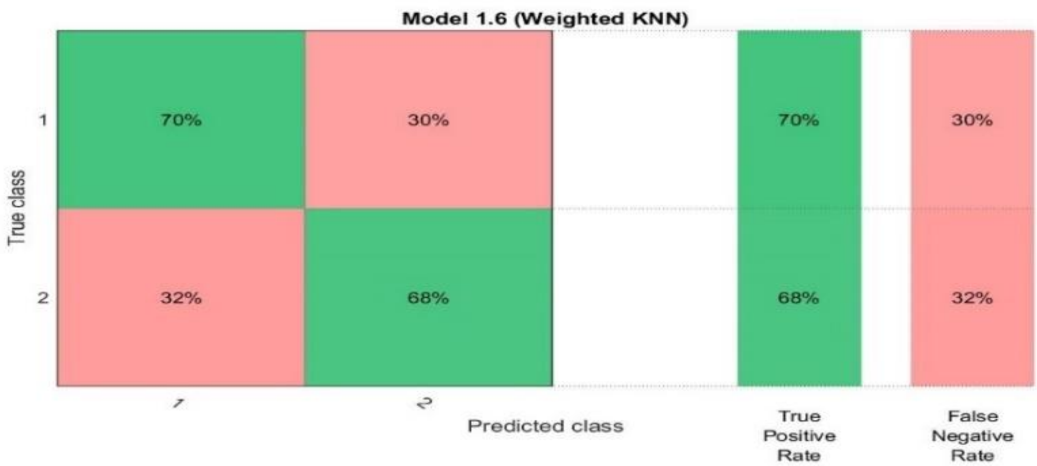


Figure 3.6 KNN classification of arousal for women.

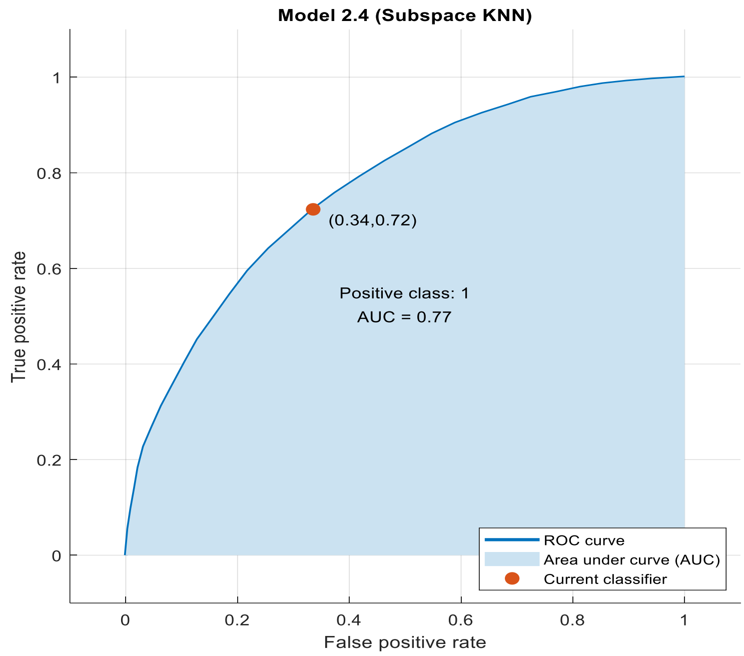


Figure 3.7 ROC curve of valence for women.

The ROC curve obtained as a result of the classification made on the valence scale of female subjects is as shown in Figure 3.7. The complexity matrix of valence with using subspace KNN for 20 women is given in Figure 3.8.

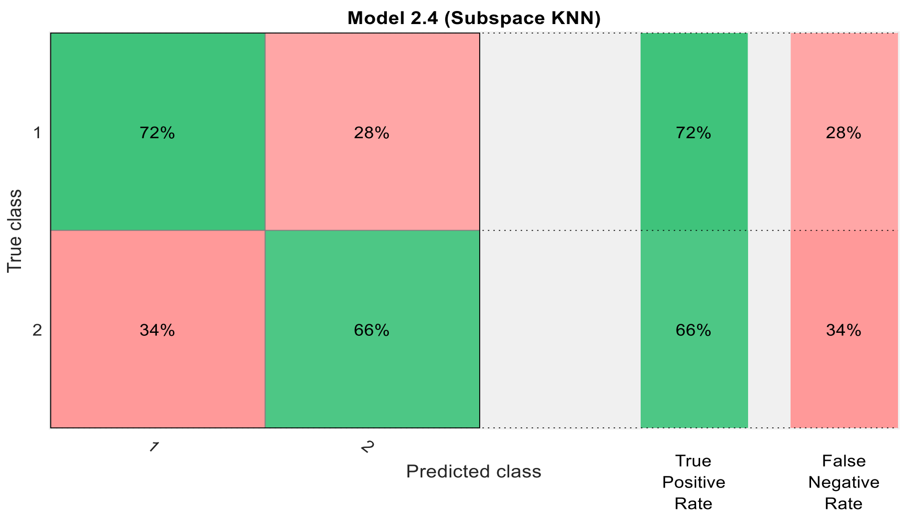


Figure 3.8 Subspace KNN classification of valence for women.

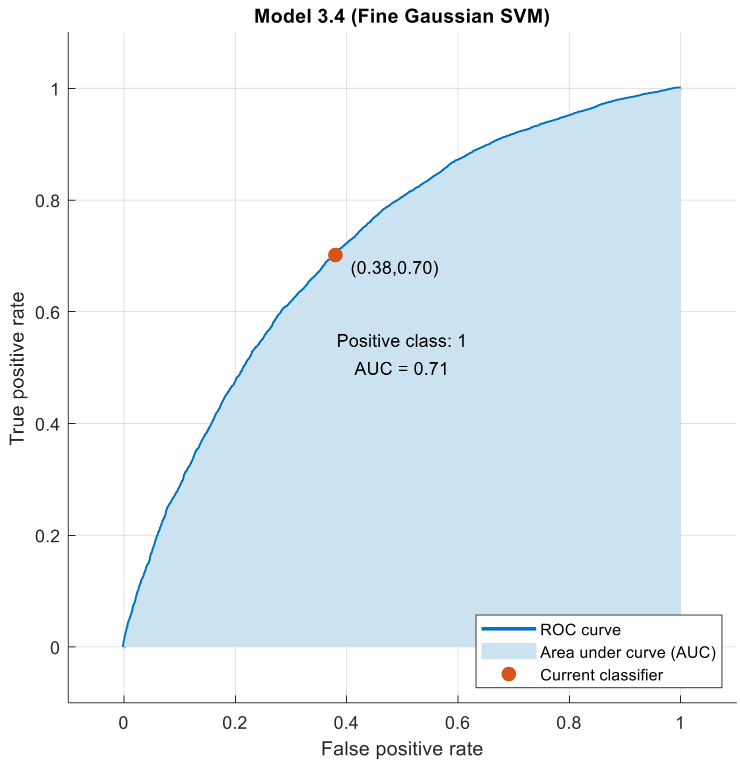


Figure 3.9 ROC curve of valence for men.

The ROC curve obtained as a result of the classification made on the valence scale of male subjects is as shown in Figure 3.9. The complexity matrix of valence with using KNN for 20 men is given in Figure 3.10.

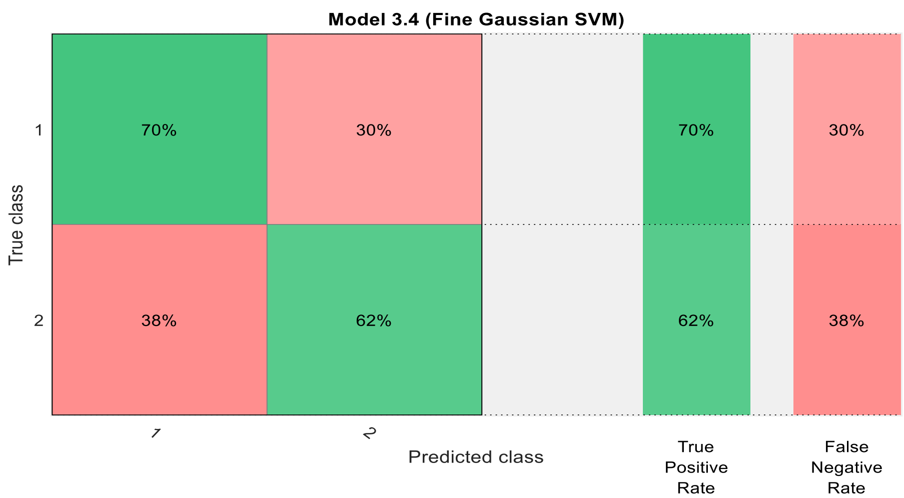


Figure 3.10 SVM classification of valence for men.

3.2 Classification After Using MEMD

Table 3.2 Classification results after MEMD.

Gender	Emotion Parameters	Mode Functions	SVM (Fine Gaussian)	KNN (Weighted)	Ensemble (Bagged Trees)
WOMEN	Arousal(High/Low)	1.IMF	**%62,0**	%61,3	%60,6
	Valence(High/Low)	1.IMF	**%65,1**	%63,6	%63,7
MEN	Arousal(High/Low)	1.IMF	**%62,9**	%61,4	%60,6
	Valence(High/Low)	1.IMF	**%62,8**	%61,9	%61,7
WOMEN	Arousal(High/Low)	2.IMF	**%65,9**	%65,8	%63,4
	Valence(High/Low)	2.IMF	**%67,1**	%66,9	%64,9
MEN	Arousal(High/Low)	2.IMF	**%62,3**	%61,8	%59,3
	Valence(High/Low)	2.IMF	**%63,5**	%62,8	%61,2
WOMEN	Arousal(High/Low)	3.IMF	%74,3	**%75,9**	%71,0
	Valence(High/Low)	3.IMF	%74,8	**%76,3**	%71,9
MEN	Arousal(High/Low)	3.IMF	**%69,5**	%69,0	%66,4
	Valence(High/Low)	3.IMF	**%69,0**	%68,8	%65,6
WOMEN	Arousal(High/Low)	4.IMF	%78,7	**%81,6**	%77,9
	Valence(High/Low)	4.IMF	%78,4	**%81,4**	%76,1
MEN	Arousal(High/Low)	4.IMF	%73,6	**%74,6**	%70,3
	Valence(High/Low)	4.IMF	%73,5	**%74,1**	%70,8
WOMEN	Arousal(High/Low)	1. and 2.	**%63,1**	%62,1	%62,1
	Valence(High/Low)	1. and 2.	**%65,4**	%64,8	%64,3
MEN	Arousal(High/Low)	1. and 2.	**%61,1**	%60,7	%60,4
	Valence(High/Low)	1. and 2.	**%62,3**	%61,7	%60,8
WOMEN	Arousal(High/Low)	1. and 3.	%66,0	**%67,3**	%66,0
	Valence(High/Low)	1. and 3.	%67,7	**%68,3**	%67,6
MEN	Arousal(High/Low)	1. and 3.	%63,6	**%64,3**	%63,8
	Valence(High/Low)	1. and 3.	%64,3	**%64,8**	%63,5
WOMEN	Arousal(High/Low)	1. and 4.	%66,0	%68,5	**%69,3**
	Valence(High/Low)	1. and 4.	%67,7	%69,2	**%70,2**
MEN	Arousal(High/Low)	1. and 4.	%64,6	**%65,1**	**%65,1**
	Valence(High/Low)	1. and 4.	%64,9	**%66,1**	%65,8

The results of the classifications made with the IMFs found using the MEMD method are given in Table 3.2. According to table, when high frequency components of EEG signal were eliminated from the EEG signal by using MEMD method, it was noticed that classification success increases.

When the combinations of the 1st IMF with other IMFs were classified, it was observed that our success rate increased. Based on this information, it can be said that the distinguishing information for the classifier is in the delta and theta frequency band of the EEG signal. Because the success rate of the 4th and 5th IMF is higher than other IMFs. To interpret the table, it is seen that SVM method is more successful than other methods in the classification of 1st and 2nd IMF. In the classification of the fourth IMF, it is seen that the KNN method is more successful than other classifications. Again, the classification success of female subjects is higher than male subjects.
As a result of the classification, we achieved the highest success rate in the 4th IMF. All EEG bands are available in the first IMF. As the number of IMF increases, starting from the gamma band with the highest frequency components, the effect of EEG bands on IMF decreased, respectively. In 4.IMF, frequency components from gamma, beta, and alpha bands are almost not seen. As a result, it can be said that frequency components in theta and delta bands are more effective in emotion classification.

Then, as described in the material and method section, we made classification with deep learning. The success rates here exceed the machine learning and reveal the advantage of deep learning. The success rates in female subjects were higher than male subjects. The use of the 4th IMF played a major role in the increase of our success rate. It seems that the most successful result is obtained when the signal processing methods are used together with deep learning.

3.3 Classification Using Deep Learning

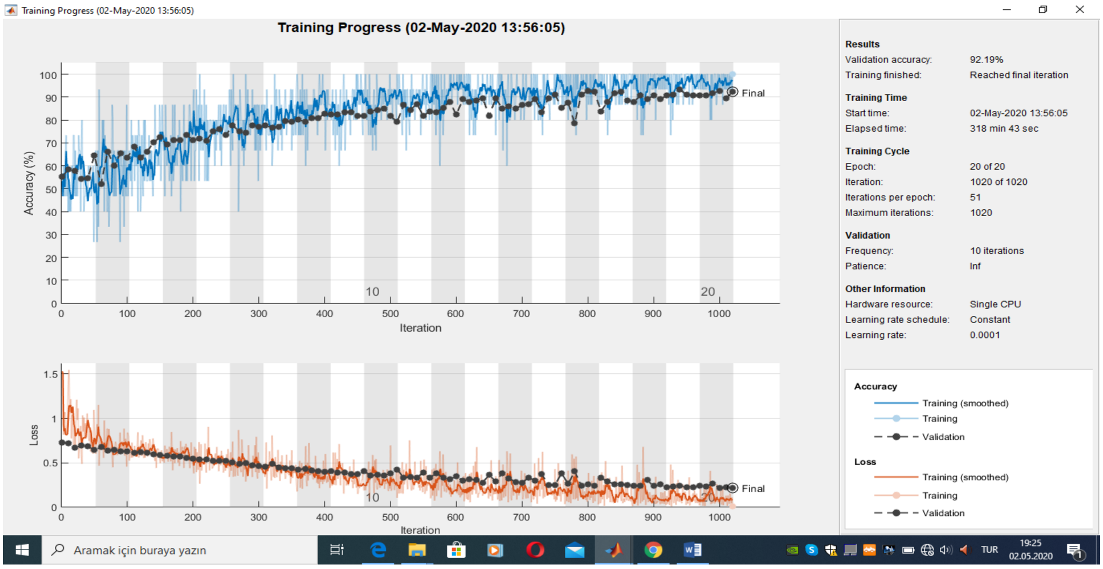


Figure 3.11 Deep learning classification of arousal for women.

We achieved a 92 percent success rate in female subjects in arousal scale. The graphics of accuracy and loss functions are shown in Figure 3.11.

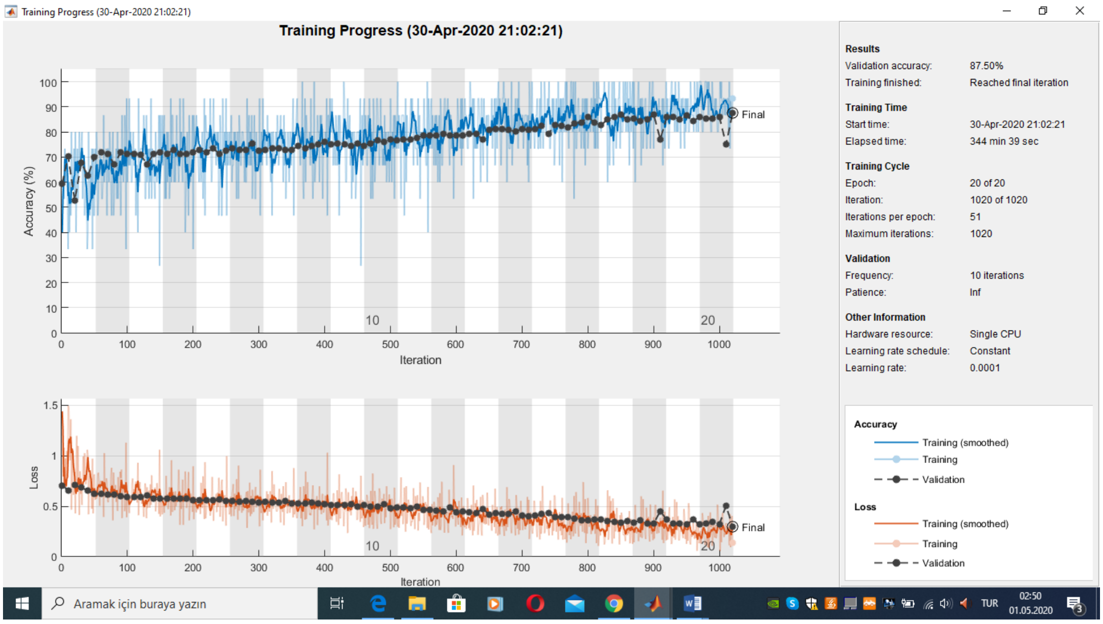


Figure 3.12 Deep learning classification of valence for women.

We achieved a 87 percent success rate in female subjects in valence scale. The graphics of accuracy and loss functions are shown in Figure 3.12.

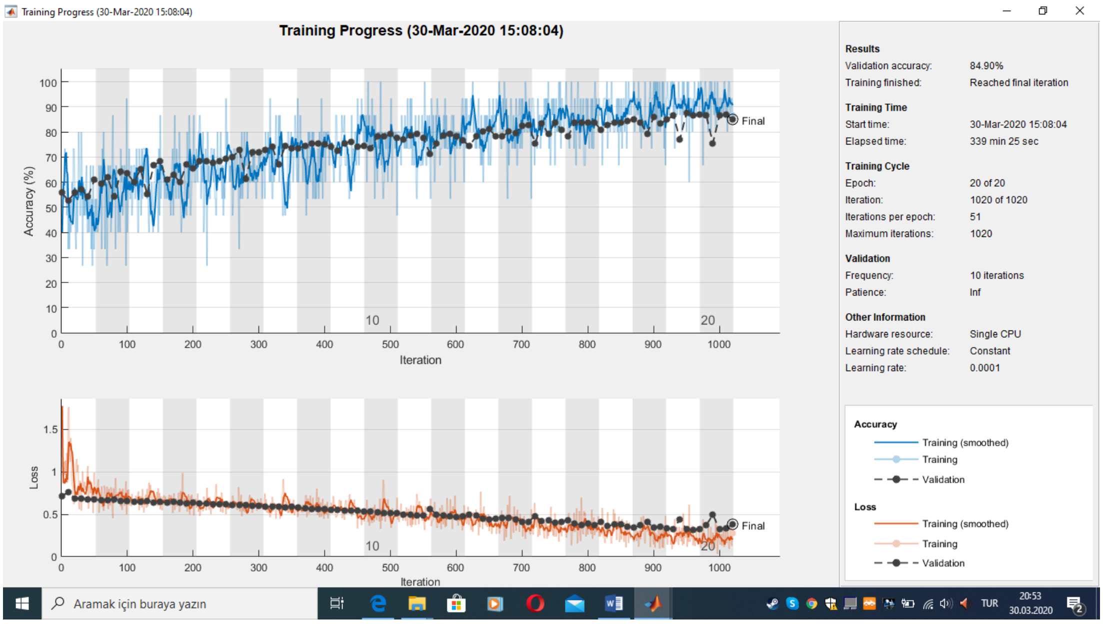


Figure 3.13 Deep learning classification of arousal for men.

We achieved a 84,9 percent success rate in male subjects in arousal scale. The graphics of accuracy and loss functions are shown in Figure 3.13.

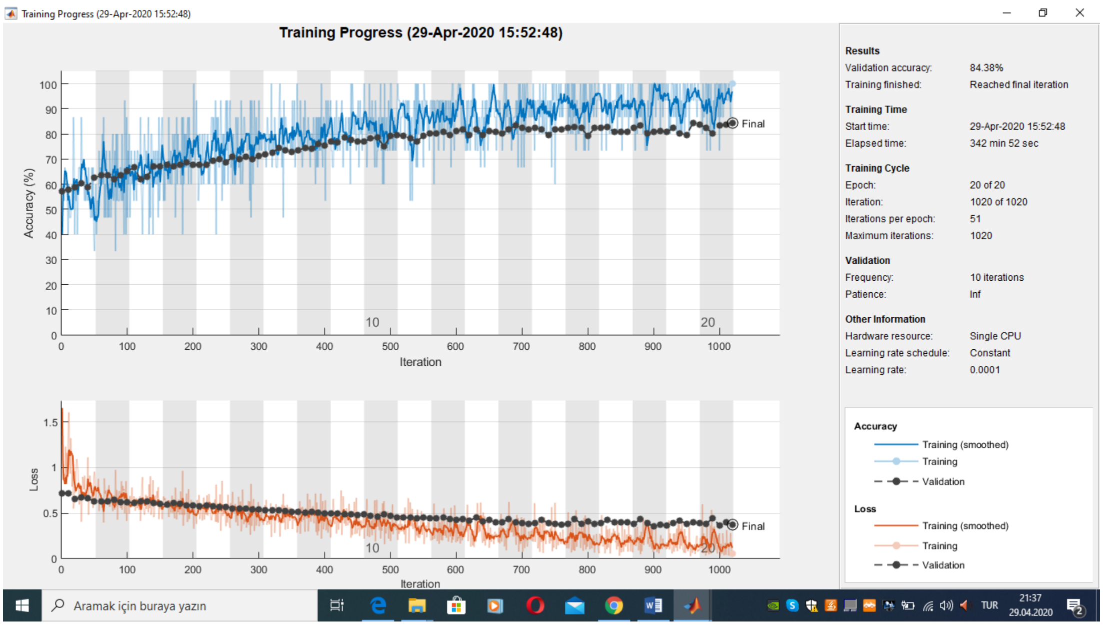


Figure 3.14 Deep learning classification of valence for men.

We achieved a 84,38 percent success rate in male subjects in valence scale. The graphics of accuracy and loss functions are shown in Figure 3.14.

4. DISCUSSION

When we increased the number of channels, we observed that the success rates increased in machine learning. As we increased the number of people, we saw that the classification success decreased in machine learning.

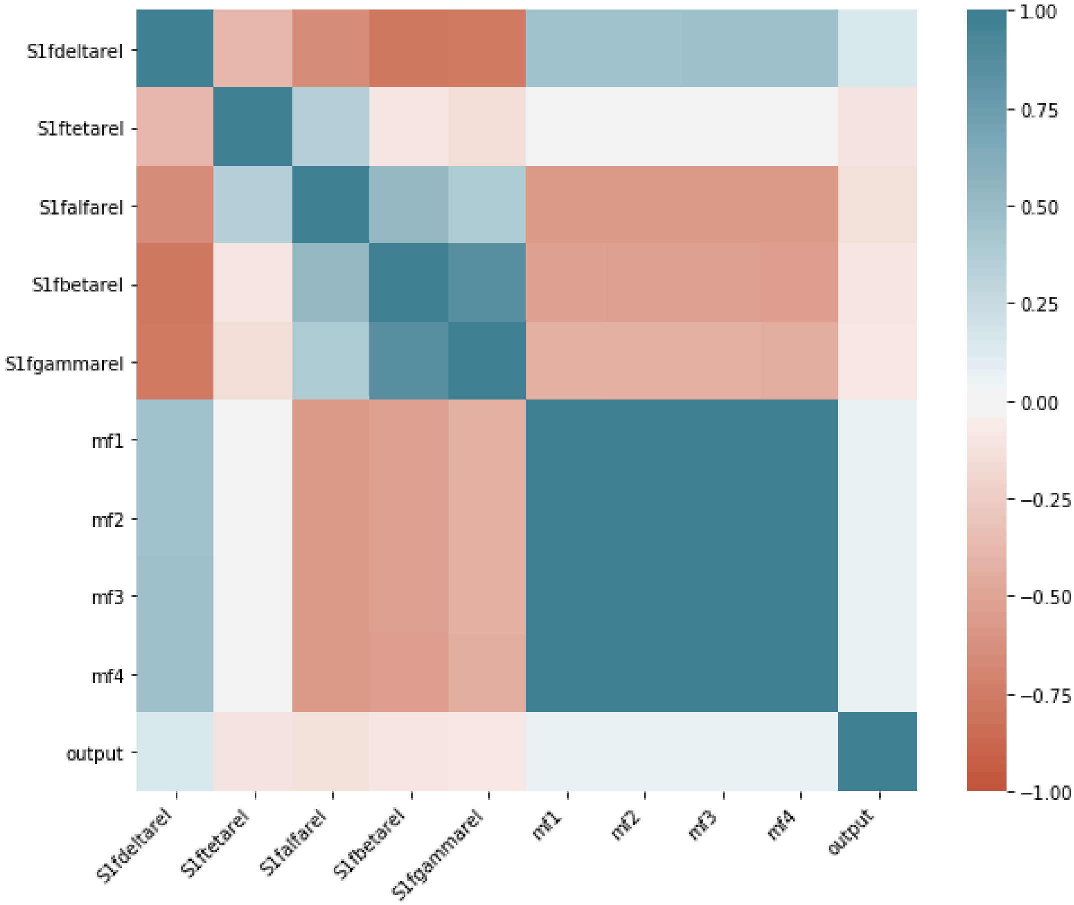


Figure 4.1 Correlation of features.

Figure 4.1 above shows the correlations of the features relative to each other. Accordingly, the relation between the two features is expressed with a color scale from blue to red. The positive correlation of the two features with each other is shown with the blue color and the negative correlation of the two features with each other is shown with the red color. If the values of one feature increase while the values of another feature increase, that is, if there is a linear relationship, this relationship is expressed in blue color. On the contrary, if the values of one feature decrease and the value of another feature increases or if there is a reverse situation, this relationship is expressed in red. If there is no positive or negative correlation between the two features, this relationship is expressed in white color. Accordingly, frequency moments show a linear feature in themselves. When we look at the correlation between frequency moments and band powers, we can say that delta and theta have positive correlations with frequency moments but alpha, beta and gamma have negative correlations with frequency moments. When the band powers are examined among themselves, it is observed that the delta band power has negative correlation with all other band powers. Time moments are not included in this table, but are known to behave in the same way as frequency moments. As a result, band powers contain more information in classification than time and frequency moments.

5. CONCLUSION

In this thesis, machine learning and deep learning models are proposed for emotional prediction from data. Various signal processing methods were used before the classifiers were applied. After applying signal processing methods, feature matrices are created and classified using machine learning and deep learning methods.

Within the scope of the thesis, firstly, emotion recognition was done by using machine learning methods. EEG signal is filtered to be taken only between 0.5-45 Hz. The data were divided into two groups as EEG record of male and female subjects. Then, two separate feature matrices for the valence and arousal scales from each group of data were created using the specified features. Then, it was classified with label values determined as high and low according to these scales. As a result of these processes, a maximum classification success of 70% was achieved. Secondly, EEG signals, which were filtered using MEMD, were opened to IMFs before classification using machine learning methods. Then, the feature matrix is classified according to the label values. It was recorded by finding out which of the IMF expansions and combinations gave the best classification results. The fourth IMF yielded the most successful classification result with a 81% success rate. As a third method, we propose using DL approach for the classification of emotions using multichannel EEG signals. We utilize the 2 dimensional spectrogram images of the IMFs calculated by the MEMD algorithm to train CNN architecture called Googlenet. The classification using the CNN yielded 84% accuracy, higher than the previous approaches.

As a result; we observed that the performance metrics improved as we developed better and more complicated methods. The signal processing methods used in the analysis and feature extraction of EEG signals play a major role in the success of the emotion estimation methods. In our experiments, classification results obtained using deep learning approach was found to be higher than machine learning methods used in our study.

REFERENCES

1. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG. Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine. 2001 Jan;18(1):32-80.
2. Devillers L, Vidrascu L, Lamel L. Challenges in real-life emotion annotation and machine learning based detection. Neural Networks. 2005 May 1;18(4):407-22.
3. Neuron structure, (2019). [online] Available at: https://www.coursehero.com /sg/anatomy-and-physiology/neuron-structure/ [Accessed 25 April 2019].
4. Abhang PA, Gawali BW, Mehrotra SC. Introduction to EEG-and speech-based emotion recognition. Academic Press; 2016 Mar 23.
5. Caton R. The electric currents of the brain. Br Med J. 1875;2:278.
6. Ahmed OJ, Cash SS. Finding synchrony in the desynchronized EEG: the history and interpretation of gamma rhythms. Frontiers in integrative neuroscience. 2013 Aug 12;7:58.
7. Alkaç Ü. Beyin araştırmaları tarihinde bir gezinti: Elektronörofizyoloji. Klinik Gelişim. 2009;3:14-9.
8. Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. Journal of Neuroscience Methods. 2003 Feb 15;123(1):69-87.
9. Hossain PS, Shaikat IM, George FP. Emotion recognition using brain signals based on time-frequency analysis and supervised learning algorithm (Doctoral dissertation, BRAC University).
10. Russell JA. Culture and the categorization of emotions. Psychological Bulletin. 1991 Nov;110(3):426.
11. Soleymani M, Pantic M, Pun T. Multimodal emotion recognition in response to videos. IEEE Transactions on Affective Computing. 2011 Dec 6;3(2):211-23.
12. Petrantonakis PC, Hadjileontiadis LJ. A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition. IEEE Transactions on Information Technology in Biomedicine. 2011 May 27;15(5):737-46.
13. Mert A, Akan A. Emotion recognition from EEG signals by using multivariate empirical mode decomposition. Pattern Analysis and Applications. 2018 Feb 1;21(1):81-9.
14. Liu YH, Cheng WT, Hsiao YT, Wu CT, Jeng MD. EEG-based emotion recognition based on kernel Fisher's discriminant analysis and spectral powers. In 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2014 Oct 5 (pp. 2221-2225). IEEE.
15. Mehmood RM, Lee HJ. Emotion recognition from EEG brain signals based on particle swarm optimization and genetic search. In2016 IEEE International

Conference on Multimedia & Expo Workshops (ICMEW) 2016 Jul 11 (pp. 1-5). IEEE.
16. Mohammadpour M, Hashemi SM, Houshmand N. Classification of EEG-based emotion for BCI applications. In 2017 Artificial Intelligence and Robotics (IRANOPEN) 2017 Apr 9 (pp. 127-131). IEEE.
17. Krisnandhika B, Faqih A, Pumamasari PD, Kusumoputro B. Emotion recognition system based on EEG signals using relative wavelet energy features and a modified radial basis function neural networks. In 2017 International Conference on Consumer Electronics and Devices (ICCED) 2017 Jul 14 (pp. 50-54). IEEE.
18. Candra H, Yuwono M, Chai R, Nguyen HT, Su S. EEG emotion recognition using reduced channel wavelet entropy and average wavelet coefficient features with normal mutual information method. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2017 Jul 11 (pp. 463-466). IEEE.
19. Mert A, Akan A. Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform. Digital Signal Processing. 2018 Oct 1;81:106-15.
20. Marchewka A, Żurawski Ł, Jednoróg K, Grabowska A. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behavior Research Methods. 2014 Jun 1;46(2):596-610.
21. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences. 1998 Mar 8;454(1971):903-95.
22. Huang W, Shen Z, Huang NE, Fung YC. Engineering analysis of biological variables: An example of blood pressure over 1 day. Proceedings of the National Academy of Sciences. 1998 Apr 28;95(9):4816-21.
23. Rehman N, Mandic DP. Multivariate empirical mode decomposition. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2010 May 8;466(2117):1291-302.
24. Junsheng C, Dejie Y, Yu Y. Research on the intrinsic mode function (IMF) criterion in EMD method. Mechanical systems and signal processing. 2006 May 1;20(4):817-24.
25. Cohen L. Time-frequency analysis. Prentice hall; 1995.
26. Qian S, Chen D. Joint time-frequency analysis. IEEE Signal Processing Magazine. 1999 Mar;16(2):52-67.
27. Shaik BS, Naganjaneyulu GV, Chandrasheker T, Narasimhadhan AV. A method for QRS delineation based on STFT using adaptive threshold. Procedia Computer Science. 2015 Jan 1;54:646-53.
28. Tacer B, Loughlin PJ. Non-stationary signal classification using the joint moments of time-frequency distributions. Pattern Recognition. 1998 Nov 1;31(11):1635-41.
29. Velez EF, Absher RG. Segmentation and classification of nasal phonemes based on their time-frequency representation. In Advanced Signal Processing Algorithms, Architectures, and Implementations 1990 Nov 1 (Vol. 1348, pp. 188-196). International Society for Optics and Photonics.

30. Akan A, Ünsal RB. Time–frequency analysis and classification of temporomandibular joint sounds. Journal of the Franklin Institute. 2000 Jul 1;337(4):437-51.
31. Khan YU, Farooq O, Sharma P. Automatic detection of seizure onset in pediatric EEG. International Joural of Embeded Systems and Applications. 2012 Sep;2(3):81-9.
32. Cecchin T, Ranta R, Koessler L, Caspary O, Vespignani H, Maillard L. Seizure lateralization in scalp EEG using Hjorth parameters. Clinical Neurophysiology. 2010 Mar 1;121(3):290-300.
33. Vourkas M, Micheloyannis S, Papadourakis G. Use of ANN and Hjorth parameters in mental-task discrimination. In 2000 First International Conference Advances in Medical Signal and Information Processing (IEE Conf. Publ. No. 476) 2000 Sep 4 (pp. 327-332). IET.
34. Hofmann T. Unsupervised learning by probabilistic latent semantic analysis. Machine Learning. 2001 Jan 1;42(1-2):177-96.
35. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. Computer. 1996 Mar;29(3):31-44.
36. Chauhan NS. Classifying Heart Disease Using K-Nearest Neighbors. Available from:https://towardsdatascience.com/implement-k-nearest-neighbors-classification-algorithm-c99be8f14052 [Accessed 14th May 2020]
37. Osowski S, Siwek K, Markiewicz T. Mlp and svm networks-a comparative study. InProceedings of the 6th Nordic Signal Processing Symposium, 2004. NORSIG 2004. 2004 Jun 11 (pp. 37-40). IEEE.
38. Hastie T, Tibshirani R, Friedman J. Springer series in statistics. The elements of statistical learning: Data mining, inference, and prediction. 2001.
39. Bishop CM. Pattern recognition and machine learning. Springer; 2006.
40. Ben-Hur A, Weston J. A user’s guide to support vector machines. In Data mining techniques for the life sciences 2010 (pp. 223-239). Humana Press.
41. Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 1998 Jun 1;2(2):121-67.
42. Soman KP, Loganathan R, Ajay V. Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd.; 2009 Feb 2.
43. Lessmann S, Baesens B, Seow HV, Thomas LC. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research. 2015 Nov 16;247(1):124-36.
44. Kim MJ, Min SH, Han I. An evolutionary approach to the combination of multiple classifiers to predict a stock price index. Expert Systems with Applications. 2006 Aug 1;31(2):241-7.
45. Tsai CF, Hsiao YC. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems. 2010 Dec 1;50(1):258-69.
46. Finlay S. Multiple classifier architectures and their application to credit risk assessment. European Journal of Operational Research. 2011 Apr 16;210(2):368-78.
47. Paleologo G, Elisseeff A, Antonini G. Subagging for credit scoring models. European Journal of Operational Research. 2010 Mar 1;201(2):490-9.

48. Baxter CW, Zhang Q, Stanley SJ, Shariff R, Tupas RR, Stark HL. Drinking water quality and treatment: The use of artificial neural networks. Canadian Journal of Civil Engineering. 2001 Jan 1;28(S1):26-35.
49. Ranzato MA, Poultney C, Chopra S, Cun YL. Efficient learning of sparse representations with an energy-based model. In Advances in Neural Information Processing Systems 2007 (pp. 1137-1144).
50. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing. 2016 Apr 26;187:27-48.
51. Kayaalp K, Süzen AA. Derin Öğrenme ve Türkiye'deki Uygulamaları. Yayın Yeri: IKSAD International Publishing House, Basım sayısı. 2018;1.
52. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436-44.
53. Bottou L, Cortes C, Denker JS, Drucker H, Guyon I, Jackel LD, LeCun Y, Muller UA, Sackinger E, Simard P, Vapnik V. Comparison of classifier methods: A case study in handwritten digit recognition. In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5) 1994 Oct 9 (Vol. 2, pp. 77-82). IEEE.
54. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016 Nov 10.
55. Özkan İN, Ülker E. Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri. Gaziosmanpaşa Bilimsel Araştırma Dergisi. 2017;6(3):85-104.
56. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
57. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 2012 (pp. 1097-1105).
58. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research. 2014 Jan 1;15(1):1929-58.
59. Polamuri S. Difference between softmax function and sigmoid function.

CURRICULUM VITAE

PERSONAL INFORMATIONS

Name-Surname: Bartu Yeşilkaya

Place of Birth: Sarayköy

Date of Birth: 19/09/1993

Driving Licence: B Class

Address: 285/1 sk. No: 10 Daire: 4 Pirireis Mah. Hatay/IZMIR

Phone: +90 5358217361

E-mail: bartu.yesilkaya@yandex.com

EDUCATIONS

2011-2016 Dokuz Eylül University, Electrical&Electronics Engineering Bachelor Degree

2018- Still Izmir Katip Celebi University, Biomedical Technology Master Degree

PROGRAMMING SKILLS

Microsoft Office, Matlab, Python, Rasspery Pi

SCIENTIFIC ACTIVITIES

IEEE 26th Signal Processing and Communication Practices Congress (SIU-2018) Cesme / Izmir/ TURKEY	02-05 May 2018
Medical Technologies Conference (TIPTEKNO'18) Gazi Magosa/ TRNC	08-10 November 2018
Medical Technologies Conference(TIPTEKNO'19) Kusadası / Aydin / TURKEY	03-05October 2018

PUBLICATIONS

A. <u>Papers presented at national scientific meetings and published in the proceedings book:</u>

A1. GD Ekimci, MA Özdemir,**B Yeşilkaya**,M Unay,UK Ercan,A Akan,Processing and Characterization of Acoustic Signals Emitted During Plasma Discharge,TIPTEKNO'18, 2018.
A2. B. Yasar, **B. Yeşilkaya**, O. Karabiber Cura, A. Akan, 'A New Method to Detect Diabetic Retinopathy',

Medical Technologies Conference, TIPTEKNO'18, Cyprus, 08-10 November 2018.
A3. RS Bastug, **B Yeşilkaya**, M Unay, A Akan, Virtual Mouse Control by Webcam for the Disabled, TIPTEKNO'18, 2018.
A4. Karakus, D., Karabiber, Cura, Ö., **Yesilkaya, B.**, ve Akan, A., "Sanal Klavye Kontrolü İçin Beyin-Bilgisayar Arayüzü Tasarımı," Medical Technologies Conference, TIPTEKNO'19, Aydin, 3-5 October 2019.
A5. Ata, A., **Yesilkaya, B.**, Karabiber, Cura, Ö., ve Akan, A., "Control of Serious Games Designed for Alzheimer's and Dementia Patients by EEG Signals," Medical Technologies Conference, TIPTEKNO'19, Aydin, 3-5 October 2019.