YAŞAR UNIVERSITY

GRADUATE SCHOOL

MASTER THESIS

EMOTION CLASSIFICATION WITH EEG SIGNALS USING CONVOLUTIONAL NEURAL NETWORKS

HAYRİYE DÖNMEZ

THESIS ADVISOR: ASSIST. PROF. DR. NALAN ÖZKURT

ELECTRICAL AND ELECTRONICS ENGINEERING

PRESENTATION DATE: 12.08.2021

BORNOVA / İZMİR
August 2021

ABSTRACT

EMOTION CLASSIFICATION WITH EEG SIGNALS USING CONVOLUTIONAL NEURAL NETWORKS

Dönmez, Hayriye

MSc, Electrical and Electronics Engineering

Advisor: Assist.Prof.Dr. Nalan ÖZKURT

August 2021

Assistive technologies for human-machine interface studies are one of the important research areas of the today's technology. Emotion perception for this interaction mechanism is one of the valuable requirements to provide communication with interface and person. Emotion classification studies are an improving field and several approaches have been proposed by the researchers. In the literature emotion definition mainly given according to two view. In this study, these views which are discrete emotion model and dimensional model-based approaches were examined.

Since EEG signals have nonlinear and nonstationary characteristics, several feature extraction and classification algorithms applied in literature. Furthermore, deep learning models have reached significant accuracy for various types of data in recent years, thus, it was aimed to get accurate results for emotion classification by examining Convolutional Neural Networks (CNN) algorithms.

In our thesis, three different study were accomplished. The first study investigates the emotion classification performance of single electrode EEG recordings with commercial Neurosky EEG device with CNN using spectrogram features. A second approach is to use spectrogram features again for multiple channel recordings with publicly accesible DEAP dataset. Finally, to improve the performance and reduce the computation complexity, Fourier and wavelet transform features of all channels were classified with CNN. To understand the performance of CNN model and make a reliable comparison with signal transformations, also raw signal image features used in this thesis.

Key Words: Electroencephalogram (EEG), emotion, convolutional neural networks (CNN), valence, arousal, human machine Interaction (HMI), emotion classification, wavelet, fast fourier transform (FFT), spectrogram, DEAP database

ÖZ

EEG SİNYALLERİNDEN EVRİŞİMSEL SİNİR AĞLARI KULLANIMI İLE DUYGU SINIFLANDIRMASI

Dönmez, Hayriye
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği
Danışman: Dr.Öğr.Üy. Nalan ÖZKURT
Ağustos 2021

İnsan-makine arayüz çalışmaları için yardımcı teknolojiler, günümüz teknolojisinin önemli araştırma alanlarından biridir. Bu etkileşim mekanizması için duygu algısı, arayüz ve kişi ile iletişimi sağlamak için değerli gereksinimlerden biridir. Duygu sınıflandırma çalışmaları gelişen bir alandır ve araştırmacılar tarafından çeşitli yaklaşımlar önerilmiştir. Literatürde duygu tanımı temel olarak iki teoriye göre verilmektedir. Bu çalışmada, dairesel model ve boyutlu model temelli yaklaşımlar olan bu görüşler incelenmiştir.

EEG sinyalleri doğrusal ve durağan olmayan özelliklere sahip olduğundan literatürde birçok öznitelik çıkarma ve sınıflandırma algoritması uygulanmaktadır. Ayrıca, derin öğrenme modelleri son yıllarda çeşitli veri türleri için önemli bir doğruluğa ulaştığından, Evrişimsel Sinir Ağları (ESA) algoritmaları incelenerek duygu sınıflandırması için doğru sonuçlar elde edilmesi hedeflenmiştir.

Tezimizde üç farklı çalışma gerçekleştirilmiştir. İlk çalışma, spektrogram özelliklerini kullanarak ESA'lı ticari Neurosky EEG cihazı ile tek elektrotlu EEG kayıtlarının duygu sınıflandırma performansını araştırmaktadır. İkinci bir yaklaşım, genel kullanıma açık DEAP veri seti ile çok kanallı kayıtlar için spektrogram özelliklerini tekrar kullanmaktır. Son olarak, performansı iyileştirmek ve hesaplama karmaşıklığını azaltmak için tüm kanalların Fourier ve dalgacık dönüşümü özellikleri ESA ile sınıflandırılmıştır. Ayrıca bu tezde, ESA modelinin performansını anlamak ve sinyal dönüşümleri arasında güvenilir bir karşılaştırma yapmak amacıyla ham sinyal görüntü öznitelikleri kullanılmıştır.

Anahtar Kelimeler: Elektroensefelografi (EEG), duygu, evrişimsel sinir ağları (ESA), valans, uyarılma, insan-makine arayüz (İMA), duygu sınıflandırması, dalgacık, hızlı Fourier dönüşümü, spektrogram, DEAP veri kümesi.

ACKNOWLEDGEMENTS

I would like to give sincere gratitude to my supervisor Nalan Özkurt for her valuable supervision, meaningful support and patience during my master education. Additionally, I would like to express my thanks to my all friends for their heartfully and valuable supports, carings and believes. Also, I give special thanks to Rabia Karagöz for her continues believe in me.

Lastly, I would like to say special thanks to my mother who always loved, supported and believed in me in every moment of my life, I am grateful for her patience to my journey.

Hayriye Dönmez

İzmir, 2021

TEXT OF OATH

I declare and honestly confirm that my study, titled “EMOTION CLASSIFICATION WITH EEG SIGNALS USING CONVOLUTIONAL NEURAL NETWORKS” and presented as a Master’s Thesis, has been written without applying to any assistance inconsistent with scientific ethics and traditions. I declare, to the best of my knowledge and belief, that all content and ideas drawn directly or indirectly from external sources are indicated in the text and listed in the list of references.

Hayriye Dönmez

August 20, 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAM	Active Appearance Model
ANN	Artificial Neural Network
AU	Action Units
CNN	Convolutional Neural Networks
DBN	Deep Belief Networks
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculography
FACS	Facial Action Coding System
FER	Facial Emotion Recognition
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
GSR	Galvanic Skin Response
HMI	Human-Machine Interaction
HMM	Hidden Markov Model
ICA	Independent Component Analysis
KNN	K-nearest Neighbors Algorithm
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LSTM	Long Short Term Memory
mRMR	minimum-Redundancy-Maximum-Relevance

PCA	Principal Component Analysis
PSD	Power Spectral Density
SAM	Self-Assessment Manikin
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TEO	Teager Energy Operator
TN	True Negative
TP	True Positive

CHAPTER 1
INTRODUCTION

1.1. Motivation of Study

In the recent years, emotion classification studies become a growing research topic for various fields in engineering since the analysis and perception of emotional reactions by machines or interfaces provide significant improvement in these fields. Human-Machine Interaction (HMI) is one of these fields which is interesting and important considering that emotional responses play an important role in social communication. It is required to provide the relevance between humans' emotional reactions and perception of these reactions by the machines. And moreover, it may be answered how can a machine or an interface earns these skills by using interaction with human brains. All these points are fundamental aims of affective computing research field. It was firstly introduced in 1997 (Picard, 1997). Affective computing includes interdisciplinary works between computer science, psychology, cognitive neuroscience etc. In this field, various healthcare, educational and daily life applications have been produced observing the main aims of affective computing studies. Humanoid robot studies, paralyzed patients' assistive applications (in order to provide communication ability), neuromarketing (in order to determine consumer reactions for products), online gaming, monitoring mental health, recommender systems and smart homes are some of the examples for this area. All these studies need adaptive systems which use the outcomes of the emotional detections and provide a user-friendly interaction. According to emotion regulation steps, first we experience the emotion (visual, auditorial etc.), and then our physiological reactions occurred (handshake, heart rate etc.) and finally our body gives physical reaction (facial expressions or body movements). To detect emotions, human body's sensory data of physiological reactions are used due to the fact that when a person experience an emotion, her or his body also affected by it.

1.2. Literature Survey

In order to detect the emotions, some features are extracted from these affected conditions with various modalities. In the literature, these modalities mainly introduced as visual, auditorial and physiological. It is considered that all these modalities include some cues related with the experienced emotion.

Facial expressions are commonly used and non-verbal visual data forms for emotion recognition. Facial expression analysis is used to extract information of affected face reaction for emotion recognition in HMI applications. Facial actions firstly introduced by Ekman in psychology. The Facial Action Coding System (FACS) was proposed, system provided a measurement standard for the emotions. Changes in the muscular movements of the face was examined due to the fact that emotional states cause these changes. Concordantly, different codes are created according to changes of muscular movements and named as Action Units (AUs). According to FACS, recorded reactions of the participants are used to make emotion detection (such as sad, happy, anger etc.) (Ekman et al., 1980). After this priori research, various face recognition analyses were made with several techniques in engineering. FER (facial emotion recognition) studies are generally divided in 2 sub-groups which are called as dynamic based and static based FER. If FER studies include spatial information (such as morphology and complexion) from single image, it is a static based FER. If they include temporal relation between frames of a video, it is a dynamic based FER. Mainly system has three stages which are face detection, feature extraction, feature selection and classification. Existing feature extraction methods mainly divided in three groups: Geometric-based approaches, Appearance-based approaches, hybrid approaches (combination of appearance-based and geometric-based features) (Ko, 2018). Several FER feature extraction and classification techniques such as Gabor Filter/Energy, Active Appearance Model (AAM) and using Local Binary Pattern (LBP), Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA), Adaboost, Neural Network and Independent Component Analysis (ICA), Deep Belief Network (DBN), Hidden Markov Model (HMM), Support Vector Machine and deep learning algorithms are used in this study area (Ko, 2018) (Vinola et al., 2015) (Revina et al., 2021). Some of the studies results examined on this area are given in Table1.1., highly preferred databases and self-generated databases used studies preferred to present general overview.

Table 1.1. Some of the literature studies performances for facial emotion classification.

Previous Studies	Approaches	Database and Accuracy	Number of Emotions
Zhang et al.	Gabor Features+Adaboost Gabor Features+SVM	JAFFE, 92,93% CK database, 94,48%	6
Happy et al.	LBP+SVM	JAFFE, CK 93,3%	6
Taylor et al.	PCA, ICA+HMM	Self-generated, 98%	6
Sujono and Gunawan	AAM+Fuzzy Logic	Self-generated, 95,87%	3
Zhao et al.	DBN+MLP	JAFFE, CK+ 90,95% 98,57%	7
Jain et al.	CNN	JAFFE, CK+ 95,23%, 93,24%	6

Although there are many studies accomplished in this study area. There are some disadvantages that affect the performance in facial expression detection studies. Occlusion, pose normalization, illumination, dynamic background, age, gender are some of the main problems that affects the performance of the FER systems since all of them affects face view thus the proper feature extractions become though.

Speech is one of the major and verbal stage of the communication, it is affected from the emotional states during the conservation of a person. Utterances which are the little portions of the speech naturally include emotional state cues of a person hence speech becomes a modality for emotion recognition. Acoustic difference which is occurred when speaking about the same thing is studied for emotion recognition under divergent emotional situations. This situation inspired engineers to study on speech emotion recognition for several HMI systems and assistive technologies (Ayadi et al., 2011). Due to the fact that speech signals have dynamic characteristics, in these studies speech signals are splitted into small frames generally. These frames can be accepted as almost stationary portions. Speech emotion recognition has two main processes feature extraction from the speech data, and a classification to decide the emotional state of an utterance. Several local and global features exist in these portions such as prosodic speech features (pitch, energy etc.) are extracted from each frame and these features grouped in local features. On the other hand, calculating the statistics of speech utterances method is grouped in global features. Although these features are grouped in terms of the local and global features, and furthermore, speech feature types can also be categorized with four main titles technically. These are continuous speech features (such as pitch, energy, formants), voice quality

features, spectral-based speech features and nonlinear TEO (Teager energy operator)-based features (Ayadi, 2011). The studies presented thus far have examined in various recognition methods. Some of the methods which are used to feature extraction and classification are PCA, LDA, acoustic features, HMM (Hidden Markov Model) Classifier, Gaussian Mixture Models, Neural Networks, Support Vector Machine, Multiple Classifier Systems and Deep learning etc. (Abbaschian et al., 2021) (Ayadi et al., 2011) (Basharirad et al, 2017). In the given Table.1.2. some of these researches presented.

Table 1.2. Some of the literature studies performances for speech emotion classification.

Previous Studies	Approaches	Database	Accuracy
Chen et al.	Energy, pitch, frequency etc. SVM +ANN	Beihang University Database of Emotional Speech (BHUDES)	86,5%, 68,5% and 50,2% for different level
Albornoz et al.	Mean of the log-spectrum (MLS), MFCCs and prosodic features+HMM, GMM, MLP and hierarchical model	Berlin dataset	HMM 68,57, Hierarchical model 71,75
You et al.	PCA+LDA SVM	Mandarin language dataset	83,4% for males and 78,7% for females
Eskimez et al.	Log-Mel Spectrogram+ Variational Autoencoder(VAE), CNN	IEMOCAP	48,54%

However, since speech recognition is needed to be studied in terms of the speech utterances, dynamics and lingual characteristics of a language should be considered, hence this prevents the production of universal solutions and creates a challenge. Additionally, expressions are highly related with speakers' cultures and environment this also effects making certain classification emotional state. Lastly, in

computational manner limited data samples creates inadequate datasets especially for deep learning applications (Abbaschian et al., 2021) (Ayadi et al., 2011).

Emotion recognition related HMI studies use some neurophysiological analysis's that include markers related with emotional experience. Physiological signals are affected from the emotional experiences by reason of the fact that human body reacts with respect to these experiences. Since, these body reactions are under unvoluntary control of human, physiological signal data becomes valuable. In these studies, specially designed experiments are preferred to induce emotions and measured with both invasive and non-invasive measurement methods. Some of these methods are as follows.

EOG (Electrooculography): Electrooculography measures the cornea-retinal potential that exists between in human eye's front and back area

ECG (Electrocardiography): Electrocardiography interpret the electrical activity of heart in voltage-time domain.

EMG (Electromyography): Electromyography provides the activity of muscles via measuring the frequency of muscle tensions in voltage-time domain.

Skin temperature: The temperature that is measured from the skin. Due to the muscles and blood vessels movements' changes, it is affected.

Respiration: Respiration sensors measures the deepness of the breathing.

Plethysmography: A plethysmograph is a device display the changes of human body volume.

Galvanic skin response (GSR): Galvanic skin response measures the conductivity of the skin.

EEG (Electroencephalography): EEG signals shows the electrical activity of human brain, in voltage-time domain (Haag. et al., 2004).

Physiological signal used studies are mainly categorized according to which type of data is selected for the recognition task. Generally, recognition steps are data preprocessing, feature extraction and classification. Data recordings are preprocessed with filtering operations because due to the nature of these signal types they might carry environmental and bodily noises. Various feature extraction methods are examined for this area in engineering field, especially time-frequency domain

analysis (Wavelet Transform), statistical calculations (Mean, Standard Deviation etc.), and unsupervised methods (PCA, LDA, Autoencoders etc). Additionally, diverse classification algorithms are selected such as SVM (Support Vector Machine), ANN (Artificial Neural Network), KNN (k-nearest neighbors algorithm), CNN, DBN (Deep Belief Networks), LSTM (Long Short Term Memory) etc. (Shu et al., 2018)

In this thesis, EEG signals are chosen to recognize emotional experience of human brain for affective state aspect in this research. Accordingly, for this purpose, DEAP (A Database for Emotion Analysis using Physiological Signals) dataset was applied (Koelstra et al., 2012)

1.3. Aim of the Study

The main goal of this thesis study is to provide an emotion classification system via EEG signals to improve the performance of the previous classifiers by using the learning ability of deep learning tools and basic signal processing methods such as Fourier transform and time-frequency transformations such as Short Time Fourier Transform (STFT) and wavelet transform. EEG signals are highly nonstationary and it is hard to find patterns simply. Thus, our aim is to extract some useful information using the time-frequency information. Also, convolutional neural network models were employed to grasp the hidden patterns inside data to improve the classification accuracy.

The study includes three parts

- The classification of basic emotions while watching movie clips with spectrogram and GoogleNet CNN model by using single electrode EEG recordings from Neurosky Mindwave Mobile 2 device were implemented.
- Binary classification of valance and arousal states with spectrogram of multiple channel recordings from trusted and highly chosen public database DEAP with custom designed CNN was proposed as a second study. The novelty of the study comes from the construction of the input matrices from wavelet transform coefficients.

To improve the accuracy and to reduce the computational complexity in binary classification of valance and arousal states for DEAP dataset, a new classification

system was proposed. This system uses all channel recordings together and classifies with custom designed CNN model. The input image was constructed from FFT and wavelet coefficients individually. Raw data was also used in classification tasks to better understand the performances of transformed signals.

1.4. Outline of the Thesis

Thesis outlines are as follows in below:

- Chapter 1 – Provides a brief explanation of the study's motivation. Moreover, introduces the state-of-the art in emotion recognition. Also gives information with the general usages and mentions the challenges of them.
- Chapter 2 – Gives the general information for emotion and guided emotion model which DEAP database created accordingly. Then mentions about the chosen DEAP database. Also defines the EEG devices
- Chapter 3 –Presents general overview for the signal processing techniques that were used in this study which are STFT, FFT and Wavelet transforms. Additionally, theoretical background of our study's classification method which is CNN algorithm was given briefly.
- Chapter 4 – Includes the emotion classification implementation approaches which were examined in this thesis. Spectrogram images of our own single electrode EEG recordings were used in the first task. Then spectrogram task was examined with multichannel electrode public DEAP dataset selected. Lastly, DEAP database was classified with CNN with three different approaches such as raw images, FTT and wavelet transform images features.
- Chapter 5 – Draws the conclusions of our study and covers some future development ideas for these tasks.

CHAPTER 2
EMOTION AND EEG

2.1. Emotion

Emotions are biological reactional states for the external stimulus which relate to various neurophysiological, behavioral and cognitive processes. For a general emotion definition, scientists do not have a common judgement or opinion yet. Mainly, literature definitions can be divided into two groups. One of it, discrete and limited set of basic emotions point of view and the other one, dimensional models' point of view. The prior study in discrete set of basic emotions theorem was firstly introduced by Ekman et al. according to this theorem emotions are expected to be experienced separately from one another and people experience same emotions under same situations (Ekman et al., 1980). Also, as it is mentioned in Chapter 1, he proposed the idea that expressions are results of the emotional experiences and patterns of these expressions are considered related with the corresponded emotions. He defined emotions in six basic titles such as happy, sad, anger, fear, surprise and disgust. He claimed the other emotions are the combinations of these basic emotions.

Although Ekman proposed such kind of theorem which is discrete emotion experiences as it is stated in above, it is pointed out by Feldman Barrett that generally subjects in researches describe their emotional experiences not only with a single state but also with other close positive or negative emotions (Feldman Barrett, 1998). On account of this situation, dimensional models presented, and these models provided an opportunity to describe affective experience of emotional experience without limiting to give a name for a single and distinct emotion. Affect is accepted as a consequence of the interactive relation of experienced emotions. Russel's circumplex model is one of the models in these theorems and associated with valence (which an emotion is pleasant or unpleasant) and arousal (which it is behaviorally active or deactive) dimension scales and describes affective states according to these levels (Russel, 1980). According to this model, participants experience emotions active / deactive and pleasant/ unpleasant intensity levels.

This approach makes the emotion classification task results more trustable since it provides a general expression with intensity levels. The database we selected for our study was also created according to this model.

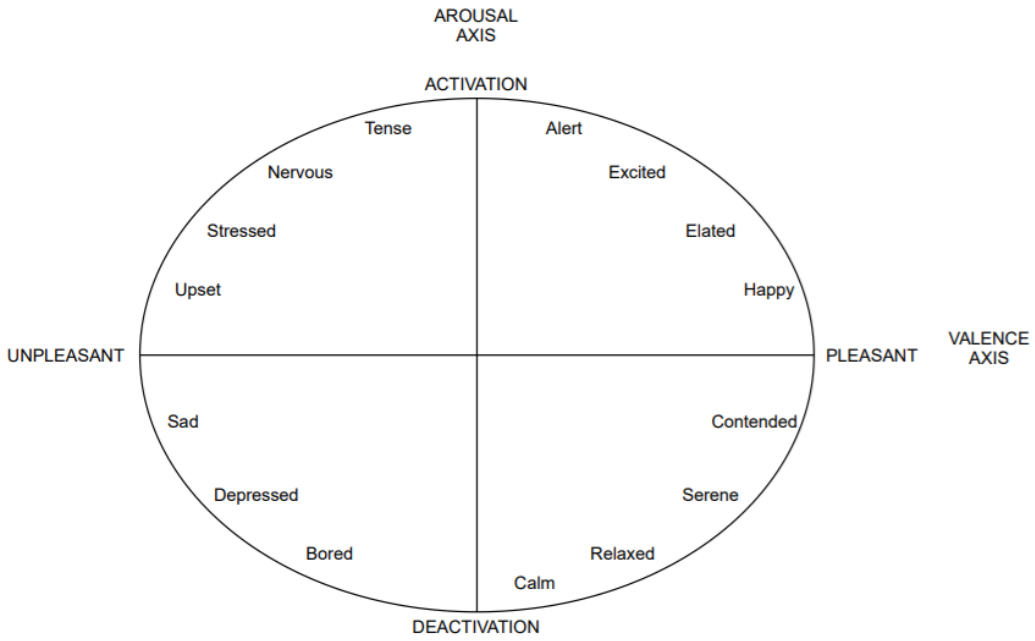

Figure 2.1 The circumplex model of emotion (Posner et al., 2005)

2.2. Electroencephalography

Electroencephalography (EEG) is an electronical device which measures and records the activity of brain electrical potentials. A German psychiatrist Hans Berger invented the EEG device for humans in 1924. EEG device provides the recording via metal electrodes on it, these electrodes work as a receiver-nodes, they might be placed outside or inside of the brain and collect data from the points which they touch on the skin. If electrodes are placed inside of the head skin it is called as invasive and if it is placed on the headskin, it is called as non-invasive type. Invasive types are placed bottom of head skin by surgical operations. On the other hand, non-invasive electrodes are placed on the head skin, using licra scalps and sticky jels, so it provides a significant usage practice if we compare with invasive types. EEG might have several channel-numbers; such as 1, 16, 32, 64, 128 etc. Electrodes are positioned referencing standard electrode position systems. Signal processing is done with different sampling frequencies (Dzedzickis et al, 2020).

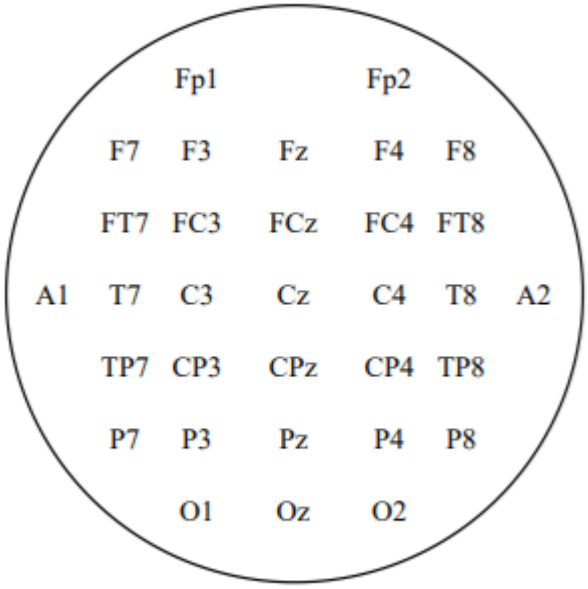

Figure 2.2. EEG 10-20 electrode position system.

Figure 2.3. EEG device (Koelstra et al., 2012).

2.3. DEAP Dataset

Our proposed study examined on the DEAP dataset. DEAP dataset was presented in 2012 for human affective state analysis (Koelstra, 2012). It is a multimodal dataset which includes 32 channels EEG signals and 8 peripheral signals channels (EOG, EMG, Respiration, GSR, Plethysmography, Temperature) also some of the

participants' facial response video recordings. While the dataset was created, a specially designed experiment was presented for total 32 participants (16 males and 16 females). Each participant watched 40 music videos that were intentionally selected to induce dimensions of emotion's circumplex model. Duration of the music videos were 60s. Since the selected videos are music videos, both auditorial and visual stimulus were used during emotion induction. For each participant recorded data file includes 40x40x8064 matrix which is named as 'data' and 40x4 matrix which is named as 'label'. For data matrix, first dimension represents the video number, second dimension represents the channel, and third dimension represents the data length. For the label matrix, 40 represents the video numbers and 4 represents the labels.

Table 2.1. Data file description.

Name	Shape	Contents
Data	40x40x8064	video number x channel x data
Labels	40x4	video number x label

While the participants watched the videos, simultaneously they rated their feeling's scale for the videos choosing the scores between 1-9 by using SAM (Self assestment manikin). Since SAM does not require verbal expression it provides valuable fastness hence provides the continuous experiment trial without delays. SAM is picture based rating system and provides rating scores for valence, arousal and dominance labels. In DEAP database liking label also included to this rating system via thumbs are up or down options. Therefore, ratings were recorded for 4 different labels as follows valence, arousal, dominance, liking.

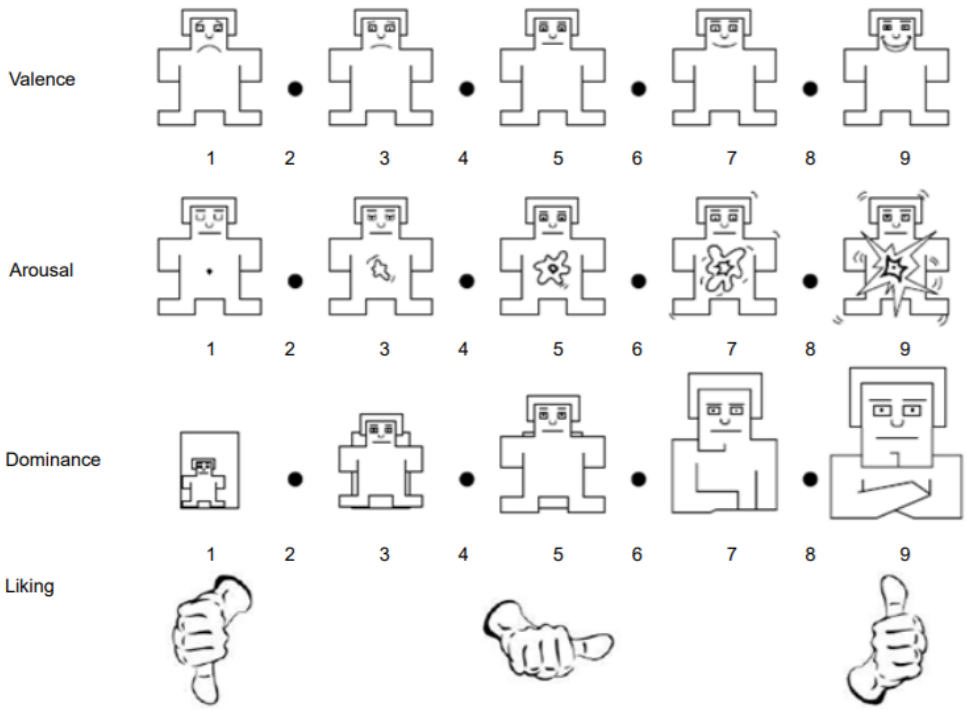

Figure 2.4. The Self-Assessment Manikin (SAM) rating pictures the affective dimensions of valence (top line), arousal (second line), dominance (third line), liking (bottom line) (Koelstra et al., 2012).

Enabling both the physiological peripheral responses and subjective views provided an extensive study for scientists. During the recordings of signals, Biosemi active 2.0 EEG device was used at 512 Hz sampling frequency. It is a non-invasive clinical EEG device and has 32 channels. After the recording, band-pass filtering and re-referencing in common average mode was examined for preprocessing and preprocessed data presented in the dataset for the usage. For our thesis study, this extensive database was selected, and we examined two different approaches on this dataset for arousal and valence scale classifications.

CHAPTER 3
EEG PROCESSING AND MACHINE LEARNING

3.1. Frequency and Time-Frequency Domain Transforms

Signal processing methods are commonly used in various fields of science since last century. Traditional transform methods such as Laplace and Fourier transform are most widely used spectral analysis methods of digital signal processing. Furthermore, to provide time-frequency domain features Short Time Fourier Transform and Wavelet Transform were proposed. Therefore, with these analyses, characteristic or abnormal patterns can be obtained, and these patterns provide a great potential when they joined with various machine learning studies.

3.1.1. Fast Fourier Transform

Fast Fourier transform is one of the signal analysis method and basically derived from the discrete Fourier transform (DFT). Computational complexity of DFT is reduced with FFT and transform process became faster therefore FFT is accepted as a fast computational algorithm of DFT. DFT algorithm of time signal $x(j)$ is,

$$FFT\,(k) = \sum_{j=1}^{n} x(j) * W_n^{(j-1)(k-1)} \qquad k = 0, \ldots . n. \qquad (1)$$

where $W_n = exp^{(-2\pi i)/n}$

n is roots of the unity.

Computation of this DFT equation wit n points sequence includes O (n^2) of multiplications and additions on the other hand FFT algorithm computes the DFT using O ($n \log n$) numbers of operation so this provides a significant speed advantage especially if n is a big number (Heckbert, 1998).

3.1.2. Short Time Fourier Transform (STFT)

Short-time Fourier transform (STFT) is another spectral analysis technique. It calculates Fourier transforms of portioned forms of signal with sliding window. STFT provides the frequency resolution in time domain, while FFT provides frequency resolution in frequency domain. STFT equation for time signal $x(t)$ is defined as,

$$STFT\,(t,\,\omega) = \int x(t')\,w(t'-t)\exp\{-j\omega t'\}dt' \qquad (2)$$

As it can be seen from the above equation, main difference between the FFT and STFT is w(t) which is called as window function. This function provides the resolutions of small pieces of signal. This definition provides a shifting operation on signal and divides it into sequential portions; thus, it can be seen the transformations of the small-time durations of signal. Additionally, magnitude squared of STFT calculation gives us spectrogram images. It is important to choose a proper window size since if it is chosen too big or small for the pattern of the signal, meaningful data can be missed (Gonzalez, 2007).

3.1.3. Wavelet Transform

Another type of time-frequency analysis method is wavelet-transform. The Wavelet-transform splits the signal into its frequency components and computes each of them with their pattern resolutions in their scales. Hence, it provides multiresolution information. Wavelet analysis creates wavelets from mother wavelet in different scales and translations. To express a function or image as a linear combination of the wavelets and scaling function discrete wavelet transform (DWT) uses those wavelets. Besides variation of the wavelets play an important role capturing of different scaled patterns. Wavelet-transform differs from STFT in the point of varying scales windows. In our study, 2D discrete wavelet transform was computed, it can be seen in the following equation,

$$\begin{aligned} W_{\Phi}\,(j_o,\,k) &= \frac{1}{\sqrt{M}}\sum_t x(t,z)\,\Phi_{jo,k}\,(t,z) \\ W_{\Psi}\,(\mathrm{j},\,k) &= \frac{1}{\sqrt{M}}\sum_t x(t,z)\,\Psi_{j,k}\,(t,z) \end{aligned} \qquad (3)$$

where $W_{\Phi}\,(j_o,\,k)$ and $W_{\Psi}\,(\mathrm{j},\,k)$ represent scaling and wavelet coefficients, respectively. They are determined by the inner product of the 2D function or $x(t,z)$ with the scaling function and wavelets at that specific scale and translation. The set of scaled and translated functions are called as scaling function. These are basis functions of composed translations and scalings in wavelet transformation. Main wavelet which is called as mother wavelet and its variously scaled functions can be seen in the following two equation

(Gonzalez, 2007).

$$\Psi^{i}_{j,m,n}(t,z) = 2^{\frac{j}{2}}\Psi^{i}(2^{j}\mathrm{t}-\mathrm{m}, 2^{j}\mathrm{z}-\mathrm{n}) \tag{4}$$

$$\phi_{j,m,n}(t,z) = 2^{\frac{j}{2}}\phi(2^{j}\mathrm{t}-\mathrm{m}, 2^{j}\mathrm{z}-\mathrm{n}) \text{ , where I = \{H, V, D\}} \tag{5}$$

2D wavelet transforms are given as,

LL= $\phi(t,z) = \phi(t)\phi(z)$

LH= $\Psi^{H}(t,z) = \Psi(t)\phi(z)$

HL= $\Psi^{V}(t,z) = \phi(t)\Psi(z)$

HH= $\Psi^{D}(t,z) = \Psi(t)\Psi(z)$

Figure 3.1. describes the wavelet decomposition in 2D. After applying the suitable wavelet and scaling filters, diagonal, vertical, horizontal and approximation coefficients are created in one level of decomposition.

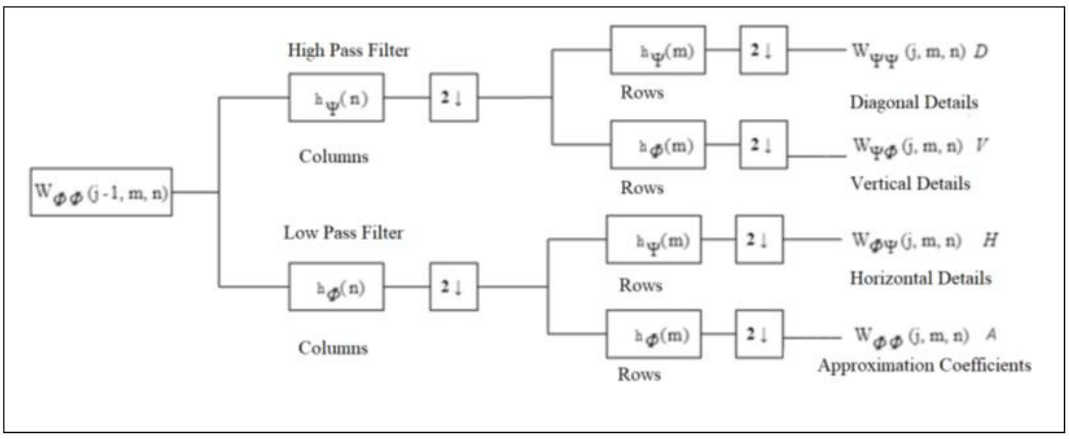

Figure 3.1. Decomposition steps of wavelet transformation.

Because of the downsampling operation at each step size of the matrix will be reduced into its half at both directions. Thus after decomposition the coefficents can fit into the original matrix as shown in Figure 3.2.

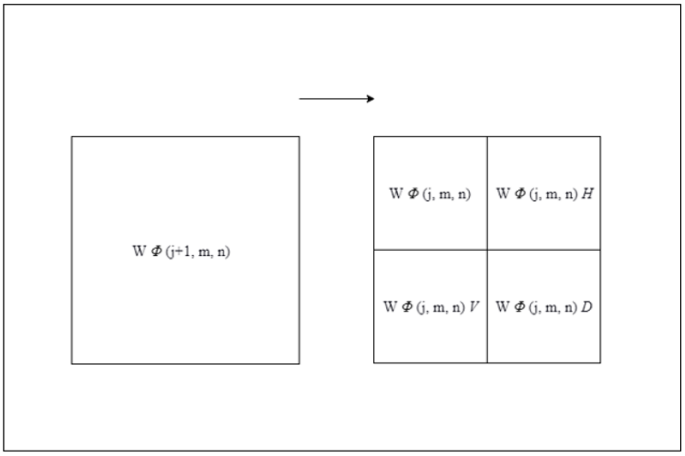

Figure 3.2. Resulting decomposition of wavelet transformation.

In our study, two level of decomposition of raw 40 channel of EEG signals was obtained with Symlet 4 wavelet and then reconstructed into single image to feed CNN.

3.2. Convolutional Neural Network (CNN)

Convolutional neural networks are one of the popular neural network algorithms in nowadays. It was inspired from human vision system. It can extract valuable abstract features. CNN reduces the complexity of systems and improves generalization. CNN algorithm may include one or more convolutional layers, pooling layers, dropout layers, batch norm layer, fully connected layers and softmax layer in its structure. Algorithm starts with convolutional layer, in this layer after the convolution operation, extracted feature maps go through the pooling layer. Convolution operation is made with different sized and different number kernels and so it provides filtering of the feeded inputs of the network and as model continue to train, in each step of the training these features become to include related patterns of the inputs. Kernel is slides with specified intervals on the image or feature map layer and these intervals are named as “strides”. For each stride step weighted sum of the filtering operation is calculated and written as the convolution result on the feature map. This process reduces the size of the features. However, in some implementations feeded input size might be wanted to protect for such cases “padding” operation is used. This operation puts zeros around the feature map matrixes and convolution operation is executed on this padded form, with this way at

the end of the convolution operation result size can be stay same with the input. A sample convolution operation was given in Figure 3.3.

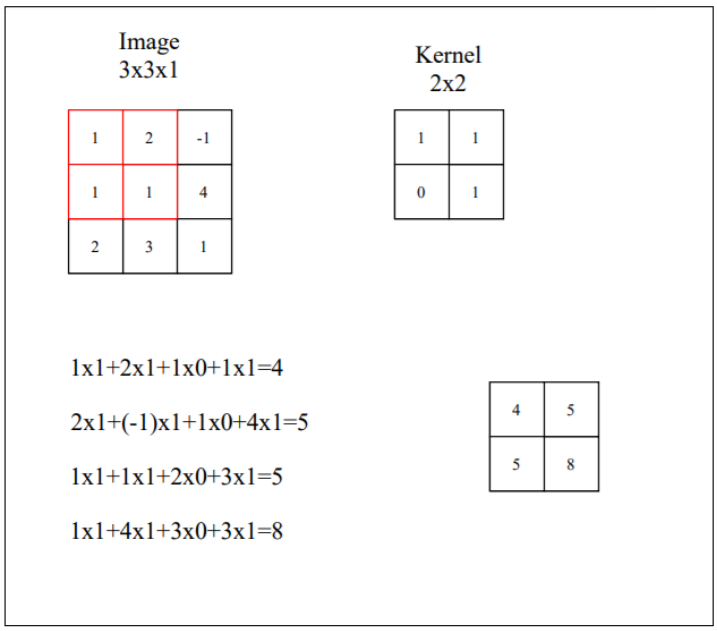

Figure 3.3. Convolution operation.

After the convolution operation, feature maps are applied to Relu activation function. The reason behind the usage of this process is introducing the non-linearity to the model (Bengio, 2016). After this layer feature maps feeded through the pooling layers. Pooling layers are used to reduce the dimensions of the feature map in width and height scales. This process is used to decrease the computational complexity and overfitting since it reduces the size of the feature maps. Size and stride of the pooling operation is adjustable it can be executed in various parameters. An example of max pooling operation was given in Fig. 3.4.

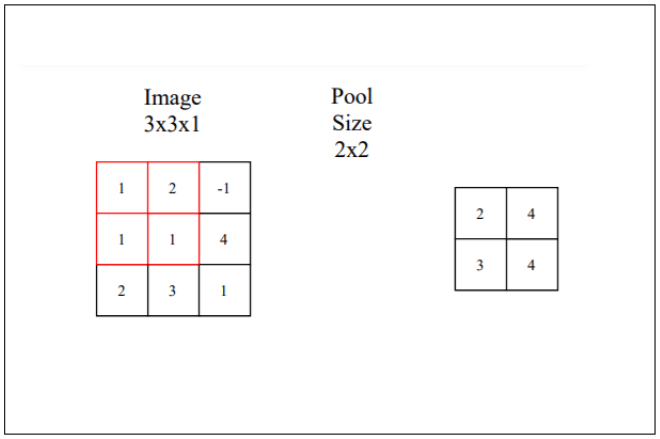

Figure 3.4. Max pooling operation.

Another CNN layer is dropout layer. Dropout layers deactivate some neurons during training with certain probability values. It prevents model to become overfit. Batch normalization layer is another CNN layer. While model is in the training stage, input distributions change hence it is needed to adjust it, batch normalization is a step that calibrates each layer's inputs.

The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer. Finally, softmax layer provides keeping total probability distribution summation to 1 hence limits the outputs in a specific range (Li et al., 2016).

CHAPTER 4
EEG EMOTION CLASSIFICATION WITH CNN

This chapter summarizes the EEG emotion classification studies accomplished with CNN. The first part includes an emotion classification experiment implemented with single electrode measurements and the results of classification using spectrogram features. The second part expands spectrogram study into multiple electrode recordings from DEAP dataset. Finally, the last part introduces the valance and arousal state classifications with CNN using DEAP dataset by using different feature sets such as FFT and wavelets compared with raw data.

4.1 Single Electrode Spectrogram Emotion Classification

As it is mentioned in Chapter 1, emotion classification is one of the important areas of HMIs. For this purpose, various methods are examined in the literature. During our emotion classification research, we implemented single channel electrode emotion classification task. A special experiment was designed where the participants watched a video, meanwhile EEG signals were recorded with single channel Neurosky headset (Donmez et al., 2019). Then STFT of segmented EEG signals were obtained for labeled time durations of the emotional scenes. After that extracted images classified in pretrained Googlenet CNN.

4.1.1. Materials and method

It was stated in Chapter 2 emotion mainly defined with two titles, in this implementation discrete emotion theory followed and specially edited video was chosen. The video was created for another study in the literature and three emotion elicitation examined which are fear, sadness and fun. 10 participants involved in the experiment all of them were female and ages varied between 24-33. Neurosky Mindwave Mobile 2 single channel EEG device is used in order to collect EEG data as shown in Fig. 4.1.

Figure 4.1. Neurosky Mindwave Mobile 2.

Experiment was applied in specially isolated room and participants are informed to do not make any body movement. EEG device had a single dry-sensor electrode attached to the forehead at position frontal polar - left hemisphere (Fp1) position according to 10-20 electrode placement system. Fp1, plus a reference electrode on the ear clip, having a sampling rate of 512 Hz was used to record raw EEG data. Neurosky's thinkgear chip library for MATLAB was used to take raw data.

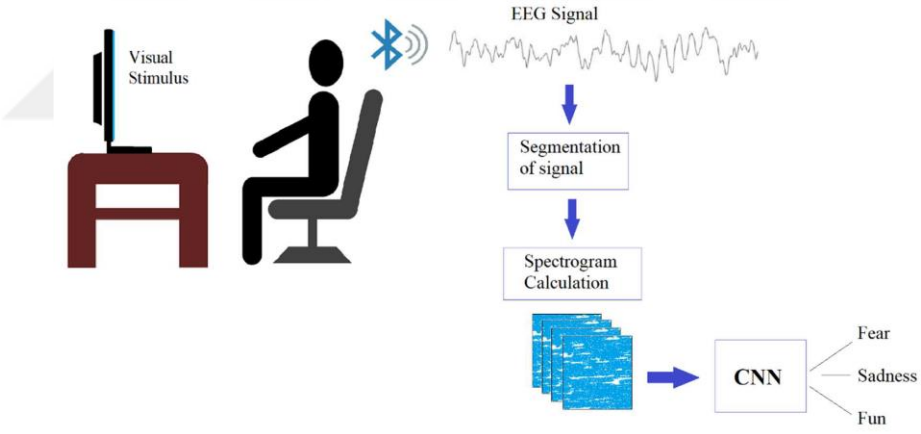

Figure 4.2. The flow chart of single channel electrode used EEG.

The diagram of the proposed system is given in Fig. 4.2. After the recording data was segmented and segmented portions transformed with STFT. Spectrogram images extracted and resized in 224x224x3 size form. In the next step, extracted data feeded the pretrained CNN network Googlenet in MATLAB with deep learning toolbox. Googlenet includes 22 deep layers, 9 inception modules and it was trained on subset of the ImageNet database. During the training learning rate used as 0.0001 and model

trained with 20 epochs. Stochastic gradient descent used as an optimizer. The layers of GoogleNet is given in Fig. 4.3.

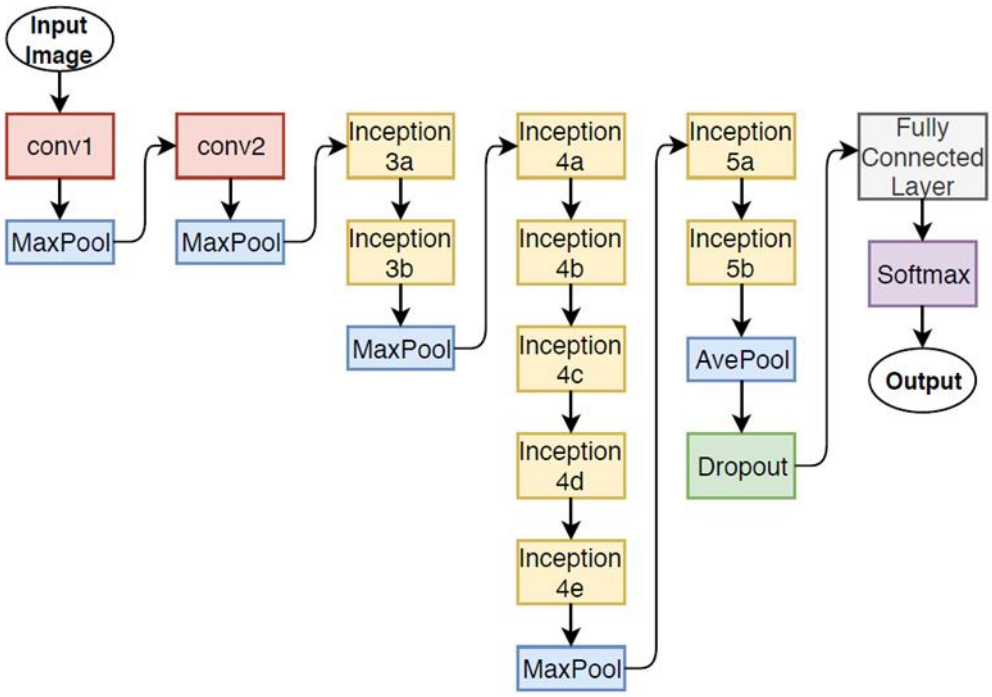

Figure 4.3. The structure of Googlenet.

Figure 4.4 show a segment from the raw signal. As shown in the graph, blink signals are visible in the recording.

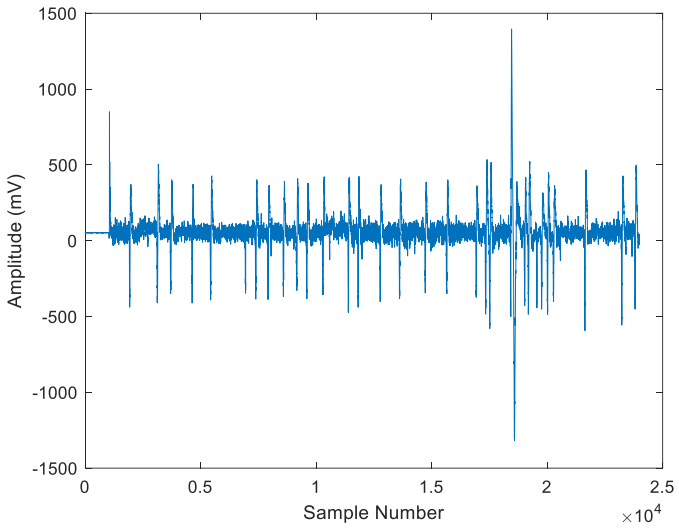

Figure 4.4. Raw signal plot.

Data was segmented for spectrograms according to time duration of each emotion in

the video. Approximate sample numbers calculated for three of the emotions by checking the sampling frequency and durations. MATLAB's spectrogram comment used for this purpose, it used with default parameters so signals were divided into 8 windows. A sample of the obtained spectrogram is shown in Fig. 4.5.

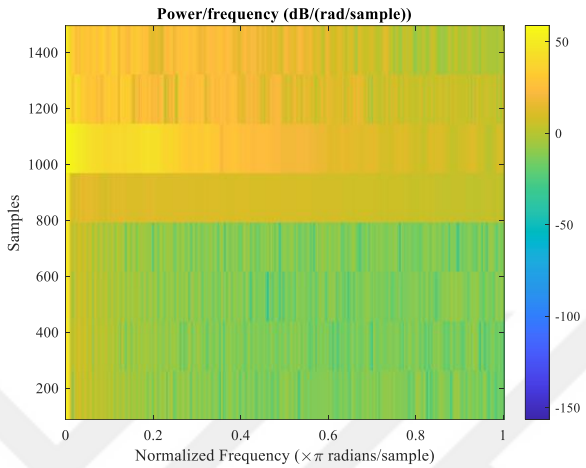

Figure 4.5. Spectrogram image.

4.1.2. Results and Discussion

After STFT extractions, each participant contributed with 49 spectrogram images and so in total 490 created for study. 80% of images was used for training and 20% of it used for testing. CNN executed in MATLAB with 20 epochs and 26 iterations per each epoch.

Classification performance was evaluated in terms of accuracy sensitivity, specificity which are defined as

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN}*100 \quad (6)$$

$$Sensitivity=\frac{TP}{TP+FN} \quad (7)$$

$$Specificity=\frac{TN}{TN+FP} \quad (8)$$

where TP (True positive) indicates that the correct prediction of given class. TN (True negative) states that actual emotion does not belong to given class and the prediction is not the given class. FP (False positive) is the number of instances which

are not classified correctly, where prediction is the given class and actual emotion does not belong to the given class. Similarly, FN (False negative) occurs when prediction is not the given class and actual emotion belongs to the given class.

Figure 4.6 summarizes the results of the tests as a confusion matrix. As can be seen, 84.69% overall training accuracy was accomplished

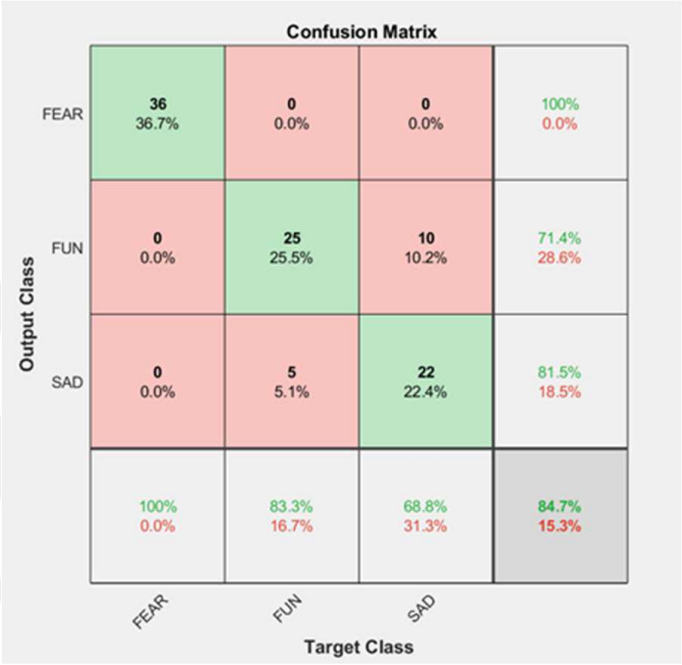

Figure 4.6. Confusion matrix of implementation.

Whole performance metrics related with the emotional states is given in Table 4.1.

Table 4.1. Performance measurements for classification.

Emotional State	Accuracy	Sensitivity	Specificity
Fun	0.84	0.83	0.85
Fear	1.0	1.0	1.0
Sad	0.84	0.68	0.92

According to the results, it was indicated that fear is an intense and generally includes common reactions for people. Although, the number of participants is limited, this study shows that single electrode carries information about emotional state. The study can be expanded with more participants and changing the video with

more proper for academic usage to induce emotions. Additionally, self-evaluation systems thought as should be placed in the future studies also.

4.2 Multi Electrode Spectrogram Emotion Classification

After the previous study, it was concluded that single electrode was limited for extracting big amounts of data, so it was required to use multichannel EEG device usage. However, since 2020 March, world have been struggling with Covid-19 pandemic distribution meanwhile social distancing and isolation became one of the prerequisites for people to survive because of this situation. We couldn't make our own data recordings with multichannel electrodes. However, a suitable dataset for this purpose in the literature was searched and DEAP dataset was chosen. It was the one of the commonly used dataset of emotion classification studies and provides 40 biosignals.

For this task, spectrogram feature extraction made to DEAP dataset in the same manner that was applied for the single electrode implementation. Since participants made a rating choice between 1 - 9 for their experiences during the recording as it is pointed out in dataset description in Chapter 2, these ratings were followed. 5 was selected as threshold and any value greater than 5 is assumed as valence or arousal and any value less than 5 or equal is assumed as not valence and not arousal. So, there is one binary classifier for arousal and one binary classifier for valance. All recordings segmented with 3.75 s windows which is heuristically found that optimal in computational sense. 40 channel recording of 32 participants of the DEAP dataset was included in our study.

Since total length of the data had 8064 samples and total duration was 60s, after the segmentation each window had 504 samples. MATLAB's spectrogram tool is used, it calculates the spectrogram with 8 default segments for windowing, we did not change this and followed it. Spectrogram image size is 224x224x3, its model's input has 224x224x3 dimension. Training image numbers varied for every participant since we define 5 as the upper boundary limit for rating results of each participant. However, for each participant, extracted feature images splitted 75% for training, 15% for validation and 10% for testing executions.

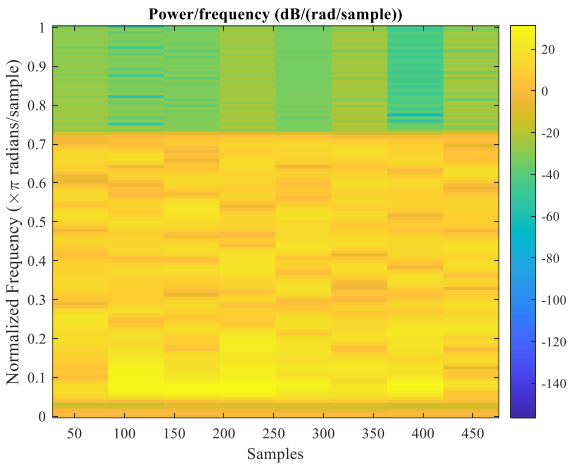

Figure 4.7. Spectrogram image sample.

The last block of the system is the binary classification of the constructed images. Six CNN models were created and executed for the classification purpose. All models were following the same main order for different layers (convolutional + relu layer, dropout layer, max pooling layer) and 3x3 kernels. However, four of them had three convolutional layers with (4-4-4), (8-8-8), (4-8-16), (16-16-16) numbers of kernels in each layer and rest of the two models had four and five convolutional layers with (4-8-16-32) and (4-8-16-32-64) numbers of kernels. After the execution of these models, it was seen that the model that had (16-16-16) numbers of kernels with three convolutional layer model has relatively better performances than others. Due to this fact in this thesis three convolutional layers with (16-16-16) numbers of kernels used structure selected. In Figure 4.8, selected CNN structure is presented.

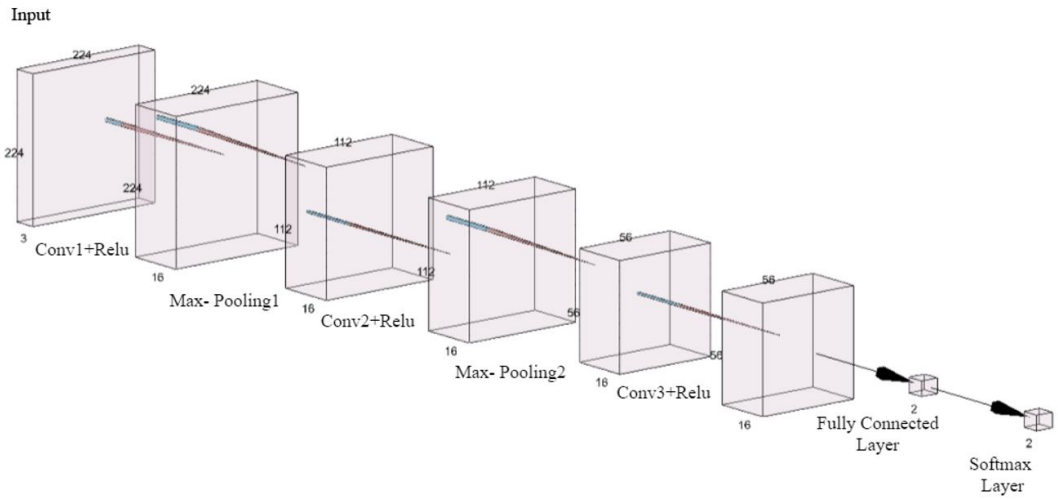

Figure 4.8. CNN architecture for spectrogram inputs feeded tasks.

For this task, minimum computation time was approximately 45 minutes and maximum 160 minutes. Results for most of the subjects were close to each other and presented in Table 4.2 and Figure 4.9. For some subjects such as S22, a high performance such as 79.23% and 79.36% for both valance and arousal were obtained. But performance changes from subject to subject. When compared to literature, we have observed that spectrogram does not contributes much considering its computational complexity. Since the signal is highly nonstationary and variable even for one subject constant window size might be the cause of this low accuracy. Additionally, since spectrograms are the square of the STFT's magnitude, phase information is also lost. Adding phase as another matrix, on the other hand, would increase the complexity.

Table 4.2. Spectrogram features used valence and arousal classification accuracy results for all participants.

Spectrogram	**Valence**	**Arousal**
S1	52,71	57,46
S2	65,57	67,58
S3	62,86	65,87
S4	73,11	76,32
S5	64,11	57,15
S6	61,10	61,67
S7	58,32	57,95
S8	57,72	59,80
S9	59,60	64,48
S10	58,80	62,29
S11	75,77	73,92
S12	63,02	62,70
S13	63,85	66,54
S14	58,98	62,82
S15	54,36	56,45
S16	61,37	54,84
S17	60,02	61,95
S18	64,45	61,37
S19	60,89	62,01
S20	62,65	59,33
S21	55,44	64,07
S22	79,23	79,36
S23	57,36	56,57
S24	64,06	69,78
S25	68,33	73,94
S26	62,63	60,96
S27	59,95	60,19
S28	65,73	64,80
S29	59,94	59,78
S30	60,06	55,20
S31	60,05	54,82
S32	55,99	62,45
Average	62,13	62,95

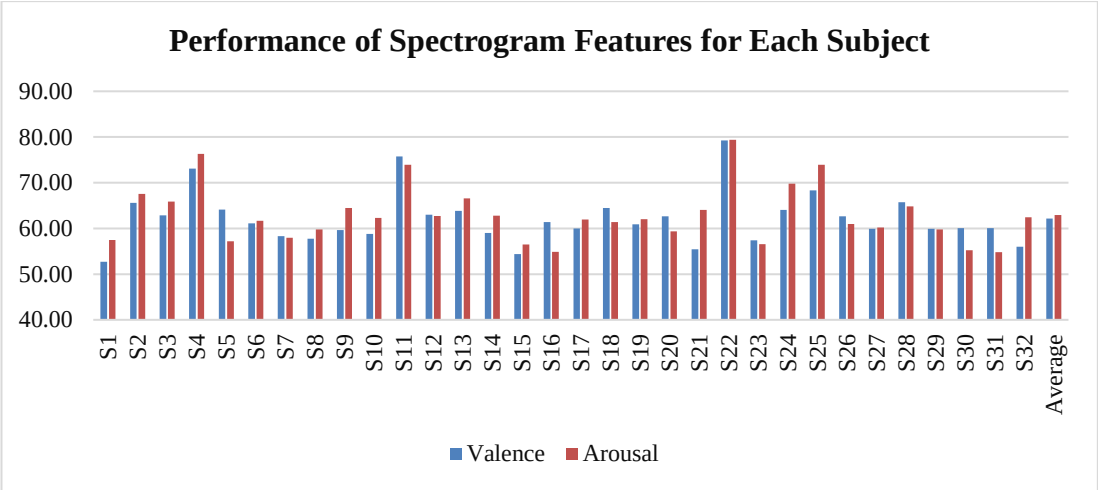

Figure 4.9. Spectrogram performance results for each subject.

4.3 Emotion Classification for DEAP Dataset with CNN

Although spectrogram was accomplished the classification, its computational efficiency is very high. For this reason, more efficient methods researched and fundamental signal processing methods such as FFT and wavelet transform were selected to examine other tasks.

This section describes the proposed emotion classification system which uses CNN fed with raw signals, FFT, and wavelets. As can be seen in Figure 4.10, the input is 40 channel raw recordings for 32 participants and their emotional labels of valance and arousal.

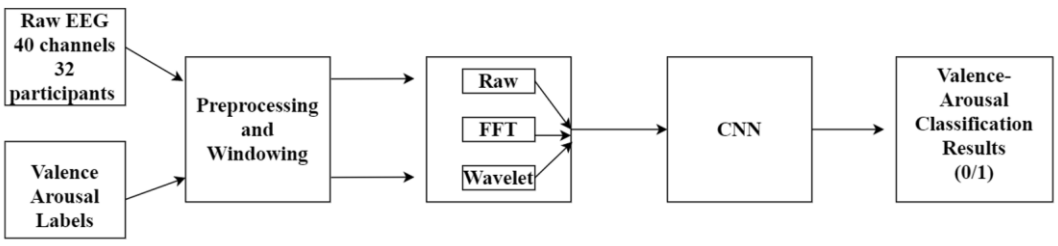

Figure 4.10. The block diagram of emotion classification system.

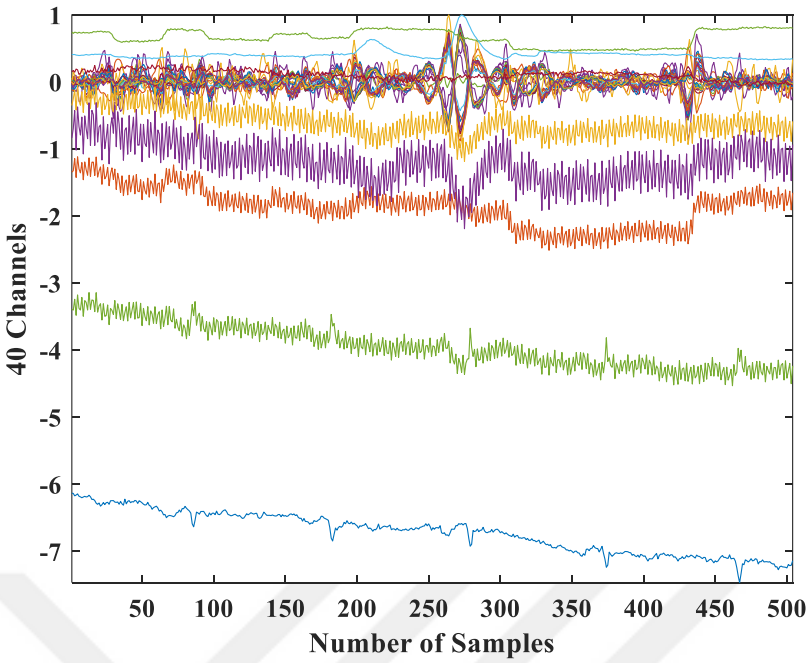

Figure 4.11. Raw signal sample.

A sample from windowed raw EEG signals are illustrated in Figure 4.11. The second block in Figure 4.10 constructs the image for the classifier. The first mode is to construct a matrix from windowed raw EEG recording of size 40x504 without applying any transformation. In second mode, FFT of raw signals were calculated for each channel. Then, the magnitude of FFT coefficients were put together to construct a matrix of size 40x504. Figure 4.12 and Figure 4.13. shows a sample of magnitude of FFT coefficients and constructed FFT image.

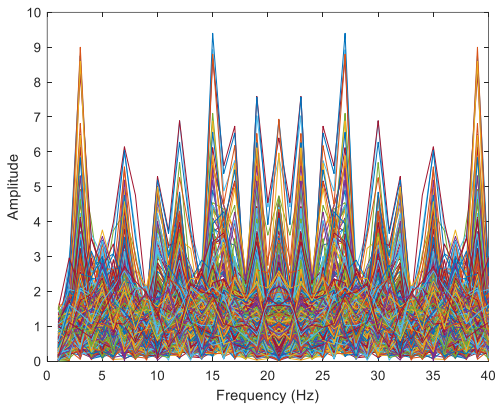

Figure 4.12 Sample of FFT coefficients magnitudes for channels.

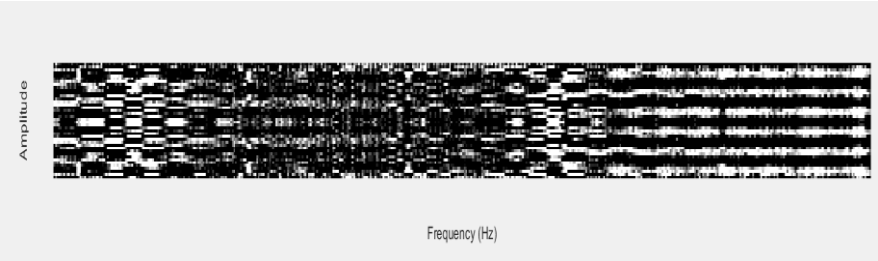

Figure 4.13 FFT image for classification

The final image transform used for features is wavelet transform. In this application, it was seen that Symlet 4 wavelet is commonly used for image classification algorithms and similarly best result were obtained for this mother wavelet. Then, the 2D wavelet transform of 40x504 raw EEG matrices were calculated. As a result four matrices approximation, horizontal detail, vertical detail and diagonal detail of 20x252 were obtained as in Figure 4.14.

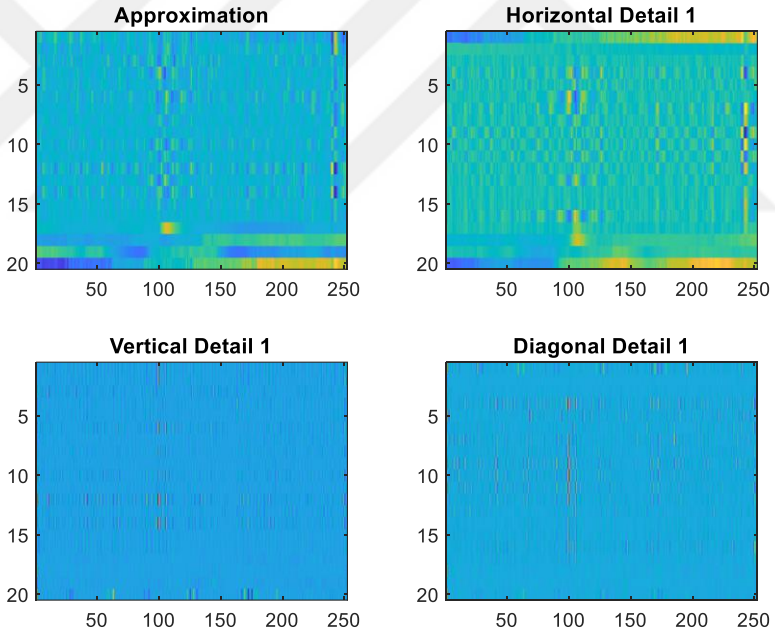

Figure 4.14. A sample from first level decomposition of EEG with Symlet 4 mother wavelet.

The first level of approximation matrix was then decomposed again to obtain for second level wavelet coefficient matrix of 10x126 as can be observed in Figure 4.15.

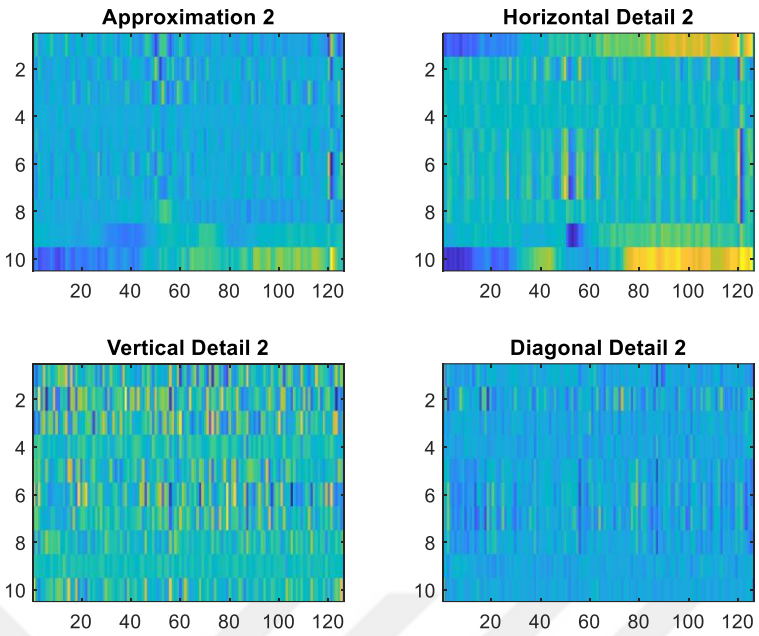

Figure 4.15. Second level decomposition of same EEG with Symlet 4 mother wavelet.

The matrices of sizes 20x252 for level 1 and 10x126 for level 2 were combined together to obtain 40x504 matrix as described in Chapter 3. This construction is illustrated in Figure 4.16. A, H, V and D stands for approximation, horizontal detail, vertical detail and diagonal detail respectively, where the number shows the decomposition level. The matrices were then saved as png images.

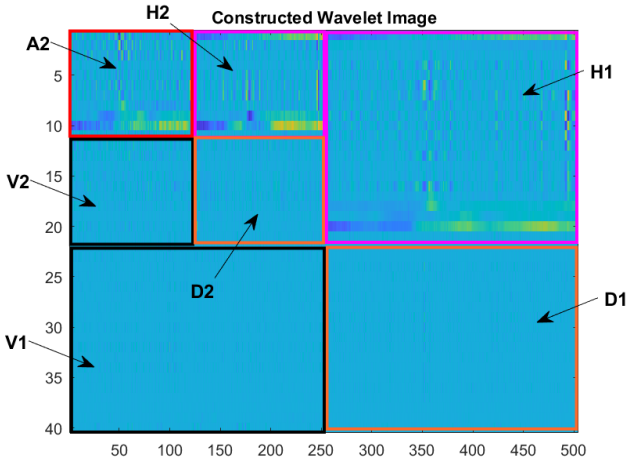

Figure 4.16. Wavelet decomposition illustration.

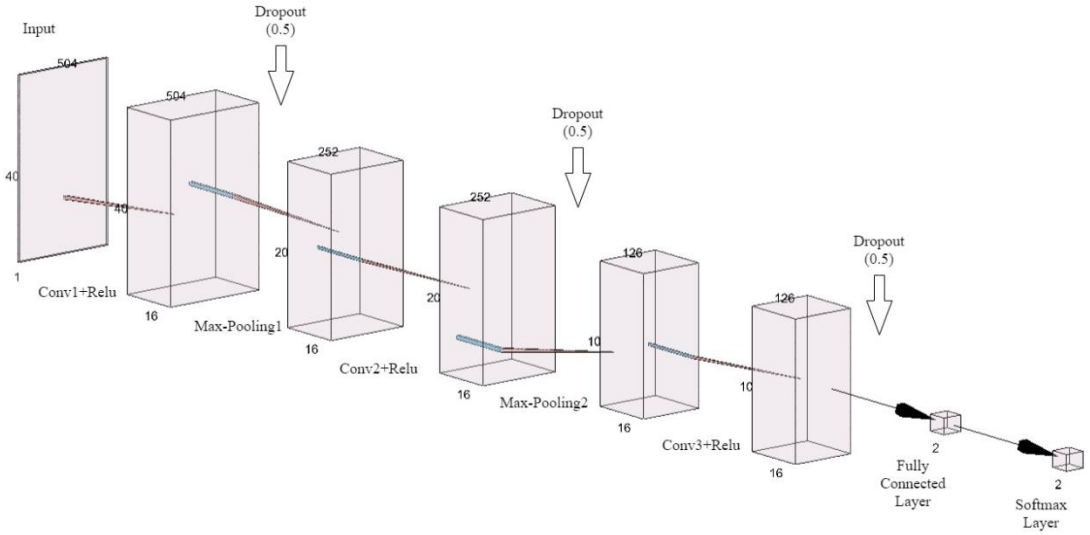

Figure 4.17. CNN architecture for raw, FFT and wavelet transform inputs.

In Figure 4.17, raw, FFT and wavelet transform inputs used CNN model structure can be seen. As we mentioned previously, for these methods each of the sample sizes was 40x504x1 and so input size for model is 40x504x1. In both usages of our model, after the convolutional layers, dropout layer structured also. As it is mentioned in Chapter 3, these layers help model to prevent overfitting. Random dropping ratio was used as 0.5 in the model execution. Forwardly max pooling operation applied to reduce dimension and extracts spatial features. Finally, fully connected layer and softmax layer placed and classification examined. While we create the model, MATLAB's Deap Learning Toolbox was used. Learning rate chosen as 0.001 and ADAM optimizer used. Model compiled with 100 epochs in MATLAB and performed on Nvidia GTX1050. Longest execution time was nearly 2 minutes on the other hand slowest one 1 minute. The reason behind this huge difference between the previous case and this case is the dimensions of the created inputs from the methods. Spectrogram images training durations took relatively more time. Since algorithm includes randomness, to get general results, each task executed 10 times and averages values calculated, in the upcoming tables these results are presented detailly. Until now, we briefly introduced the methods, database and model which were used in this study. In the following figures and tables classification results are presented for different tasks.

4.3.1. Results for Raw Features

In the Table 4.4 and Figure 4.18, two classes overall classification performance results are presented for each participant for raw signals. It can be said with respect to these results, maximum and minimum accuracy results are close for each subject. This study proved that raw EEG signals have valuable performance results. The results obtained from the analysis of this implementation task also shows our CNN model's performance individually since only feature extraction process is applied in CNN structure. As can be seen from the table, it can be stated that patterns of the classes captured by our CNN model.

Table 4.3. Raw EEG signal features used valence and arousal classification accuracy results for all participants.

Raw Signal	Valence	Arousal
S1	68,50	70,58
S2	70,51	60,00
S3	70,87	69,23
S4	74,42	71,53
S5	79,03	68,51
S6	73,13	70,00
S7	75,81	63,76
S8	72,93	67,42
S9	78,30	60,77
S10	75,31	54,30
S11	76,34	61,05
S12	63,84	62,26
S13	74,64	70,50
S14	75,95	62,14
S15	80,00	72,80
S16	73,13	72,36
S17	73,62	72,31
S18	79,80	69,60
S19	76,86	70,94
S20	68,20	62,50
S21	67,83	80,39
S22	71,38	77,11
S23	76,83	73,45
S24	66,21	69,53
S25	60,00	70,55
S26	68,88	62,97
S27	70,32	85,96
S28	75,22	62,40
S29	77,19	70,02
S30	63,08	65,17
S31	65,92	68,36
S32	75,80	71,65
Average	72,50	68,44

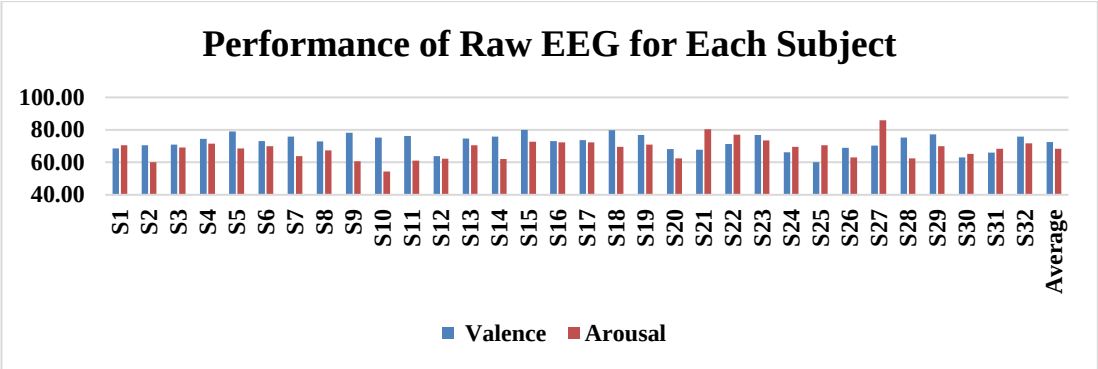

Figure 4.18. Raw EEG signal performance results for each subject.

Additionally, ROC performances of raw features are drawn for valence and arousal classes and can be seen in Figure 4.19, Figure 4.20, Figure 4.21, Figure 4.22 also, performance measurements specificity, sensitivity and accuracy calculated for two participants are given in Table 4.4 and Table 4.5 For raw features, since area under the ROC curves for all iterations and overall average is pretty high, it has good classification performance.

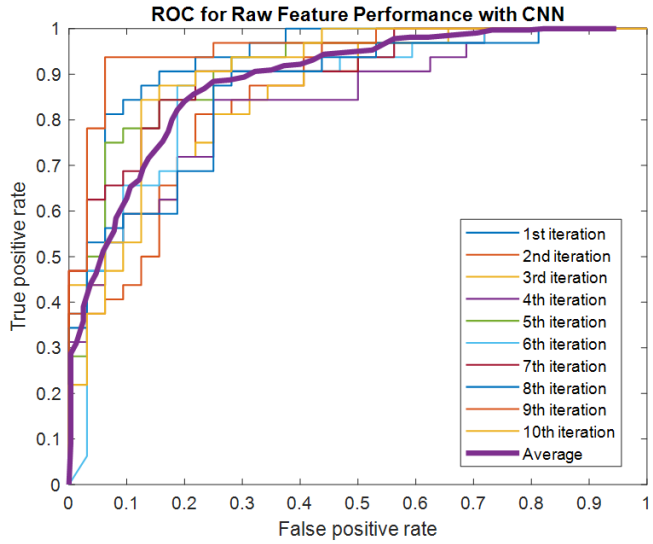

Figure 4.19. ROC for raw feature performance of subject 15 (valence class).

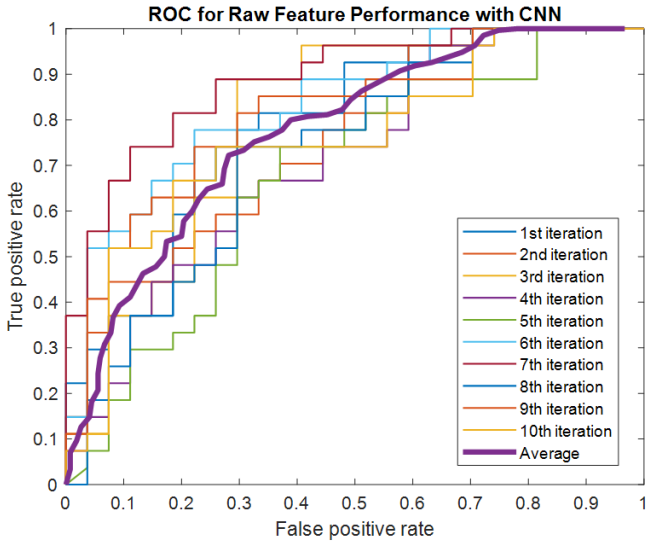

Figure 4.20. ROC for raw feature performance of subject 20 (valence class).

Table 4.4 Performance measurements for Subject 15 and Subject 20 (valence class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,82	0,79	80
S20	0,66	0,73	68,2

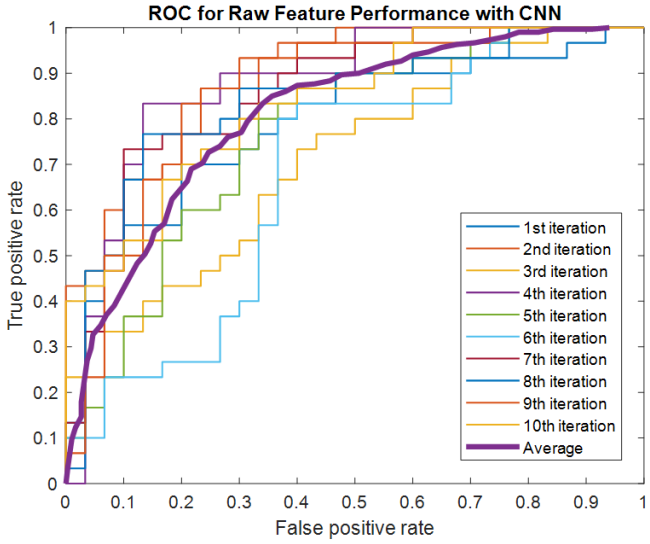

Figure 4.21. ROC for raw feature performance of subject 15 (arousal class).

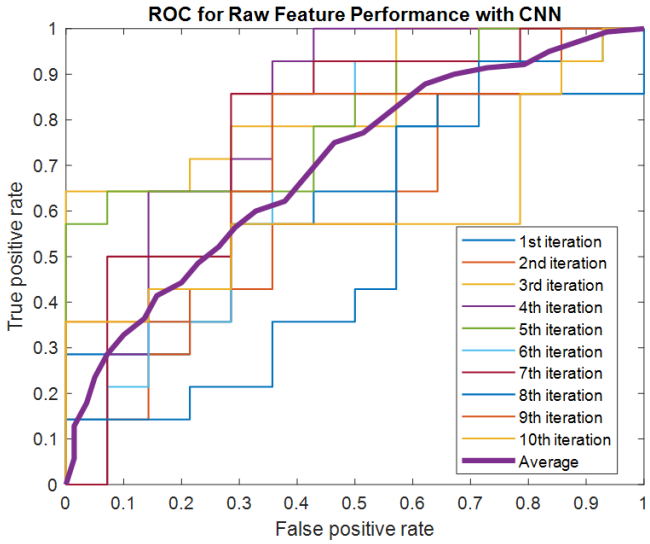

Figure 4.22. ROC for raw feature performance of subject 20 (arousal class).

Table 4.5. Performance measurements for Subject 15 and Subject 20 (arousal class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,73	0,74	72,8
S20	0,62	0,65	62,5

4.3.2. Result for FFT Signals

In the second implementation, FFT image features are used. According to these results which is FFT features used are given in Table 4.6 and Figure 4.23, FFT feature performance probably in mid-level successful accuracies both for the individual and overall average. Also, subject's maximum and minimum results are varying in a wide range for both classes. The reason behind the why FFT did not performed well high probably related with the fact that EEG signals have non-stationary characteristics which varies with time but FFT algorithm is computed by assuming the signal samples continue infinitely and steadily in periodic order and loses the time information.

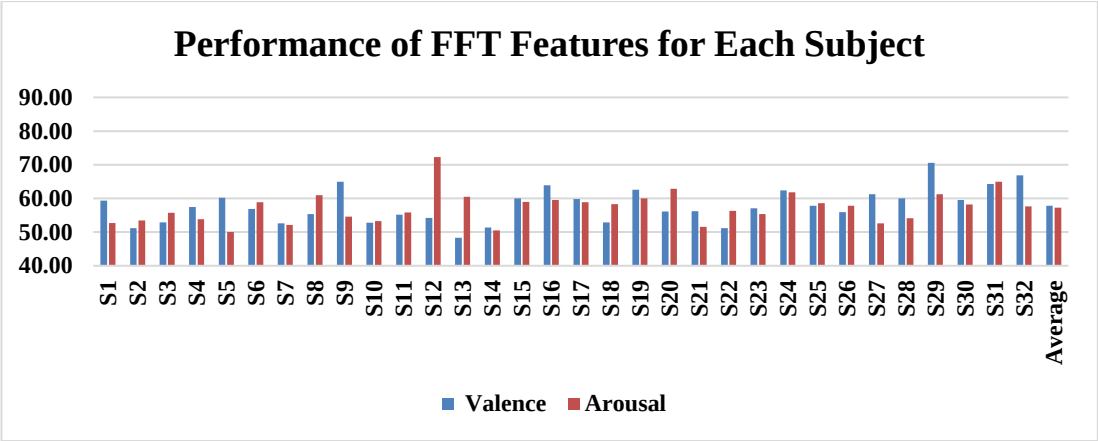

Figure 4.23. FFT features performance results for each subject.

Table 4.6. FFT features used valence and arousal classification accuracy results for all participants.

FFT Features	Valence	Arousal
S1	59,34	52,71
S2	51,20	53,47
S3	52,93	55,76
S4	57,48	53,85
S5	60,18	50,00
S6	56,87	58,89
S7	52,63	52,08
S8	55,34	60,95
S9	65,00	54,62
S10	52,81	53,27
S11	55,18	55,85
S12	54,18	72,27
S13	48,32	60,50
S14	51,40	50,47
S15	60,01	59,00
S16	63,96	59,53
S17	59,83	58,86
S18	52,90	58,34
S19	62,61	60,00
S20	56,11	62,90
S21	56,17	51,55
S22	51,21	56,35
S23	57,04	55,33
S24	62,39	61,79
S25	57,83	58,61
S26	55,90	57,81
S27	61,28	52,61
S28	60,01	54,14
S29	70,56	61,27
S30	59,52	58,18
S31	64,27	65,00
S32	66,86	57,63
Average	57,85	57,30

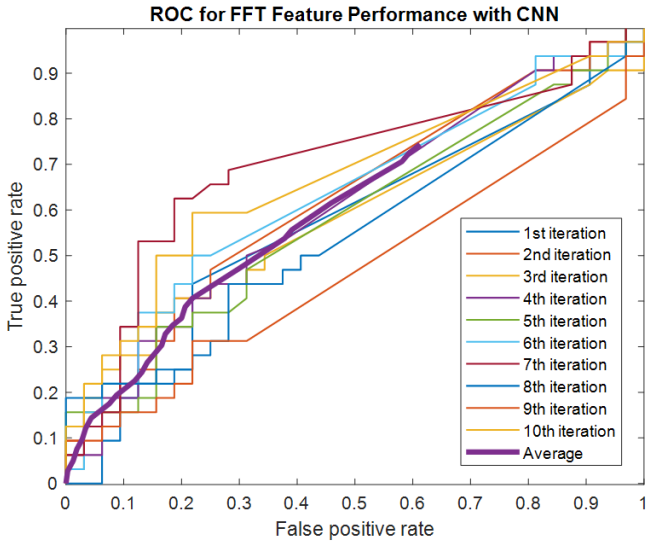

Figure 4.24. ROC for FFT feature performance of subject 15 (valence class).

ROC performances for FFT features are drawn for valence and arousal classes presented in Figure 4.24, Figure 4.25, Figure 4.26, Figure 4.27. Moreover, performance measurements specificity, sensitivity and accuracy calculated for two participants can be seen in Table 4.7 and Table 4.8. Areas under ROC curves for each iterations and overall averages are smaller if it is compared with the raw signal feature results and also, they are almost in mid-level diagonal line, this indicates performance was not held as much as it was expected to be.

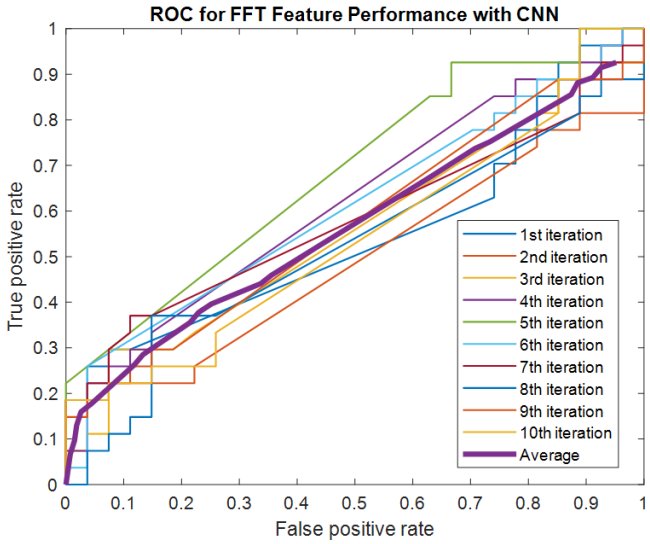

Figure 4.25. ROC for FFT feature performance of subject 20 (valence class).

Table 4.7. Performance measurements for Subject 15 and Subject 20 (valence class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,58	0,64	60,01
S20	0,55	0,63	56,11

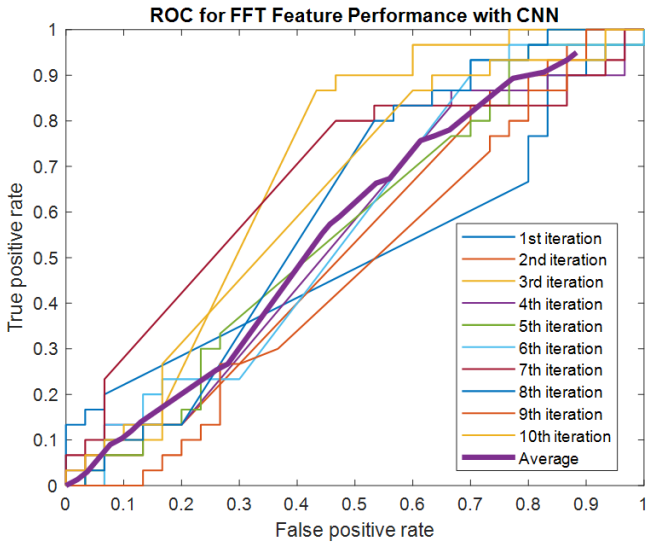

Figure 4.26. ROC for FFT feature performance of subject 15 (arousal class).

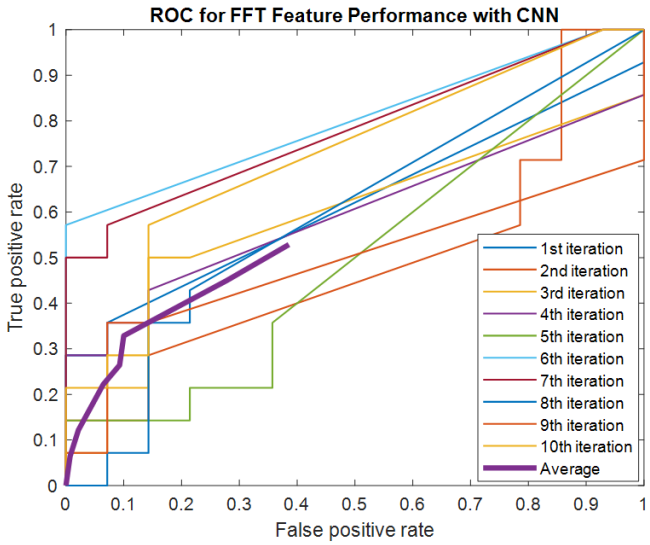

Figure 4.27. ROC for FFT feature performance of subject 20 (arousal class).

Table 4.8. Performance measurements for Subject 15 and Subject 20 (arousal class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,67	0,59	59
S20	0,59	0,75	62,9

4.3.3. Results for Wavelet Transform Features

Wavelet transform features used method is our last method in this thesis, it has the highest accuracy results for both classes and individual tasks. The results shows that wavelet transform is suitable to detect patterns of the dynamic EEG signals. It can be considered that since FFT provide frequency domain amplitudes and raw signal features provide valuable spatial information, Wavelet transform's time–frequency resolution features are suitable, since the window size in wavelet transform varies with respect to different frequencies it captures the patterns better.

Table 4.9. Wavelet transform features used valence and arousal classification accuracy results for all participants.

Wavelet	**Valence**	**Arousal**
S1	71,33	75,57
S2	72,24	73,27
S3	77,06	73,07
S4	73,08	78,84
S5	86,73	68,17
S6	74,07	67,41
S7	70,82	77,30
S8	79,15	73,14
S9	82,51	76,33
S10	81,41	71,39
S11	78,27	77,31
S12	75,34	84,99
S13	75,54	71,00
S14	81,42	75,95
S15	87,04	74,80
S16	76,67	76,42
S17	77,25	75,96
S18	86,54	80,43
S19	77,03	75,95
S20	68,20	77,50
S21	76,00	81,17
S22	76,04	71,15
S23	83,18	86,26
S24	72,24	84,55
S25	74,84	76,10
S26	77,72	81,84
S27	67,83	81,44
S28	78,55	68,62
S29	81,30	73,75
S30	70,47	73,35
S31	81,31	80,83
S32	77,66	78,36
Average	77,15	76,32

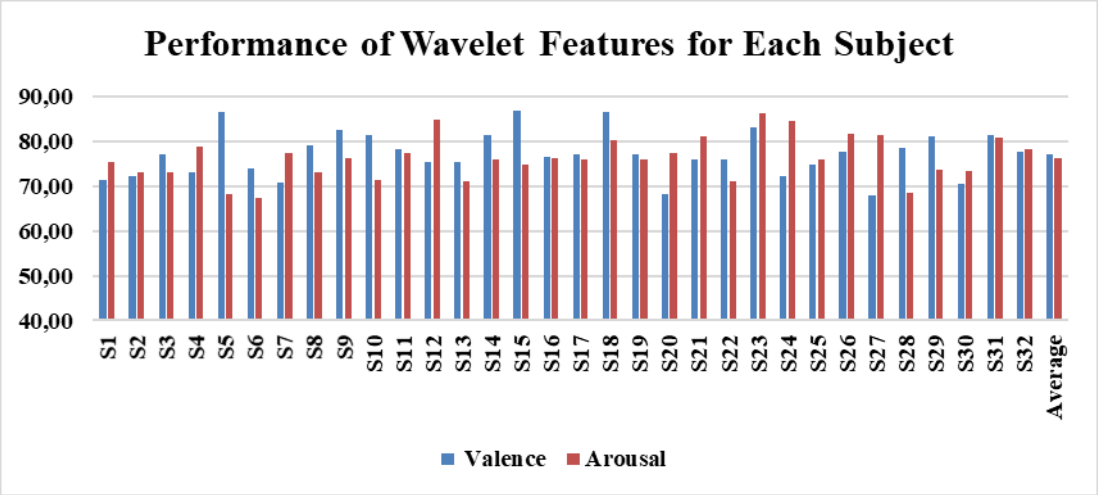

Figure 4.28. Wavelet features performance results for each subject.

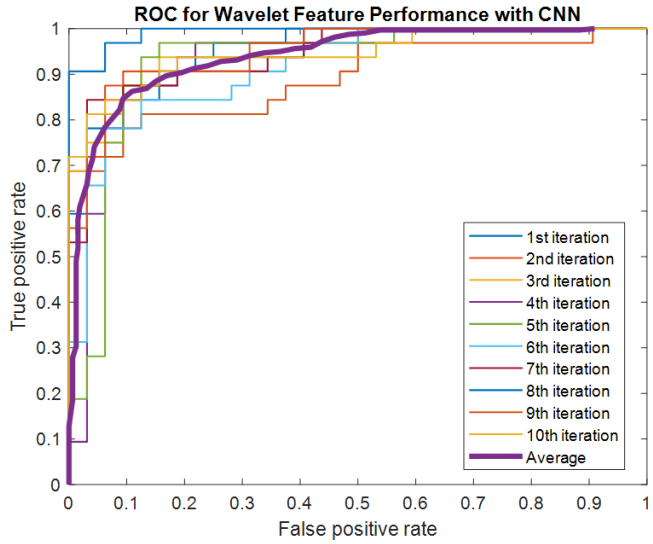

Figure 4.29. ROC for wavelet feature performance of subject 15 (valence class).

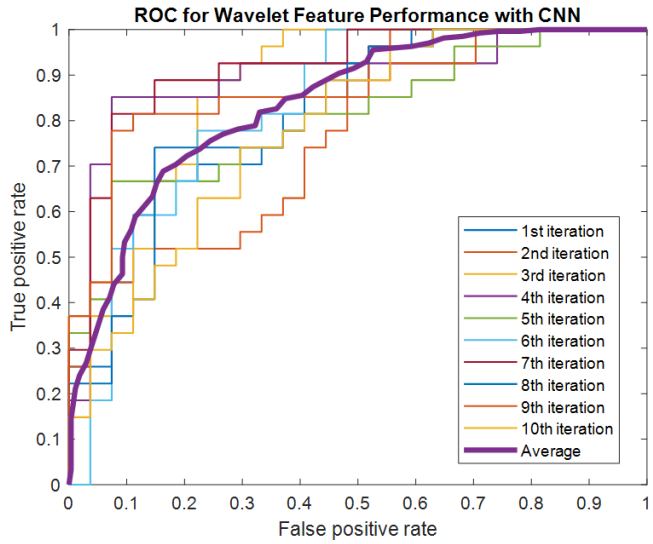

Figure 4.30. ROC for wavelet feature performance of subject 20 (valence class).

Finally, ROC performances for wavelet features were created for valence and arousal classes presented in Figure 4.29, Figure 4.30, Figure 4.31, Figure 4.32. Moreover, performance measurements specificity, sensitivity and accuracy calculated for two participants can be seen in Table 4.10 and Table 4.11. ROC curves for each iterations and overall averages are bigger than the previous cases, this situation shows the performance of wavelet features for classification with pretty well results.

Table 4.10. Performance measurements for Subject 15 and Subject 20 (valence class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,86	0,88	87,04
S20	0,63	0,86	68,2

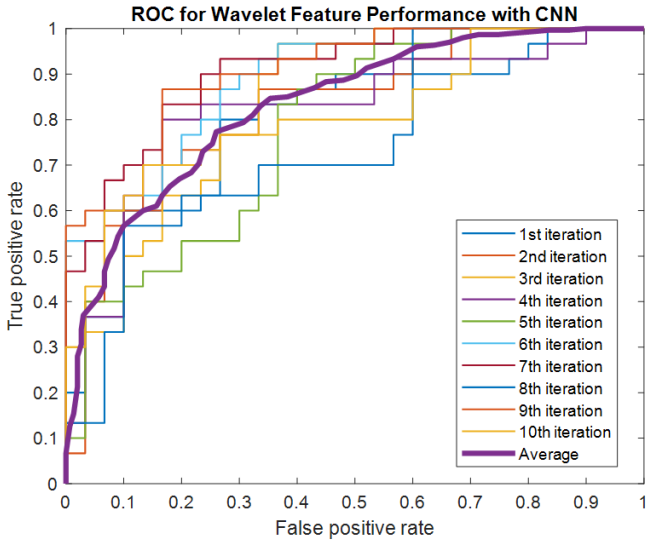

Figure 4.31. ROC for wavelet feature performance of subject 15 (arousal class).

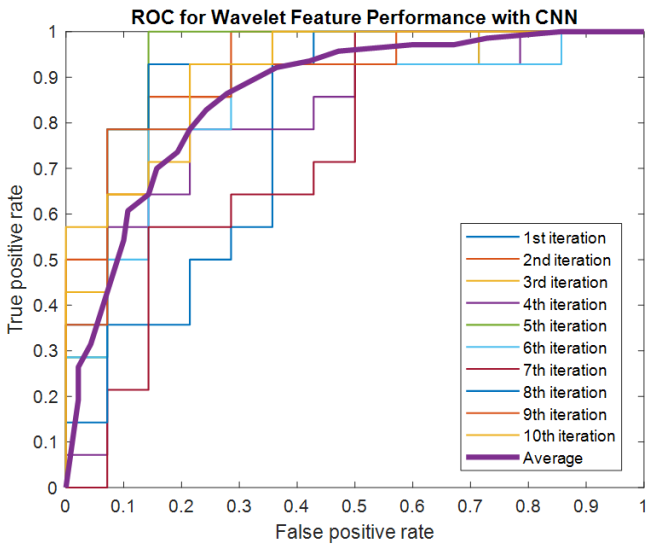

Figure 4.32. ROC for wavelet feature performance of subject 20 (arousal class).

Table 4.11. Performance measurements for Subject 15 and Subject 20 (valence class) classification.

Case/Subject Number	Specificity	Sensitivity	Accuracy
S15	0,72	0,8	74,8
S20	0,91	0,76	77,5

4.3.4. Comparison of the General Results

When we checked the performances for classes, it can be seen from the Figure 4.33 and Figure 4.34, raw signal image features and wavelet transform feature results dominate the performance comparison scales, on the other hand FFT features were not performed effectively.

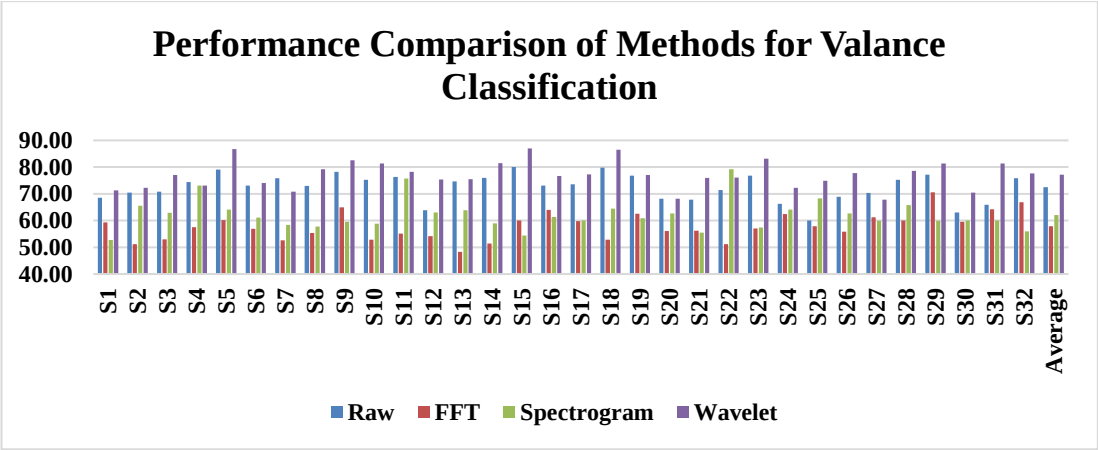

Figure 4.33. Comparison of all methods for valence class individually.

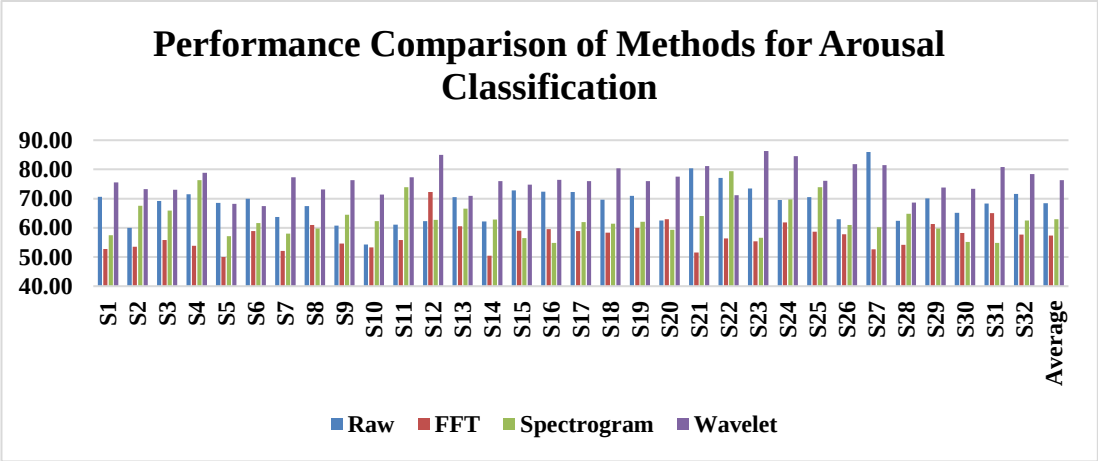

Figure 4.34. Comparison of all methods for arousal class individually.

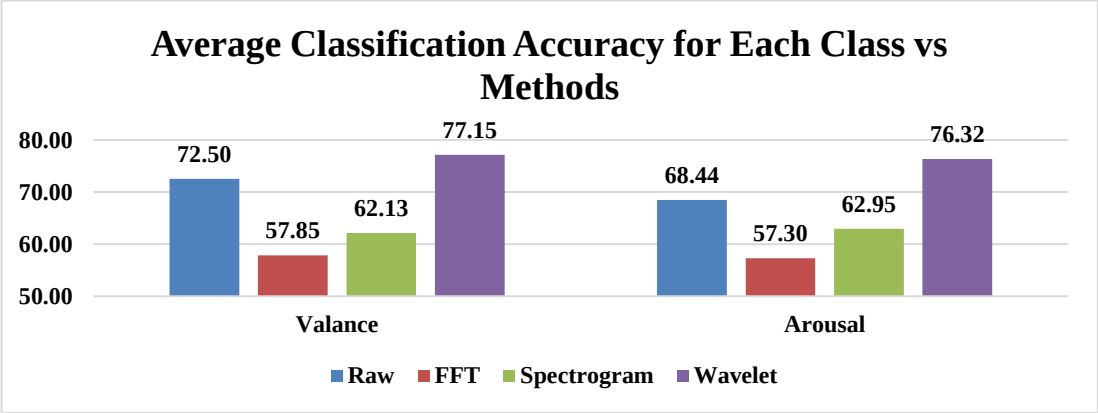

Figure 4.35. Average performance comparison of all methods for valence and arousal classes.

Figure 4.35 shows that, wavelet and raw signal image features outperforms FFT and spectrogram features. A few important conclusions can be made about the results:

- All channels contain valuable information in time domain and CNN successfully extracts this pattern for raw information.
- Since time information is lost in Fourier coefficients classification fails.
- The calculation of spectrogram matrices for each signal individually both increases data size and classifier complexity. Even though spectrogram results are better than Fourier features, it cannot reach the success rate of raw signals.
- Since the wavelet features includes both time and frequency information and all channel information were constructed together in wavelet coefficient matrix, a more efficient and accurate classification results were obtained.

Moreover, average performance results of 4 feature methods are close each other for both of the classes. This indicates the consistency of the method performances for valance and arousal.

4.4 Performance Comparison with Literature Studies

EEG signal used emotion recognition studies investigated several feature extraction and classification techniques. Some of the studies that uses DEAP database are as follows. Koelstra et al. used PSD and power asymmetry and reached 62,7% and 62% accuracies for valence and arousal classes (Koelstra, 2012). Li et al. applied to deep belief networks (DBN) for raw EEG signals and reached arousal and valence scale accuracies 58,4% and 64,2% respectively (Li et al., 2015). Yoon and Chung examined probabilistic classifier based on Bayes' theorem and a supervised learning using a perceptron convergence algorithm and used PSD and power asymmetry for feature extraction (Yoon et al., 2013). Atkinson et al. examined SVM classifier by using minimum-Redundancy-Maximum-Relevance (mRMR) feature extraction method reached 73,06/73,14% (arousal/valence) accuracies respectively (Atkinson et al., 2016). Liu and Sourina used fractal dimension and statistical features and Support Vector Machine (SVM) is used as a classifier 76,51/50,8% (Liu et al., 2014). In the recent years, Chao et al. applied CapsNet by using multiband feature matrix and succeded 66,73/68,28% accuracies (Chao et al., 2019). On the other hand Xing

et al. used Stack autoencoder and LSTM and provided 81,1/74,8% accuracies (Xing et al., 2018). Additionally, Wang et al., applied another deep learning method 3D CNN and reached 72,1/73,1% accuracies (Wang et al., 2019). Lastly, this year in 2021 Joshi et al. used Linear Formulation of Differential Entropy and BiLSTM structure and succeeded 76 and 75,5 percentage accuracies. Our proposed study on the other hand, reached 77,18/76,44% with wavelet features. The results are summarized in Table 4.12.

Table 4.12. Performance comparison of previous studies on DEAP dataset and proposed method.

Previous Studies	Number of Channels	Methods	Valence Accuracy	Arousal Accuracy
Koelstra et al.	32	PSD and power asymmetry	62,7	62
Li et al.	32	Raw+DBN	58,4	64,2
Yoon and Chung	(32+61*) 93	FFT+Bayes Theorem	70,9	70,1
Atkinson et al.	14	Mrmr+SVM	73,14	73,06
Liu and Sourina	32	FFT+SVM	50,8	76,51
Chao et al.	32	Multiband Feature Matrix+CapsNet	66,73	68,28
Xing et al.	32	Stack AutoEncoder+ LSTM	**81,10**	74,38
Wang et al.	32	3D CNN	72,1	73,1
Joshi et al.	32	Linear Formulation of Differential Entropy+BiLSTM	76	75,5
Proposed study multi electrode spectrogram	40	Spectrogram+CNN	62,13	62,95
Proposed Study	40	Raw+CNN FFT+CNN Wavelet+CNN	72,49 58,01 **77,18**	68,35 57,37 76,44

** 61generated virtual channels*

As can be observed from table, the best results were observed with wavelet features in valance classification whereas Xing have a slightly better performance at valence classification. However, for arousal classification performance our proposed study with wavelet features has better accuracy. Additionally, although Liu and Sourina's arousal accuracy is slightly better than our proposed method, their valence scale result is in mid-level and two class accuracies far away from each other. Since results

of proposed study are close to each other we can say that proposed study has a good performance both in terms of accuracy and consistency.

CHAPTER 5
CONCLUSIONS AND FUTURE STUDIES

In this thesis, it was aimed to provide emotion classification by using EEG signals for future studies of HMI field since it has a significant place in today's and future's technology. To examine this task, emotion classification studies and emotion models in literature firstly reviewed, and it was seen that there were many feature extraction, classification and data recording methods. Also, emotion definitions mainly done according to two theories. To examine emotion classification properly for HMI field, we aimed to provide an accurate, realiable and efficient study. Thus, one of the valuable performance achiever models of literature which is CNN is selected to perform our study.

Therefore, one of emotion definitions followed in our first study of our emotion classification study, as it is stated in Chapter 4.1., data recording was done by using cost efficient and simply usable single channel EEG device and spectrogram image features extracted by using STFT. To understand the pattern performance of these features, CNN model was chosen and were classified in a pretrained network Googlenet. It was aimed to provide implementation with a cost efficient and simple to use EEG device since reachability is also important for future works. Spectrogram images selected in order to get time domain frequency changes since EEG signals have unstable characteristics and their performance was successful for three emotion labels and reached 84,69% general accuracy, but it was seen that there was a difference between labels' sensitivity, specificity results. Since fear is an intensive emotion, results of fear were better noticeable.

Therefore, it was concluded that the emotion model that was followed in this study is not proper for general usage of study and more data needed. As it was indicated in Chapter 2, responses can be affected from individual factors. Thus, followed emotion theory changed and instead of it, circumplex model selected and DEAP database which is created according to this emotion theorem is selected. Actually, at first we considered to create our own dataset according to this emotion model but due to the

Covid-19 pandemic, we followed a publicly available dataset from the literature. Another reason for the choice of this dataset was, it was one of the commonly used dataset in the literature for emotion classification, so it provides a comparison chance also. Moreover, it was created with 32 channel EEG device and includes 40 biosignals in total. Although single channel EEG device maintains a cost efficiency and simple usage chance, it is better the acquire more data for this kind of study. Classification model selected as CNN once again but for this time our own model created. The spectrogram images crated with valance and arousal states for all channels were fed to CNN for classification. Although successful result were obtained with spectrogram and CNN approach, the training times were too long. Then, more efficient models were searched.

Two different feature extraction methods namely the FFT and wavelet transforms were employed. Thus, physiological signal patterns were examine in different resolutions. Accordingly, new CNN models were created. It was aimed to see its individual performance so raw signals image features also classified with same CNN structure. It was observed that, wavelet transform's resolution has a significant success to capture nonstationary EEG dynamics for valence and arousal classes. The classification results were close to each other thus it can be concluded that this method is suitable to extract meaningful data from EEG signals for emotion classification. Moreover, feature tasks examined with ROC curves for two subjects and as it was expected, AUC of ROC curves for FFT feature used task had smallest area contrarily wavelet had the biggest and curve line was too close to 1. This was also another proof for the success of proposed classifier. Additionally, proposed CNN structure and parameter values also can be stated as suitable for EEG emotion classification. One of the contributions of our thesis is, to provide comparison between time, frequency and time-frequency features on same CNN classification algorithm. It was proved that besides the frequency features time domain information also has major importance for emotion classification tasks. Another contribution is to use all wavelet coefficients in one matrix formation. Thus, different resolutions were obtained in one input. Additionally, since all these executions applied to 40 signal channels of DEAP dataset, this might be a standard usage for a multimodal study for this research area.

As further studies, a multi-channel EEG emotion capture experiment can be designed, and raw and wavelet features might be used for emotion classification. Except for the binary classification of valance and arousal, different settings could be made. Also, those studies could be combined with cognitive and neuromarketing researches.

REFERENCES

Abbaschian, B.J., Sierra-Sosa, D., & Elmaghraby, A.S. (2021). Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models. *Sensors (Basel, Switzerland), 21*.

Albornoz, E., Milone, D.H., & Rufiner, H. (2011). Spoken emotion recognition using hierarchical classifiers. *Comput. Speech Lang., 25*, 556-570.

Atkinson, J., & Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl., 47*, 35-41.

Ayadi, M.M., Kamel, M., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit., 44*, 572-587.

Basharirad, B., & Moradhaseli, M. (2017). Speech emotion recognition methods: A literature review. *AIP Conf. Proc*, *1891*.

Bengio, Y. (2016). *Deep Learning*. MIT Press.

Chao, H., Dong, L., Liu, Y., & Lu, B. (2019). Emotion Recognition from Multiband EEG Signals Using CapsNet. *Sensors (Basel, Switzerland), 19*.

Chen, L., Mao X., Xue, Y. and Cheng, L. L. (2012). Speech emotion recognition: Features and classification models. *Digit. Signal Process*., *22 (6),* 1154–1160.

Donmez, H., & Ozkurt, N. (2019). Emotion classification from EEG signals in convolutional neural networks. *2019 Innovations in Intelligent Systems and Applications Conference (ASYU),* 1-6.

Dzedzickis, A., Kaklauskas, A., & Bučinskas, V. (2020). Human Emotion Recognition: Review of Sensors and Methods. *Sensors (Basel, Switzerland), 20*.

Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*(6), 1125–1134.

Eskimez, S. E., Duan Z. and Heinzelman, W. (2018). Unsupervised Learning Approach to Feature Analysis for Automatic Speech Emotion Recognition. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5099-5103.

Feldman Barrett, L. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cogn. Emot.* 12, 579–599.

Gonzalez, R. and Woods, R., (2007). Digital Image Processing. New Jersey: Prentice Hall.

Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004). Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. *ADS*.

Happy, S., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing, 6*, 1-12.

Heckbert, P.S. (1998). Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm.

Jain, D., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognit. Lett., 120*, 69-74.

Joshi, V. & Ghongade, R., (2021). EEG based emotion detection using fourth order spectral moment and deep learning. *Biomedical Signal Processing and Control.*, *68*.

Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors (Basel, Switzerland), 18*.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Transactions on Affective Computing, 3*, 18-31.

Li, F., Karpathy A., and Johnson, J. (2016). *CS231n: Convolutional Neural Networks for Visual Recognition*, Stanford University, [online] Available: https://cs231n.github.io/.

Li, X., Zhang, P., Song, D., Yu, G., Hou, Y., & Hu, B. (2015). EEG Based Emotion Identification Using Unsupervised Deep Feature Learning. *Neurocomputing, 165,* 23–31.

Liu, Y., & Sourina, O. (2014). EEG-based subject-dependent emotion recognition algorithm using fractal dimension. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3166-3171.

MathWorks, İzmir, TURKEY. (2021). *Deep Learning Toolbox: User's Guide (r2020a).*

Retrieved August 20, 2021 from https://www.mathworks.com/help/deeplearning/.

Neurosky, (2014),

Retrieved August 20, 2021 from http://developer.neurosky.com/docs/doku.php?id=developer_tools_2.5_development_guide/.

Picard, R. W. (1997). *Affective computing.* The MIT Press.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, *17*(3), 715–734.

Revina, I., & Emmanuel, W. (2021). A Survey on Human Face Expression Recognition Techniques. *J. King Saud Univ. Comput. Inf. Sci., 33*, 619-628.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A Review of Emotion Recognition Using Physiological Signals. *Sensors (Basel, Switzerland), 18*.

Sujono, & Gunawan, A.A. (2015). Face Expression Detection on Kinect Using Active Appearance Model and Fuzzy Logic. *Procedia Computer Science, 59*, 268-274.

Taylor, P., Siddiqi, M.H., Ali, R., Sattar, A., Khan, A.M., Siddiqi, M.H., Ali, R., Sattar, A.,Khan, A.M., Lee, S. (2014). Depth camera-based facial expression recognition system using multilayer scheme. *IETE Tech. Rev. 31*, 277–286.

Vinola, C., & Vimaladevi, K. (2015). A Survey on Human Emotion Recognition Approaches, Databases and Applications. *Electronic Letters on Computer Vision and Image Analysis, 14*, 24-44.

Wang, Y., Huang, Z., McCane, B., & Neo, P. (2018). EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1-7.

Xing, X., Li, Z., Xu, T., Shu, L., Hu, B., & Xu, X. (2019). SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG. *Frontiers in neurorobotics*, *13*, 37.

Yoon, H.J., & Chung, S.Y. (2013). EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Computers in biology and medicine, 43 12*, 2230-7.

You, M., Chen C., Bu, J., Liu, J. and Tao, J. (2006). A Hierarchical Framework for Speech Emotion Recognition. *2006 IEEE International Symposium on Industrial Electronics*, 515-519.

Zhang, L., & Tjondronegoro, D. (2011). Facial Expression Recognition Using Facial Movement Features. *IEEE Transactions on Affective Computing, 2*, 219-229.

Zhao, X., Shi, X., & Zhang, S. (2015). Facial Expression Recognition via Deep Learning. *IETE Technical Review, 32(5), 347–355.*