YAŞAR UNIVERSITY

GRADUATE SCHOOL

MASTER THESIS

CARDIAC ARRHYTHMIAS CLASSIFICATION BASED ON SPECTROGRAM AND CONVOLUTIONAL NEURAL NETWORKS

SENA YAĞMUR ŞEN

THESIS ADVISOR: ASSIST. PROF. DR. NALAN ÖZKURT

ELECTRICAL AND ELECTRONICS ENGINEERING

PRESENTATION DATE: 04.02.2021

BORNOVA / İZMİR
FEBRUARY 2021

ABSTRACT

CARDIAC ARRHYTHMIAS CLASSIFICATION BASED ON SPECTROGRAM AND CONVOLUTIONAL NEURAL NETWORKS

Şen, Sena Yağmur
MSc, Electrical Electronics Engineering
Advisor: Assist. Prof. Dr. Nalan Özkurt
February 2021

In this thesis, the main objective is to classify heart rhythms which were acquired from MIT-BIH Arrhythmia Database using Convolutional Neural Networks (CNN) architecture. The classified heartbeats are normal sinus rhythm, premature ventricular contraction (PVC), left bundle branch block (LBBB), and right bundle branch block (RBBB). Two main studies are described as follows; *CNN Classification of Time-Series vs. Spectrogram Study* and *Spectrogram-CNN-Hyperparameter Tuning with Adam Optimization Algorithm Study.*

In *CNN Classification of Time-Series vs. Spectrogram* study, the classification of arrhythmias was carried out via using both raw signals and Short-Time Fourier Transform (STFT) in order to analyze the characteristic of the heartbeat signals. In time-domain classification, normal sinus rhythm, PVC and RBBB heart signals were used as a 1-D vector. The large number of electrocardiogram (ECG) time-series signals were classified with CNN by using their raw form. With the aid of STFT, Hamming window was applied in order to obtain the information from the signal. The normal sinus rhythm, PVC, and RBBB heart signals which are in the time domain were transformed into the time-frequency domain with the help of STFT. The STFT provided the acquisition of spectrograms from heart signals, and these spectrograms were classified with CNN through using as their RGB image form. The proposed CNN Classification of Time-Series vs. Spectrogram study demonstrated the high accomplishment rates in the deep learning approach, and according to accuracy, sensitivity and specificity terms, and also showed to better than traditional feature extraction methods.

In *Spectrogram-CNN-Hyperparameter Tuning with Adam Optimization Algorithm* study, heartbeat signals were examined by tuning CNN hyperparameters. The ECG heart signals which are normal sinus rhythm, LBBB, and RBBB were transformed into their corresponding spectrograms in order to acquire characteristics of the heart signals. These spectrograms were restricted with a particular time/frequency resolution rate, and the suitable time/frequency resolution rate was identified heuristically. Adam was selected as an optimization algorithm of the deep learning network in order to train the ECG spectrograms. The tuned hyperparameters were the learning rate, gradient decay factor and squared gradient decay factor of the Adam algorithm. The identification of hyperparameters was performed by using the grid search method in order to compare the results. The effect of the tuning process according to learning rate and the moment estimation coefficients were represented as their validation loss graphs. The proposed Spectrogram-CNN-Hyperparameter Tuning with Adam Optimization Algorithm study yielded great achievement rates with respect to accuracy, sensitivity and specificity terms.

Key Words: ECG heartbeat classification, arrhythmia detection, electrocardiography, short-time fourier transform, deep learning, convolutional neural network, hyperparameter tuning, grid search, adam optimization, adaptive learning

ÖZ

KALP ARİTMİLERİNİN EVRİŞİMSEL SİNİR AĞLARI VE SPEKTROGRAM TABANLI YÖNTEMLE SINIFLANDIRILMASI

Şen, Sena Yağmur
Yüksek Lisans Tezi, Elektrik Elekronik Mühendisliği
Danışman: Yard. Doç. Dr. Nalan Özkurt
Şubat 2021

Bu tezde, Evrişimsel Sinir Ağları (ESA) mimarisi kullanılarak MIT-BIH Aritmi Veritabanından elde edilen kalp ritimlerinin sınıflandırılması temel amaçtır. Sınıflandırılmış kalp vuruları normal sinüs ritmi, erken ventriküler kasılma, sol dal bloğu ve sağ dal bloğudur. İki ana çalışma şu şekilde tanımlanmıştır; *Zaman-Serisi-ESA ile Spektrogram-ESA Karşılaştırılması* ve *Adam Optimizasyon Algoritması ile Spektrogram-ESA Hiperparametre Ayarlanması.*

Zaman-Serisi-ESA ile Spektrogram-ESA Karşılaştırılması çalışmasında, kalp sinyallerinin özelliklerini analiz etmek için, aritmilerin sınıflandırılması hem ham sinyaller hem de Kısa Süreli Fourier Dönüşümü (KSFD) kullanılarak gerçekleştirilmiştir. Zaman alanı sınıflandırmasında normal sinüs ritmi, erken ventriküler kasılma ve sap dal bloğu kalp sinyalleri bir boyutlu vektör olarak kullanılmıştır. Çok sayıda EKG zaman serisi sinyallerinin ham halleri kullanılarak ESA ile sınıflandırıldı. KSFD yardımıyla sinyalden bilgi almak için Hamming pencereleme yöntemi uygulandı. Zaman alanında bulunan normal sinüs ritmi, erken ventriküler kasılma ve sağ dal bloğu kalp sinyalleri KSFD yardımıyla zaman-frekans alanına dönüştürüldü. KSFD, kalp sinyallerinin spektrogramlarının alınmasını sağladı ve bu spektrogramlar, RGB görüntü formları kullanılarak ESA ile sınıflandırıldı. Önerilen Zaman-Serisi-ESA ile Spektrogram-ESA Karşılaştırılması çalışması, derin öğrenme yaklaşımında yüksek başarı oranlarını doğruluk, duyarlılık ve özgüllük terimlerine göre ve ayrıca geleneksel özellik çıkarma yöntemlerinden daha iyi olduğunu göstermiştir.

Adam Optimizasyon Algoritması ile Spektrogram-ESA Hiperparametre Ayarlanması çalışmasında, kalp sinyalleri ESA hiperparametreleri ayarlanarak incelenmiştir. Normal sinüs ritmi, sol dal bloğu ve sağ dal bloğu EKG kalp sinyalleri, bu sinyallerin karakteristiğini elde etmek için kendilerine karşılık gelen spektrogramlarına dönüştürüldü. Bu spektrogramlar belirli zaman/frekans çözünürlük oranına göre sınıflandırıldı ve uygun zaman/frekans çözünürlük oranı sezgisel oalrak tanımlandı. Adam, EKG spektrogramlarını eğitmek için derin öğrenme ağının optimizasyon algoritması olarak seçildi. Ayarlanmış hiperparametreler, Adam algoritmasının öğrenme hızı, gradyan azalma faktörü ve kare gradyan azalma faktörüdür. Sonuçların karşılaştırılması için hiperparametrelerin tanımlanması grid arama yöntemi kullanılarak yapılmıştır. Ayarlama işleminin, öğrenme hızına ve moment tahmin katsayılarına göre etkisi, doğrulama kayıp grafikleri olarak temsil edilmiştir. Önerilen Adam Optimizasyon Algoritması ile Spektrogram-ESA Hiperparametre Ayarlanması çalışması, doğruluk, duyarlılık ve özgüllük kavramlarına göre büyük başarı oranları göstermiştir.

Anahtar Kelimeler: EKG kalp vuru sınıflandırması, aritmi tespiti, elektrokardiyografi, kısa süreli fourier dönüşümü, derin öğrenme, evrişimsel sinir ağı, hiperparametre ayarlaması, grid arama, adam optimizasyon, uyarlamalı öğrenme

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Assist. Prof. Dr. Nalan ÖZKURT for her guidance and patience during this thesis. She contributed to me different and important aspects in academic ways and believed me for producing successful studies. She always stood behind me to encourage me when I was feeling down. When I was trying to demonstrate and implement my ideas, she was there to answer my questions whenever. I felt like she was no different than my mother, and this thesis would not be possible without her. I am very thankful for her never-ending contributions and support.

Secondly, I would like to express my continual love to my family who provides unconditional trust and support, especially my brother who wants to become a software engineer, caring about me and protecting me in every situation in my life.

Thirdly, I would like to thank my friend Research Assist. Çağla SARVAN for her support and for sharing her information during this progress. I also have to thank her for answering my questions with patience.

Fourthly, I would like to thank my best friend Ali Mert CEYLAN for his support, trust, discussions, and for sharing his ideas during this thesis. He was my best support and enlightened me when I got confused.

I would like to thank my dearest friend Aslı Nilüfer ASAN for helping with some of the figures in my thesis.

I would like to thank my friend Research Assist. Ceyhan TÜRKMEN for his supportive attitude and for encouraging me during this thesis.

Thank you all who are named above, this thesis has become more meaningful for me with all of you.

Sena Yağmur ŞEN

İzmir, 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYMBOLS AND ABBREVIATIONS

ABBREVIATIONS:

ADAM	Adaptive Moment Estimation
AV	Atrioventricular
CFT	Continuous Fourier Transform
CNN	Convolutional Neural Network
DFT	Discrete Fourier Transform
ECG	Electrocardiogram
FFT	Fast Fourier Transform
FS	Fourier Series
FT	Fourier Transform
GD	Gradient Descent
HPF	High Pass Filter
LA	Left Atrium
LBBB	Left Bundle Branch Block
LPF	Low Pass Filter
LV	Left Ventricle
MLP	Multi-Layer Perceptron
N	Normal Beat
NN	Neural Network
PVC	Premature Ventricular Contraction
RA	Right Atrium
RBBB	Right Bundle Branch Block
RV	Right Ventricle

SA	Sino Atrial
SGD	Stochastic Gradient Descent
SGDM	Stochastic Gradient Descent with Momentum
STFT	Short Time Fourier Transform

CHAPTER 1
INTRODUCTION

1.1. The Motivation and Literature Review

There are various systems in the human body are carried out pivotal operations for maintaining human body health. The circulatory system is responsible for blood streaming continuously into the human body. The heart is the main component with regard to the circulatory system that is located on the chest. The health condition of the heart can be analyzed by scrutinizing the characteristics of the heart signals. Provided that information on the heart signals, the correct diagnosis and treatment can be applied to patients according to their heart disorders. In recent years, many deaths are caused by cardiac diseases across the world. Cardiac diseases are inhibited from sustaining healthy life in each person. The problem of developing automated construction to the analysis of electrocardiogram (ECG) signals according to their changing morphologies for different patients. The difficulties occurred to the identification of ECG pattern recognition because the ECG signals can have background noises, artifacts that are caused by measurements from the electrodes of the ECG medical device. In the literature, different methods are developed for the identification of pattern recognition, but artificial neural networks have a great significance in ECG signal classification. Artificial neural network-based classification techniques are well-known means of solving non-linear problems and their robustness property. The importance of artificial neural networks is also coming from noise toleration, fast and efficient computing ability. The adaptation of neural networks when faced with new pattern recognition in order to solve complex problems. From this aspect, a large number of data can be used in the neural networks for non-linear problems. The developed systems can assistance to medical experts to diagnose crucial heart rhythm disorders.

The electrical signals generated by the heart have to be scrutinized to figure out the cardiac health of the human body. All of the electrical signals measured from the human body are categorized under biomedical signals. Biomedical signals are utilized to understand the underlying physiological mechanisms of a particular biological system or event. One of the biomedical signal types is the electrocardiogram (ECG) signal which is known as a cardiac signal. The cardiac signals can be obtained from ECG medical device that provides us to identify patient heart disorders. In one cardiac cycle, the observable heart signals are characterized by the heart's condition via using ECG medical device according to cardiac signal characteristics. The medical expert can identify a heart condition according to signal characteristics from measurements by monitoring the ECG medical device. Abnormalities of the cardiac signals are called arrhythmia. These arrhythmias are categorized as their signal characteristics such as P-QRS-T segments or duration and wave structure. The entire information is obtained from signals enables medical experts to label heartbeats as abnormal or not. There are many approaches for the detection and classification of ECG signals based on feature extraction methods, machine learning methods, and deep neural network-based methods. In a deep neural network approach, suitable optimization algorithms should select according to the engineering problem to obtain high accomplishment rates such as accuracy, specificity, sensitivity, positive predicted value, etc. In these circumstances, different methods are presented in the literature studies.

In the investigation of characteristics of the cardiac signals, signal processing techniques are utilized widespread. According to the investigation of some researchers, ECG signals with high-frequency components have been evaluated as signal-beat analysis or signal averaging on 128 sequential beats based on Single-beat Spectral Variance (SBSV) by using 2-dimensional Fourier Transform (FT), and they have been divided patients into three groups as myocardial infarction that patients' have ventricular tachycardia or not. Following their results, SBSV has been categorized 29 out of 35 patients as pathologic with an accuracy of 83% in group 1. In group 2, 5 out of 50 patients have been shown abnormality with an accuracy of 48%. In group 3, there has been detected no pathologic outcome (Spiegl et al., 1998).

Feature extraction from biomedical signals is a prevailingly used technique. This feature extraction method provides us to categorize signals according to their signal characteristics. Because ECG signals have great importance of detection properly, some feature extraction methods are used in many investigations. One way of the feature extraction from ECG signals is the morphological approach, which is analyzed according to time intervals or segment durations for pivotal heartbeat structures by using mathematical consideration. Ghongade and Ghatol (2008) indicated that four types of ECG signals classification are normal sinus rhythm, atrial premature beat, premature ventricular contraction beat, and left bundle branch block beat using morphological features of QRS complexes have been classified with artificial neural network approach: Support Vector Machine (SVM), Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP). In this study, they have been obtained 99 % accuracy in both three classifiers with a limited set. On the other hand, Shufni and Mashor (2015) stated that in their study, time-domain ECG signals are used at the initial phase after applied Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) on these ECG signals for evaluating in frequency-domain by looking at statistical features as mean, variance and standard deviation. According to Shufni's results, they have been accomplished with an accuracy of 95.45% and 96.21% in both the time domain and frequency domain respectively. Wavelet Transform is especially used for non-stationary signals, such as ECG signals, that provide both time and frequency components at a high-resolution rate. The WT is outperformed with an accuracy of 99.24% compared to statistical features. It is stated in (Balachandran et al., 2014) that ECG signals are used in their study in order to improve the system constructed on a multi-resolution wavelet transform. In their system, ECG features are highly discriminative by using the Daubechies wavelet filter, and ECG signals have been taken from the MIT-BIH database. They have been asserted that Discrete Wavelet Transform (DWT) is a strongly recommended method for signal processing and its applications (Hsieh et al., 2005; Banerjee and Mitra, 2014).

Heart Rate Variability (HRV) is controlled by the power spectrum of various frequency segments as sequential single beat analysis in terms of heart rate. Jacobson (2007) stated that in his investigation ECG signals are extracted its low frequency and high frequency according to power spectral density in 24 hours duration by using Short-Time Fourier Transform (STFT). Due to the nature of ECG signals are known

as non-stationary, in their recommended study multi-component biomedical signals are examined in the time-frequency domain through utilizing STFT, and they have been aimed to obtain good performance rates for analysis of four different type of cardiac arrhythmias (Benmalek and Elmhamdi, 2015). In order to scrutinize the non-stationary signals, such as ECG signals, time-frequency analysis methods have been the great significance of reaching both spectral and time localized information from this kind of non-stationary signals (Abdeldayem and Bourlai, 2018; Krishna, 2017; Wu et al., 2019; Choudhary et al., 2018). Wu and his friends claimed that the ECG signals, which are consisting of multiple arrhythmias, applied time-frequency transformation such as STFT, Continuous Wavelet Transform (CWT), and pseudo-Wigner-Ville distribution (PWVD). As a result, STFT is outperformed on other time-frequency transformation approaches, and they have been reached a 96.65% accuracy rate in their study (Wu et al., 2019).

Deep learning models can be seen as a particular type of artificial neural network and it is copycatted brain neurons. The considerable property of deep neural networks is allowed to classifying performance on large-scale data compared to traditional approaches (Liu et al., 2018; Pourbabaee et al., 2018; Fan et al., 2018). It is stated in (Chen et al., 2019) the six types of cardiac arrhythmias are categorized by combining the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) in order to obtain features on a larger scale ECG data. Their research has been indicated that CNN outperformed compared to other classification methods, and also they have been reached 81% classification accuracy. On the other hand, how DNA has special characteristics individually, ECG signals have also their characteristics. Therefore, the detection and classification of ECG signals have great importance for personal identification in biometrics (Xu et al., 2018). The study that is aimed to develop a generic and robust system for automated real-time patient-specific classification of ECG signals. One of the studies is related to the wavelet transform in order to extract features that are projected onto feature space as lower-dimensional via using principal component analysis (İnce et al., 2009). The other method is related to the adaptive implementation of one-dimensional CNN in order to categorize long ECG records are taken from the MIT-BIH database, and they have been asserted that their study is immensely generic, according to this reason, that system is feasible for any ECG dataset. They have been classified as ventricular ectopic beat (VEB) and

supraventricular ectopic beat (SVEB) arrhythmias with an accuracy of 99% (Kiranyaz et al., 2016).

Huang and his friends (2019) stated that the classification of ECG signals approaches that depend on CNN as a classifier with five different arrhythmias such as normal rhythm (NOR), left bundle branch block beat (LBB), right bundle branch block beat (RBB), premature ventricular contraction beat (PVC), and atrial premature contraction beat (APC) transformed into spectrogram form by using Short Time Fourier Transform (STFT). Nowadays, the evolution of CNN architectures and their improvements have been distinguished according to novelties of the investigations and using state-of-the-art techniques. In that manner, CNN hyperparameters have been adjusted according to different activation functions, changing the number of hidden units or neurons, filter size, and step size (Mehta et al., 2020; Bibaeva, 2018; Rajkumar, 2019; Sun et al., 2019). In Mehta's investigation, the effect of different activation functions of both artificial and deep neural networks has been examined on EEG signals, and they have been used threshold activation function, linear activation function, sigmoid activation function, and ReLU-based activation functions in their study. On the other hand, Bibaeve's study has been demonstrated for a different architecture of the CNN model by adjusting pooling type, activation function, filter size, and step size via using evolutionary algorithms and genetic algorithms.

The optimization algorithms are responsible for reducing the network's loss function by adjusting weights or learning rates, and also ensure more accurate results depending on the selected optimizer. Non-convex optimization incorporates the function that has more than one optima instead of having just one global optimum. When the non-convex optimization problems are considered according to loss surface, reaching the global optima is still being challenging. The neural network tries to find the global optimum by reducing the loss on the error surface simultaneously. When considered the neural networks, there is always a non-convex optimization problem (Yu and Ramamoorthi, 2019). In their study, a novel video stabilization method that leads to widespread non-convex optimization problems in order to solve the problem, they have been recommended using CNN solely as an optimizer rather than the extraction of information from the data.

Many of the researchers have been suggested lots of frameworks in order to find the best optimizer with respect to their model and data. There are various optimizers such as Adaptive Moment Estimation (ADAM), Stochastic Gradient Descent (SDG), Stochastic Gradient Descent with Momentum (SGDM), AdaDelta, AdaGrad, and RMSProp that are used in the neural network model in order to demonstrate that how optimizer effects on the loss function and provide the most accurate result compared to other neural network architectures (Poojary and Pai, 2019; Kamsing et al., 2019; Sun et al., 2019). According to Kamsing's study, four types of optimizers have been used for object identification problems with implementing CNN in order to classify satellite images in the planesnet dataset.

Through using all information that was synthesized after the literature review, the detection and classification of ECG signals by considering crucial optimization techniques are used, and the state-of-the-art approaches are emphasized in this thesis. There are developed various systems for ECG signal analysis but remain the drawbacks in the literature. Accomplishment rates are directly related to the selected features and feature extraction techniques, and the selected dataset has great importance in changing the success rate. These days, the dimension of data and size of the dataset rise enormously, also the selection of feature extraction methods is becoming challenging. Because of these reasons, deep learning approaches are indispensable, and the most optimized solutions are needed in deep learning approaches.

1.2. Aim of the Study

The main purpose of this thesis is to develop a system that provides us to the diagnosis of cardiac patients from records are in accordance with arrhythmia. The records were adjusted as thirty minutes samples long in the proposed approach that depends on ECG records taken from normal patients and patients that have cardiac arrhythmias. These ECG records are analyzed in the beat-to-beat form in order to classify them according to their labels by using CNN. In order to elaborate further, model optimization techniques were investigated the network, and also model hyperparameters were scrutinized.

Thus, the contribution of the thesis includes:

- ECG signals were analyzed in the time domain and in the time-frequency domain, ECG time-series signals and ECG spectrogram images were used as input in the CNN model and compared to each other for the classification problem with high accomplishment rates.
- ECG spectrogram images were used as an input in the CNN model, and the model hyperparameters were investigated.
- The optimization algorithms of the model were scrutinized according to optimizers.

1.3. Outline of the Thesis

In Chapter 1, literature investigations are given in detail associated with this thesis. This thesis was given the form of the related literature studies, and also the purpose of the thesis was carried out in this chapter.

In Chapter 2, heart anatomy was represented as a short form. This chapter explains the heart structure, working principle of the heart and its connected components, electrocardiography device that enables us to measure and recording of cardiac signals, and cardiac arrhythmias with the corresponding signal structure which are utilized in this thesis.

In Chapter 3, time-frequency analysis related to signal processing techniques which are Fourier Transform (FT), Continuous Fourier Transform (CFT), and Short-Time Fourier Transform (STFT) were expressed in detail. The theory about Neural Networks (NNs) with its subtitles Multi-Layer Perceptron (MLP) and Convolutional Neural Networks (CNNs) was mentioned particularly. The optimization algorithms were also scrutinized related to this thesis in this chapter.

In Chapter 4, the dataset was explained that was used in this study for arrhythmia detection, also the ECG signal classification test results were demonstrated in both the time domain and the time-frequency domain.

In Chapter 5, the detection and classification of ECG signals in the time-frequency domain according to tuning Adam algorithm hyperparameters, and the results were demonstrated.

In Chapter 6, the contribution of the thesis was debated according to the given results, and future works were explained.

CHAPTER 2
PHYSIOLOGICAL BACKGROUND

The human body is composed of organisms from cells to systems. The cell is the smallest structural and functional unit of an organism. After that, the tissue is made up of specialized cells. More than one kind of tissue is organized to become an organ for performing unique functions in the human body. Various functions of organs are coming up together for being a system that is dependent on each other. The human body is actually a closed loop that has twelve different systems, and these are mainly circulatory, digestive, endocrine, immune, lymphatic, nervous, muscular, reproductive, skeletal, respiratory, urinary, and finally integumentary systems. All these systems are responsible for sustaining body health for a living.

In order to perform blood streaming, the circulatory system manages that operation continually in the human body. Thanks to the circulatory system, nutrients, oxygen, and carbon dioxide are being transported. In this chapter, the anatomy of the heart which structure and working principle of the heart are explained briefly. Additionally, the electrocardiography device that helps to the measurement of cardiac signals, and arrhythmia types that are used in this thesis are explained as well.

2.1. Heart Anatomy

The heart is made up of special muscles to pump the blood through the body and has 4 chambers, which are namely the atrium and ventricles. The unique name of the upper ones is an atrium, and the lower ones are ventricles. These 4 chambers and cardiac valves are interconnected to each other. The right atrium (RA) and right ventricle (RV) helps to collect blood enriched by carbon dioxide, out of the body, and pump it to the lungs. On the other hand, the left atrium (LA) and left ventricle (LV) helps to collect blood enriched by oxygen, out of the lungs and pump it to the body.

In order to initiate blood streaming, the heart muscles have to be contracted. The electrical pulses are occurred by contraction of the heart in the human body. These electrical pulses are created by an enormous pacemaker in each heart particularly, is

also known as the Sino-Atrial (SA) node (Guyton and Hall, 2010).

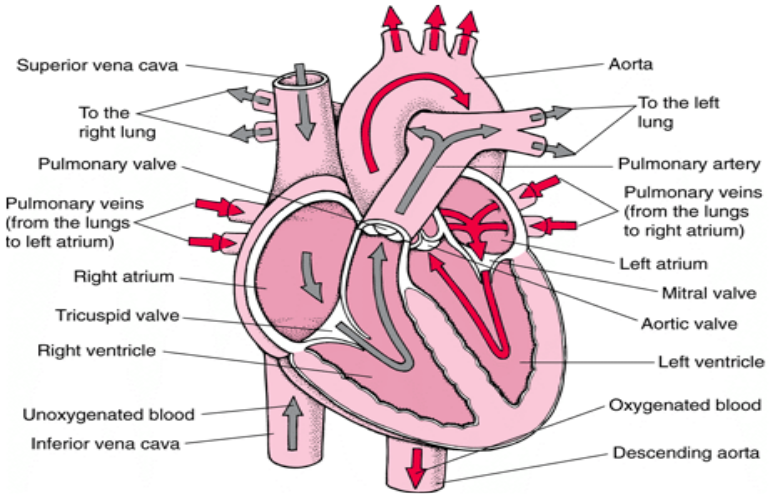

Figure 2.1. Blood Flow Direction and Structure of the Heart (Webster, 1998)

There are 3 leading elements of the heart in order to provide the cardiac cycle:

- The SA node is located in the right atrium near the gap of the superior vena cava.
- The atrioventricular (AV) node is located between the atria and the ventricles.
- The His-Purkinje fibres are located in the inner ventricular walls of the heart.

The blood flow of the heart is related to the contraction of the heart and the conformance of the cardiac valves. There are 4 different valves of the heart:

- Tricuspid valve reaches among RA and RV.
- Mitral valve reaches among LA and LV.
- Pulmonary valve reaches among RV and pulmonary vein towards the lung.
- Aortic valve reaches among LV and aorta.

The cardiac cycle includes a duration of relaxation is named diastole, throughout that the heart receives the blood, followed by a duration of contraction, in order to pump out, is named as systole.

The RA collects the blood which enriched by carbon dioxide from both vena cava as superior and inferior that are fundamental veins reaching the heart. Atrial depolarization leads to contraction of the atrial muscles. With atrial contraction, the blood is moved from the RA to the RV together with the tricuspid valve. Then in ventricular contraction, the blood, which is poor by carbon dioxide in the RV, in order

to clean the blood, the RV is sent out the blood to the lungs via the pulmonary valve. In other words, contraction is pumping the blood out and known as diastole while relaxation following contraction is known as systole.

In atrial contraction, the LA distributes the blood, which is enriched by oxygen from the lungs, towards LV via the mitral valve. The strongest diastole is actualized in LV with respect to the other chambers of the heart, and oxygenated blood spread the whole body via the aortic valve.

The cardiac contraction depends on SA node stimulations in order to produce a normal heart rhythm. The path of the conduction system shows the impulses are moved from the SA node to the AV node. AV bundle is responsible for the conduction of impulses which are transmitted from atria to ventricles. The left and right bundle branches of Purkinje fibres transmit these impulses to each part of the ventricles. Heart rhythmicity and impulse conduction are controlled via cardiac nerves such as sympathetic and parasympathetic. Parasympathetic nerve stimulation causes the release of acetylcholine hormone. Acetylcholine has two main actions on the heart which are diminished SA node's rhythm rate and low transmission of impulses to the ventricles. Sympathetic nerve stimulation causes the release of the norepinephrine hormone. Norepinephrine hormone has increased the power of heart contraction and the whole action of the heart. The conduction system of the heart can damage by any heart disease, but the normal heartbeat is 72 beats per minute (bpm) (Guyton and Hall, 2010).

The successive operations are accomplished in one heartbeat cycle:

1. The SA node is stimulated by a neurotransmitter which is known as acetylcholine.
2. The P-wave is occurred by electrical potentials produced during the depolarization of the atria at the beginning of the atrial contraction. The voltage of the P-wave is relatively between 0.1 and 0.3 mV for 60-80 ms.
3. PR/PQ interval is a period which is the duration between the starting of the P-wave and the starting of the QRS complex. PR interval is nearly 0.16 seconds. This period supports to fulfill the blood transmission from the atria into the ventricles. His-Purkinje fibres stimulate ventricles and outstretch to left and right bundle branches.

4. His-Purkinje fibres stimulate ventricles and outstretch to left and right bundle branches.

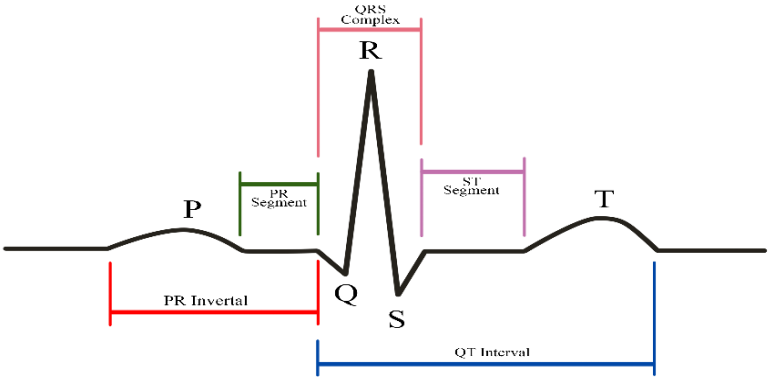

Figure 2.2. Single P-QRS-T Structure of Normal Heartbeat

5. The QRS complex is the period of the starting of excitation of the atria and the starting of excitation of the ventricles QRS complex voltage is about 1.0 to 1.5 mV amplitude and 0.83 seconds from the top of R-wave to the base of S-wave.
6. ST interval is a duration of which is the end of the QRS complex to the starting of the T-wave. This interval is explained as the duration between ventricular depolarization and repolarization, also known as the isoelectric period. The duration of the ST interval is nearly 0.08 seconds.
7. T-wave is described as a repolarization wave, it occurs ventricles are relaxed. The T-wave shape has among 0.2-0.3 mV amplitude for approximately 120-160 ms with respect to Figure 1.

2.2. Electrocardiogram

The human body is composed of different systems in itself in order to maintain the body's function completely. Heart functions are measured and monitored by ECG. Heart signals can be obtained as real-time data by using Holter monitoring. The monitoring of heart signals is done together by engineers, cardiologists, and clinicians.

P-QRS-T wave structure represents voltage versus time of the cardiac cycle of the heart by looking at the ECG monitor. As mentioned, the preceding section about one cardiac cycle (heartbeat), the first step is SA node is stimulated by the neurotransmitter, and then electrical impulse spreads along with the heart as a regular wave shape. This operation can be seen on the ECG device via using electrodes which are placed on the patient's limb and chest. The wave structure of the heart signals obtained from the ECG device is analyzed according to the P-QRS-T wave structure. In order to interpret ECG rhythm also known as Heart Rate (HR), atrial rate counts the number of P-wave and ventricular rate counts the number of QRS complexes in 60 seconds.

All types of heart signals contain both time and frequency components individually. For that reason, analyzing heart signals has crucial importance on human life. When considering the real-time data monitoring and processing together with the digital or statistical signal processing approaches, it is possible to make logical extractions from the real-time ECG signals with respect to time and frequency components and wave shapes.

The diagnosis of cardiac disease in earlier stages is the most important level of treatment. There is a statement about detection and diagnosis on time, correct diagnosis on time provides to apply the correct treatment, the correct treatment provides to decrease the rate of death. Therefore, we need a suitable and efficient system in order to not only detect but also classify the heartbeats according to their labels such as normal and abnormal by using state-of-the-art methods.

2.3. Cardiac Arrhythmias

The electrical impulses of the heart not being conducted as it is supposed to be through the heart ventricles' is called arrhythmia. Over the past decades, investigation of ECG heartbeat signals has been of great importance in order to develop the robust automated cardiac arrhythmia detection and classification system, and also the treatment of cardiac diseases (de Albuquerque et al., 2018).

In order to detect and classify ECG heartbeat signals as normal or arrhythmia, some researchers have developed successful classifiers in the literature, and the investigation about developing such classifiers has not been finished yet. The developments and findings of classifiers are still proceeding with investigation today.

Depending on the labeling of the Association for the Advancement of Medical Instrumentation (AAMI), there are 5 main heartbeat classes, and their 15 sub-clusters are indicated in Table 2.1. In Table 2.1 heartbeat classes in the MIT-BIH database are categorized with respect to AAMI standards.

Table 2.1. Heartbeat Classes according to AAMI Standards

AAMI Class	**Type of beat**
Normal (N)	Normal beat Left bundle branch block beat Right bundle branch block beat Atrial escape beat Nodal (junctional) escape beat
Supraventricular ectopic beat (S)	Atrial premature beat Aberrated atrial premature beat Nodal (junctional) premature beat Supraventricular premature beat
Ventricular ectopic beat (V)	Premature ventricular contraction Ventricular escape beat
Fusion beat (F)	Fusion of ventricular and normal beat
Unknown beat (Q)	Paced beat Fusion of paced and normal beat Unclassifiable beat

In this investigation, 4 types of heartbeats are used such as Normal (N) beat, LBBB beat, RBBB beat, and PVC beat by implementing the STFT approach.

2.3.1. Normal

N beat term defines who has a healthy heart condition in itself. Additionally, the PR interval is 0.12-0.20 seconds, it is approximately 0.16 seconds. The QRS complex interval has a 0.06-0.12 seconds duration. The normal beat is also known as a normal sinus rhythm and has its unique P-QRS-T wave pattern. Figure 2.3 represents single beat of normal ECG signal for one-second duration which is taken from the MIT-BIH Arrhythmia Database and plotted in MATLAB.

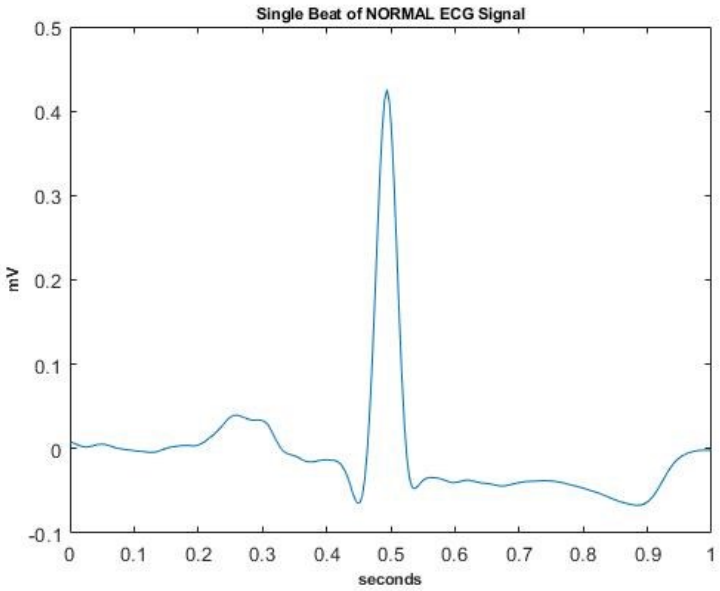

Figure 2.3. Normal Heartbeat of ECG Signal

2.3.2. Left Bundle Branch Block

LBBB is the activation of the heart's left LV being delayed and this status leading to delaying LV contraction more than the RV. LBBB is an important arrhythmia type since it indicates abnormal cardiac rhythms on the left side of the heart.

Just as mentioned in the previous part, the normal heartbeat QRS pattern duration is 0.06-0.12 seconds. On the other hand, the QRS pattern of the LBBB beat is greater than 0.12 seconds. Also, it is hard to interpret on an ECG monitor but analysis and diagnosis of this arrhythmia are also possible for doctors and clinicians by looking at the QRS pattern. Moreover, LBBB has a normal P-wave and is followed by a T-wave after the QRS structure. In order to notice LBBB arrhythmia, unforeseen downward or upright fluctuations must be seen in the QRS pattern which changes in ST-segment. Figure 2.4 represents single beat of the LBBB ECG signal for one-second duration which is taken from the MIT-BIH Arrhythmia Database and plotted in MATLAB.

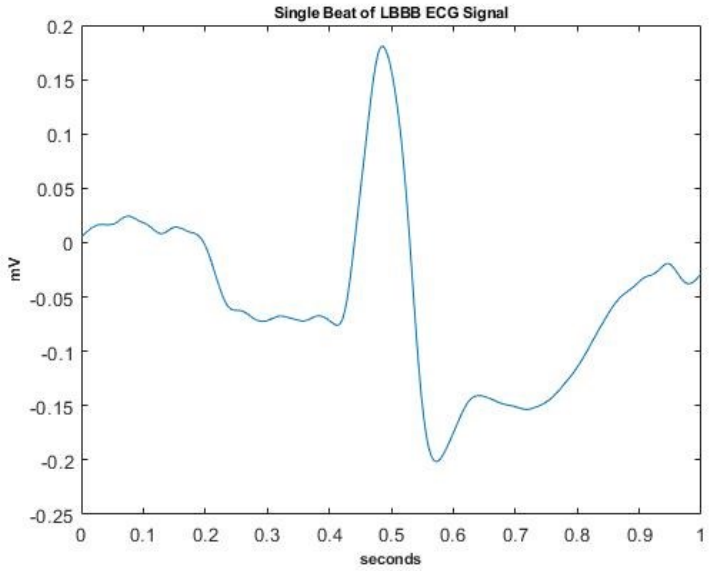

Figure 2.4. LBBB Heartbeat of ECG Signal

2.3.3. Right Bundle Branch Block

RBBB is a complete or incomplete heartbeat block of the electrical signal conduction system to the RV. RV is not stimulated directly via electrical impulses coming from the right bundle branch. The electrical signal excitation of the RV is delayed by RBBB arrhythmia. Because of this, the QRS complex is seen wider than normal sinus rhythm, and the transmission takes a longer time than normal sinus rhythm conduction because of delay in contraction.

By looking at the QRS pattern characteristic, if it is an incomplete block, then the QRS duration should be at least 0.1 seconds. On the other hand, if it is a complete block, then the QRS duration should be at least 0.12 seconds. Also, at the end of the QRS structure, S-wave is seen as a prolonged shape. RBBB is not as critical as LBBB arrhythmia since LBBB has considerable indications about cardiac disease, and also RBBB can be diagnosed in medical ways more easily as opposed to LBBB by doctors or clinicians Figure 2.5 represents single beat of normal ECG signal for one-second duration which is taken from MIT-BIH Arrhythmia Database and plotted in MATLAB.

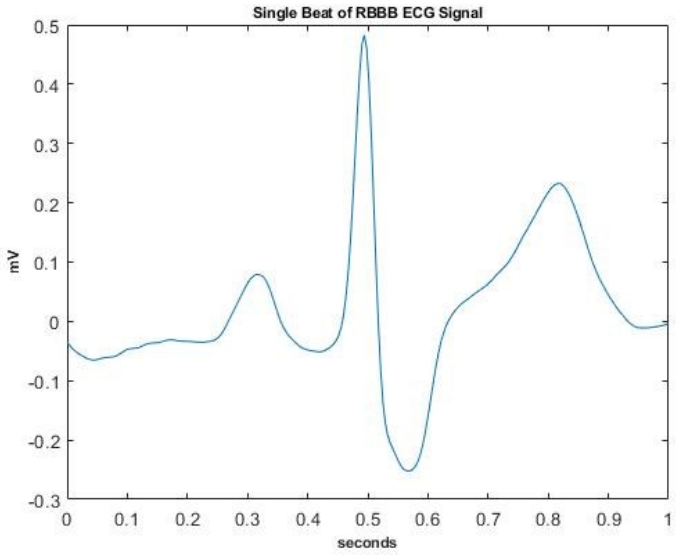

Figure 2.5. RBBB Heartbeat of ECG Signal

2.3.4. Premature Ventricular Contraction

PVC is a common arrhythmia caused by starting from the wrong stimulation point of the heart such as the Purkinje fibres in the ventricles. Stimulation should fire at the SA node via helping neurotransmitters.

Single PVC beat does not indicate a hazardous situation, but if 3 or more PVC heartbeats occur, it may be a symptom of ventricular tachycardia which is mostly fatal arrhythmia. What is more, electrocardiographic features of PVC beats are extended QRS structure duration greater than 0.12 seconds with the abnormal morphological structure according to the normal sinus rhythm, incompatible ST segment, and T-wave changes Figure 2.6 represents single beat of PVC ECG signal for one-second duration which is taken from MIT-BIH Arrhythmia Database and plotted in MATLAB.

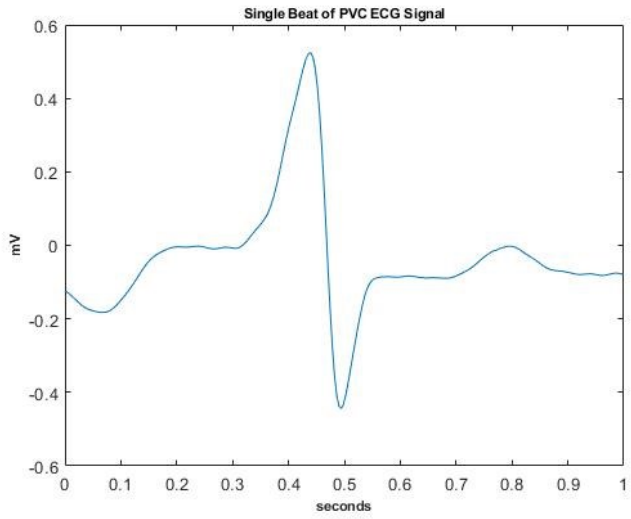

Figure 2.6. PVC Heartbeat of ECG Signal

CHAPTER 3
SIGNAL PROCESSING AND MACHINE LEARNING

In this chapter, the time-frequency analysis, the definition, notion and subtopics of Fourier Transform (FT), its important properties will be given in detail. Additionally, the machine learning techniques and their significant points for classification problems will also be given in this chapter.

3.1. Time-Frequency Analysis

The time-varying spectra are the representation of the time-frequency analysis. To exemplify, consider that any sound signal is a function of time. The energy density spectrum is how the intensity changing with time, and also it is the absolute square of FT. Additionally, the main concept of joint time-frequency representation can be shown as a time versus frequency plot. On the other hand, there is a difference between spectrum and time-frequency representation. The spectrum tells us which frequencies occurred, but joint time-frequency analysis explains which frequencies occur at a specific time. Clearly, by looking at the spectrum method, it can be said which frequencies available in the signal but cannot realize when these frequencies happen at a specific time.

Why spectrum or plural version of spectra change with time? There are two reasons in order to explain this question. The first reason is that the generation of specific frequencies based on parameters might be a change in time. For illustration, there is a mechanical oscillation system at the mechanical vibration rate in the air. When the length or tension varies with time, several frequencies will be generated in time because the string is going to vibrate with various frequencies and hit the air. The other reason is that the propagation of waves in a medium is shown as its dependency on frequency phenomena such as glass is a quite fine filter of the X-rays (Cohen, 1995).

3.2. Fourier Series and Fourier Transform

FT is one of the most major mathematical approaches to transform that a function of time for representation into its frequency domain. FT incorporates both periodic and non-periodic signals. In the real-world, square wave, sinusoidal wave, triangular wave, and saw tooth wave is known as periodic signals. Fourier Series (FS) indicates that whole signals are able to be expressed as periodic elements summation with varying amplitude and phase. For the purpose of analyzing signals or functions in a matter of frequency domain, FS expansion is used as a series of sines or cosines form which is continuous-time periodic signals. The FS is generalized into complex-valued functions or advance levels functions, in order to use FT both time and frequency domain. FT is used for the representation of non-periodic signals. In the digital signal processing world, this kind of linear transformations are mapped to a single frequency element as a localized frequency spectrum, and also linear transformations are invertible. Because signals have finite energy, signals can be considered as a certain frequency band called the spectrum. That is why forward FT and inverse FT can be used for translation between the time and frequency domain. Also, FT is appropriate for stationary signals for analyzing signal characteristics. In this thesis, the STFT method is used for converting time-domain signals into the time-frequency domain with a limited resolution property by using window operation (Auger, 1996).

3.2.1. Continuous Fourier Transform

One of FT's most significant contribution is that the representation of non-periodic signals in continuous time. By looking at the idea of Continuous-Fourier Transform (CFT), non-periodic signals can be considered as the limited form of periodic signals when a period turns into a large amount. For that reason, the non-periodic signals can construct from periodic signals in an infinite duration. FT pair consists of both analysis and synthesis parts which are also known as forwarding FT and inverse FT. Both of them show that the signals can be written as a linear combination of complex exponentials. The complex exponentials amplitude and their connected harmonic frequencies exist as a discrete set for periodic signals.

On the other hand, the complex exponentials also have amplitude, but their frequencies emerge in a continuous way with respect to the synthesis part of the FT for non-periodic signals. FS coefficients of a periodic signal ensure us to characteristic

information of $x(t)$ qua linear combination of sinusoidal waves at various frequencies, and $X(\omega)$ is generally used term that spectrum of $x(t)$. FT provides us to compute the analysis of frequency spectral for non-periodic signals. It is characterized by,

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt, \tag{1}$$

And the inverse Fourier Transform is defined as,

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{-j\omega t} d\omega, \tag{2}$$

where $x(t)$ indicates that a non-periodic signal and $X(\omega)$ represents the applied FT on the non-periodic signal $x(t)$. The spectrum term is referred to as $X(\omega)$ and it is a complex-valued function, can be defined as,

$$X(\omega) = |X(\omega)| \angle \phi(w), \tag{3}$$

where the continuous spectrum amplitude is expressed as $|X(\omega)|$ and $\angle \phi(w)$ indicates that the phase relation.

3.2.2. Short-Time Fourier Transform

STFT is the most common technique for analyzing non-stationary signals. Since FT is not sufficient for analyzing non-stationary signals, FT is proper for stationary signals such as white Gaussian noise. Also, FT will not give a suitable spectrum that exhibits what frequencies occurred at a particular time because FT cannot cope with the non-stationary signals. Therefore, STFT should apply to non-stationary signals in order to analyze them individually. The main concept of STFT is the determination of dividing the signal into several segments for Fourier analysis in order to see how spectrum changes in these segments. The underlying consideration of STFT separates signals into small time intervals and applies FT on each segment in order to figure out the frequencies occurred in these intervals. STFT also provides to analyze the localized area for non-stationary signals. As a consequence, the spectra show how the spectrum

is changing in time. When the signal is divided into short-duration signals, the short duration signals will have a large bandwidth. For this reason, the spectra of short-duration signals are very little related to the features of the original signal. There is a big distinction between STFT and FT, which is known as a window function in order to analyze signal as localized in time and apply FT on small intervals of the original signal. On the contrary, FT provides us information about whole frequencies along the whole time period in the digital signal. The window size should select according to Heisenberg's uncertainty principle. Heisenberg's uncertainty principle is stated that there is a constraint between resolution in time and frequency domains. According to this principle, low frequencies can be represented more detailed in time duration, and high frequencies can be demonstrated less detailed in time duration. This principle comes from that aspect; when window operation is localized as in the time domain, it is concluded with uncertainty in the frequency domain. If the window size is selected as wide enough, time uncertainty will raise according to Heisenberg's uncertainty principle, and vice versa. Consequently, STFT is indicated that both time and frequency resolution in order to obtain good enough information from the digital signal, but it is compulsory to select that, high frequency with low time resolution or low-frequency resolution with high time resolution.

In other words, STFT helps to overcome the limitation of FT-like analysis by using window function of a certain length that slides along the time axis to operate a "time-localized" FT for the whole digital signal. The STFT utilizes the window function $g(t)$, that is centered at time τ by using a sliding approach. The time-localized FT is applied to the digital signal $x(t)$ together with the window for each particular τ value. Afterwards, the window is slid by τ throughout the time axis, and one more FT is carried out. Thus, the window function can be seen as a successive operation in order to implement FT on the entire signal. There is an important point that the small-signal portion within the window function is assumed to be nearly stationary. The stationary term means that the statistical properties of the signal does not change with respect to time. On the other hand, the non-stationary signals are formed of frequency components that are haphazard and change with time. As a consequence, the STFT separates the time domain signals into 2-dimensional time-frequency representation. Figure 3.1 indicates that the alteration of the frequency situation of the signal together with the window function.

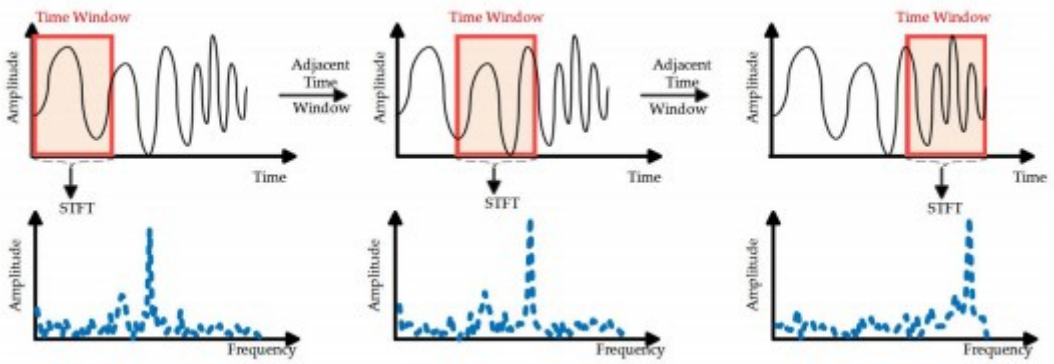

Figure 3.1. STFT by Sliding Time Window Function (Kumar, 2017)

The STFT can be denoted as,

$$STFT(\tau, f) =< x, g_{\tau,f} >= \int x(t)g(t-\tau)e^{-j2\pi ft}dt, \quad (4)$$

According to Equation (4), it indicates that the measurement of the similarity between the original signal $x(t)$ and time-shifted and frequency-modulated window function $g(t)$. In the last few decades, different kinds of window functions have been improved and all of the window functions are designed for specific applications with respect to signal characteristics. For instance, the Gaussian window is designed for transient signal analysis, Hamming and Hann's windows are designed for narrowband, random signals analysis, the Kaiser-Bessel window is designed for dividing two signals into components as its frequencies which are close to each other, and broadly various amplitudes. It should be considered that the selection of window function is directly related to the time and frequency resolutions. The STFT technique should not be selected arbitrarily according to Heisenberg's uncertainty principle (Cohen,1995).

Explicitly, the product of the time and frequency resolutions is lower bounded by,

$$\Delta\tau.\Delta f \geq \frac{1}{4\pi}, \quad (5)$$

where $\Delta\tau$ represents the time resolution and Δf denotes the frequency resolution. Mathematically, the time resolution $\Delta\tau$ is measured by the root-mean-square time width of the window function and described as,

$$\Delta\tau^2 = \frac{\int \tau^2 |g(\tau)| d\tau}{\int |g(\tau)|^2 d\tau}, \tag{6}$$

Correspondingly, the frequency resolution Δf is measured by the root-mean-square bandwidth of the window function and defined as,

$$\Delta f^2 = \frac{\int f^2 |G(f)| df}{\int |G(f)|^2 df}, \tag{7}$$

where $G(f)$ is the FT of the window function $g(t)$. In conclusion, the time-frequency resolution ensured by the window function when examining the signal $x(t)$ according to Equation (7).

Since the time and frequency resolutions of a function are just dependent on τ value, when the window function is selected, the time and frequency resolutions are being adjusted on the whole time-frequency plane.

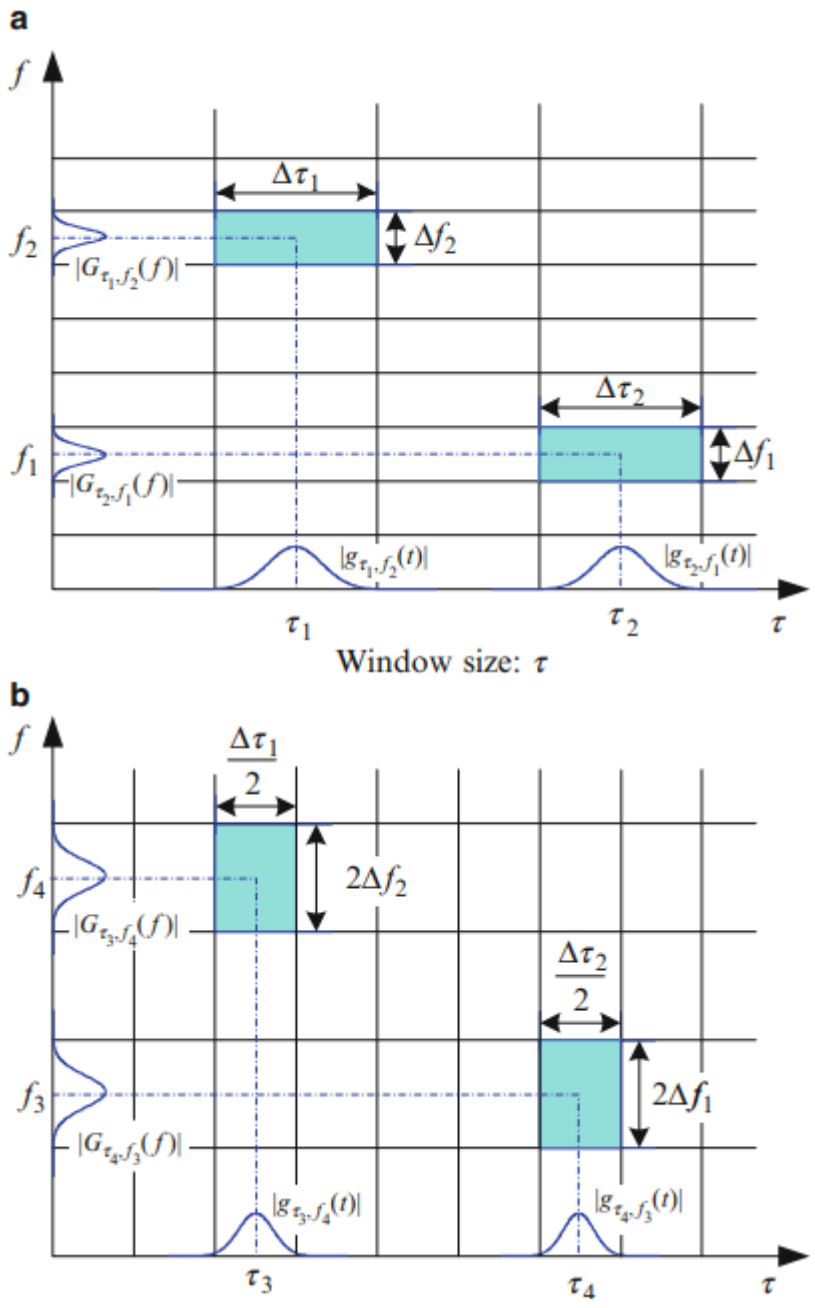

Figure 3.2. Time-Frequency Resolutions according to the STFT approach a) window size τ and b) window size $\frac{\tau}{2}$ (Gao,2011)

Figure 3.2 indicates that the window sizes τ and $\frac{\tau}{2}$ are described by the product of $\Delta\tau$ and Δf are independent of selected real window size as equal. Despite checking both time and frequency information at the same time is an enormous foundation, adjusting the fixed window size is still being hard enough.

In this thesis, STFT is applied to ECG signals which are obtained with the help of MATLAB.

3.3. Neural Networks

Neural Networks (NNs) have arisen from the human nervous system. The human nervous system can be seen as a perfect mechanism that is explained in the three main parts: brain, receptors, and effectors. The brain is the center of the human nervous system that receives and perceives information continuously and applies the decisions in itself. For that reason, the brain of the human nervous system is represented by a neural net. The receptors provide the transmission of stimulus from the human body or external environmental factors by transforming electrical impulses into the neural net. The effectors transform electrical impulses created by the neural net for the reaction as system output. Figure 3.3 indicates that the forward and backward transmission of the nervous system as a block diagram.

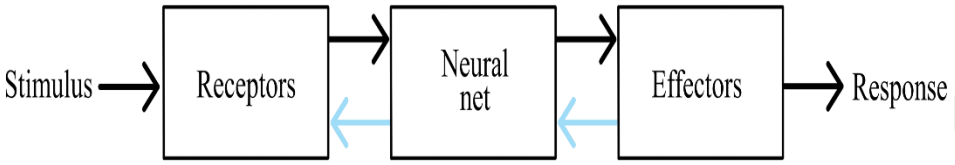

Figure 3.3. Representation of Nervous System

A neuron is a fundamental unit which maintains the process of a neural network as an information-processing relation. The array of synapses is viewed as connecting links between nerve cells, and also known as connecting links. The connection links are characterized by weight in themselves. In a neural network, neurons are responsible for receiving, processing, and transmitting information. The fundamental neuron model is composed of three main components such as connecting links, an adder, and an activation function. The adder performs as a summation of input signals according to their weights, also known as synaptic strengths of the neuron. The adder principle depends on the construction of the linear combination. The activation function operates in order to limit the amplitude range of the output signal to the finite value.

The normalized amplitude range of the output of a neuron is closed interval as [0,1] or [-1,1]. Figure 3.4 represents a non-linear neuron model that consists of input signals, synaptic weights, bias, summing junction, activation function, and output.

In Figure 3.4, externally applied *bias* is denoted by b_k which affects the input of activation function as increasing or decreasing. The neurons are indicated by k, successive input signals are $x_1, x_2, \dots, x_m$, and the synaptic weights are indicated by $w_{k1}, w_{k2}, \dots, w_{km}$ that is the k^{th} neuron weights, $\varphi(.)$ is an activation function of the model and y_k is the k^{th} neuron's output signal.

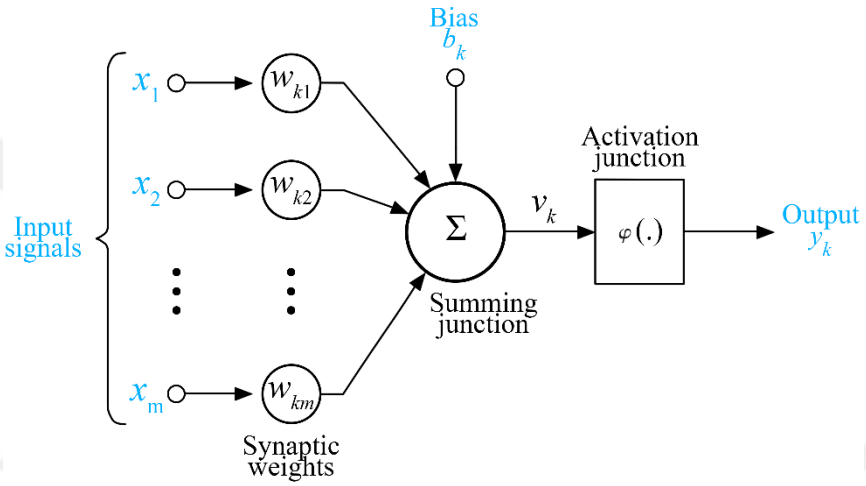

Figure 3.4. Model of Non-Linear Neuron

The neural network structures are closely related to the learning algorithms that are used in the training process of the network. Thus, the learning algorithms are also used in the design of neural networks for classification problems.

The different forms of learning operate on neural networks similar to human learning. There are different learning techniques and each of them is dissimilar. These learning techniques are able to categorize into three sections such as supervised learning, unsupervised learning, and reinforcement learning. All of them clarify that the instances can be categorized or not.

Supervised learning is related to the information about instances are known. Thus, the neural network is going to be trained according to the supervised learning approach.

Supervised learning problem can be written as Equation (8) and have desired output for input samples.

$$\tau = \{(\boldsymbol{x}_i, \boldsymbol{d}_i)\}_{i=1}^{N}, \tag{8}$$

where τ represents the instances for training with the number of N, $\boldsymbol{x}_i$ denotes the input signal vector and $\boldsymbol{d}_i$ also denotes the desired output that an array of labeled instances.

The difference between the desired output and actual output is known as the error of the neural network. The mean-squared error can be calculated from Equation (9) by taking average values of N instances.

$$E_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(d_i - \boldsymbol{w}x_i)^2, \tag{9}$$

where $\boldsymbol{w}$ is the synaptic weights multiplication, x_i denotes the actual output and d_i denotes the desired output.

Supervised learning updates weights by using error minimizing for each instance. In each training process, the network error updates itself until the network reaching the optimum point which is an error approximately zero. On the other hand, if the network performance wanted to be tested, the new samples are never used in the training process. Then, the performance of the network can discuss according to the evaluation criteria.

The neural networks have remarkable solutions for wide-ranging problems such as pattern recognition and classification. In the different layers of neurons, each layer receives input information from previous layers, and transfer the output layer. The input for the next layer is related to its weights and additional bias, which depends on the cost function and the optimizer. The neural network iterates for a pre-arranged number of iterations, called epochs until the cost function is minimized. The main concept of classification is that the features are extracted from the input signal in order to have information about the data, and the extracted features are given into the neural network for reduction data dimension or size in order to have steady and meaningful results.

Figure 3.5 indicates that the summary of the classification process for neural networks as a block diagram.

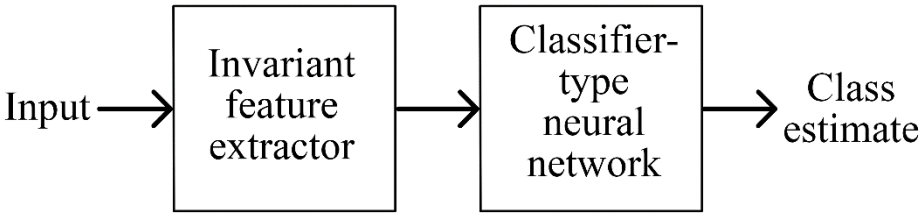

Figure 3.5. Neural Network Classification Block Diagram

There are many kinds of neural networks according to their purpose of design. The first design of the neural network is perceptron. The perceptron rule is used for the classification of linearly separable data as a binary classification technique. On the other hand, if data is not linearly separable, then an alternative neural network design has to be utilized for the classification of non-linear problems. So as to classify non-linear models, Multi-Layer Perceptron (MLP) is a well-known neural network architecture.

3.3.1. Multi-Layer Perceptron

MLP is a well-known neural network structure composed of one or more layers and shown in Figure 3.6.

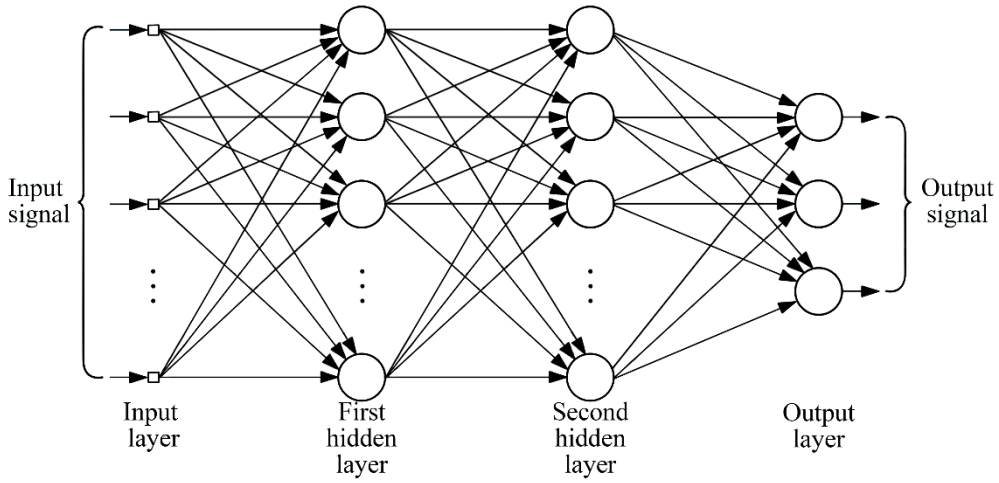

Figure 3.6. The Architecture of MLP with Two Hidden Layers

The fundamental features of MLP can be written as:

- Each neuron of the model includes a non-linear activation function which is differentiable.
- There are hidden layers on both input and output sides.
- The network demonstrates high degree connectivity in order to synaptic weight identification.

Signal flows in the forward direction which is left to right in the network on a layer-by-layer basis. Hidden neurons are the most invaluable element in order to function of MLP. The hidden neurons act like feature detectors. Since the learning process operates across the MLP, the hidden neurons start to search for discriminative features that qualify the training data. Also, applying non-linear transformation on the input signal into a new space, called a feature space that provides easy division from each other according to the original input space.

Each hidden neuron of an MLP is created to operate two computations:

1) The function signal occurring at the output of each neuron, which is explained as a continual non-linear function of the input signal and its weights related to that neuron.
2) The estimate of the gradient vector is the gradient of the error surface with respect to weights connected to the inputs of the neuron, which describes the backward transmission of the network.

In supervised learning, there are two basic approaches such as batch learning and online learning. In the batch learning approach, the synaptic weights of MLP are implemented for all N instances in the training set τ for one epoch of the training process. In addition to this, the cost function is defined by average error energy. Arrangements of weights depending on the epoch-by-epoch way. The instances of the training set are randomly shuffled for each epoch of the training process, and different initial conditions are set in a random way. Therefore, it is suitable for solving non-linear regression problems. In the online learning approach, training instances are demonstrated randomly in the neural network, and the multi-dimensional investigation is carried out. The synaptic weights arrangement of MLP according to example-by-example way. The online learning approach is

simple to implement and ensure powerful solutions for wide-ranging problems and complex pattern-classification problems (Haykin, 2009).

3.3.2. Convolutional Neural Network

Convolutional Neural Networks (CNNs) are suitable for data which has a spatial relationship. It is a special class of MLP and well-suited for pattern classification. Also, both MLP and CNN can be used for image and video classification, however; MLP uses input as a vector form, CNN uses images according to spatial relation. Specifically, CNN is designed for image classification and video classification. Spatial information means that having a location-based connection with other information, and space represents the two-dimensional plane in an image. In other words, CNN is a deep MLP structure designed for the identification of two-dimensional shapes with a high degree of invariance to translation, skewing, scaling, and some kind of distortion. CNN rely on in order to learn tasks in a supervised manner via the neural network has the following constraints:

- Feature extraction can be explained as each neuron takes its own synaptic inputs from a local receptive field in the previous layer in order to extract the local features. When features are extracted, their actual location becomes less significant, provided that its position relative to other features is nearly conserved.
- Feature mapping, when analyzing the neural networks are consists of multiple feature maps in each layer for computation. Each feature map can be seen as a plane that sharing the same bunch of synaptic weights, but it is actually a constraint. This constraint has also useful effects such as shift-invariance and reduction of the free parameters.
- Subsampling can be expressed as each convolutional layer is followed by a computational layer that operates local averaging and subsampling which provides to reduce the resolution of the feature map. Also, this provides to reduces the sensitivity of the feature map's output considering distortion issues. Figure 3.7 indicates the feature extraction and classification parts of the CNNs structure.

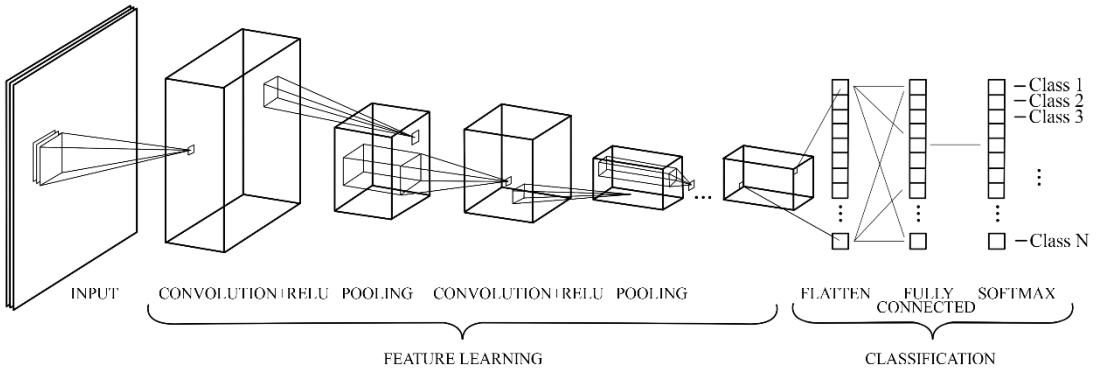

Figure 3.7. A CNN Architecture with Corresponding Layers

Day by day, CNNs have been accomplished to adapting into image classification and segmentation, object detection, video processing, natural language processing, speech recognition. The main importance of CNN is to have powerful learning ability due to the use of multiple feature extraction stages that can automatically learn representations from the data. During the whole training process, whole weights in all layers of the CNN by updating itself defines the learning process. CNN requires the usage of a big amount of data. CNN architecture includes multiple layers similar to feed-forward neural networks. The outputs and inputs of the layers are considered as a set of image structures. Moreover, CNN can learn automatically its own features without the need for any feature extraction techniques. On the other hand, CNN architectures can be constructed as a combination of some layers and each layer makes the different level of abstraction. The CNN layer description is explained as follows according to the structure in itself:

Input Layer: Images are given as inputs into the network.

Convolution Layer: This layer is the main component of the CNNs. Feature extraction is done by convolution layers. The reason for using the convolution process is to decrease the complexity of the network instead of using matrix multiplication in traditional neural networks. Kernel filters are applied to image input in the same way and are also used in image processing. In convolution layers, features are extracted and then sequentially transferred to other convolution layers. Thus, the obtained features will be the input of the next convolution layer.

Figure 3.8 represents the convolution operation applied kernel as 3-by-3 on 7-by-7 image input. The resulting pixel value is demonstrated after filtering as 5-by-5.

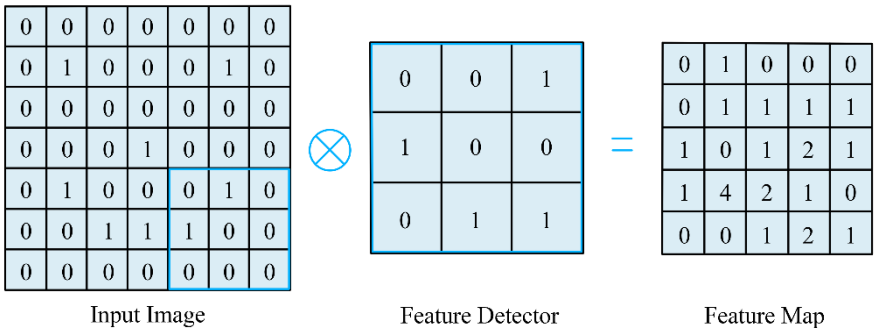

Figure 3.8. Representation of Convolution Operation

Pooling Layer: The pooling layer is used for the reduction of the feature dimensions. Therefore, the resolution of the feature map is reduced. More complicated features are extracted through each convolutional layer. After that, the input images are divided into non-overlapped rectangular sets. Each set is down-sampled with maximum pooling or average pooling. The pooling layer provides fast convergence and is preferable for generalization. This type of layer is generally located between the consecutive convolution layers. Figure 3.9 indicates that the average pooling returns the average of all values in the image covered by the kernel filter. Figure 3.10 indicates that the maximum pooling returns the maximum value in the image covered by the kernel filter. Stride means that sliding kernel filter according to stride value.

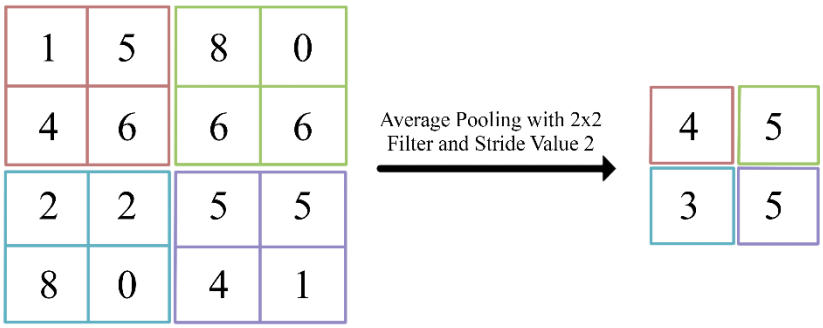

Figure 3.9. Representation of Average Pooling Process

However, maximum pooling better than average pooling because maximum pooling has a denoising effect in addition to dimensionality reduction.

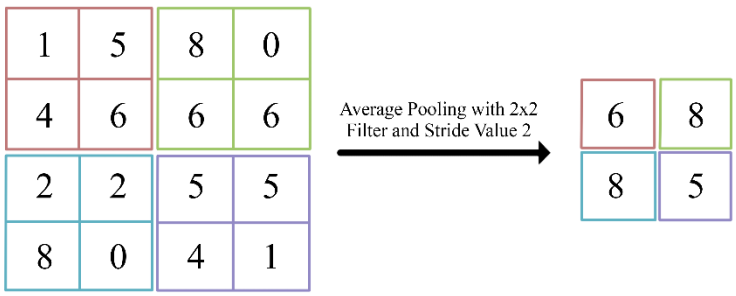

Figure 3.10. Representation of Maximum Pooling Process

Rectified Linear Unit (ReLU): This layer is mainly used as an activation function. It can be described as in Equation (10),

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}, \tag{10}$$

where $f(x)$ denotes the activation function of the CNN, and which means that a ReLU layer performs a threshold operation to each element of the input, where any value less than zero is set to zero. According to this function, the gradient descent algorithm does not backpropagate if the input to the activation layer is negative. This means that in a positive region and non-negative gradient is obtained. This provides us to increase the accuracy rate of CNNs.

Batch Normalization Layer: A batch normalization layer normalizes each input channel across a mini-batch. In order to speed up the training of CNNs and reduce the sensitivity to network initialization. The batch normalization layers between convolutional layers and nonlinearities, such as ReLU layers. Thus, normalized activations can be calculated as in Equation (11),

$$x_i' = \frac{(x_i - \mu_B)}{\sqrt{\sigma_B^2 + \varepsilon}}, \tag{11}$$

where the batch normalization normalizes its input x_i by first calculating the mean μ_B and variance σ_B^2 on related a mini-batch and related each input channel. ε develops numerical stability when a mini-batch variance is too small.

Fully Connected Layer: This layer is located at the end of the network. The fully connected layer is processed after applied are convolution layer, max-pooling layer, and ReLU layer consecutively. The neurons of the fully connected layer are linked to the previous layer activations. Hence, this is the final feature selection layer. After the final feature selection proceeds, the output is calculated as the average of multiplication of matrices and including additional bias. The property of this layer is the weights of the layers are forecasted with minimizing training error.

Softmax Layer: A softmax layer applies a softmax function to the input. For classification problems, a softmax layer and then a classification layer must follow the final fully connected layer. In the softmax layer, class probabilities are determined. The softmax function is the output unit function after the last fully connected layer for multi-class classification problems can be expressed as in Equation (12),

$$P(c_r|x,\theta) = \frac{P(x,\theta|c_r)P(c_r)}{\sum_{j=1}^{k} P\left(x,\theta\middle|c_j\right)P(c_j)} = \frac{\exp\left(a_r(x,\theta)\right)}{\sum_{j=1}^{k}\exp\left(a_j(x,\theta)\right)}\ , \tag{12}$$

where $0 \le P(c_r|x,\theta) \le 1$, and $\sum_{j=1}^{k} P\left(c_j\middle|x,\theta\right) = 1$. Furthermore,
$a_r = \ln\left(P(x,\theta|c_r)P(c_r)\right.$, $P(x,\theta|c_r)$ is the conditional probability of the instance given class r, and $P(c_r)$ is the class prior to probability.

Classification Layer: A classification layer computes the cross-entropy loss for multi-class classification problems with mutually exclusive classes. The cross-entropy function can be written as in Equation (13),

$$loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} t_{ij}\ln\left(y_{ij}\right), \tag{13}$$

where N is the number of instances, K denotes the number of classes, t_{ij} indicates that i^{th} instance belongs to the j^{th} class, and y_{ij} is the output for instance which is i for class j.

Parameters are the coefficients of the model, and they are adjusted by the model. When considered the algorithm during the training process, the model coefficients are optimized so as to minimize the error. In other words, the model parameters can be estimated by the model individually, and weights and biases are known as the model parameters in the neural networks. The parameters of the model do not require set manually, it can be estimated by optimization algorithms such as Stochastic Gradient Descent, Adam, and AdaGrad so on. In last, when the training process is finished, the model can decide how operates on unseen data.

Hyperparameters have to be adjusted manually. The model will not update the values of selected parameters. Therefore, hyperparameters should set before the model starts the training process. These parameters cannot be learned by the model. Moreover, the number of hidden layers, learning rate, momentum, mini-batch size, activation function, and epochs are the most common hyperparameters of the neural network. In order to clarify the idea, the learning rate represents the algorithm of how significant effect of the gradient on the weight. If an unsuitable learning rate is selected for the neural network, gradients may vanish or diverge.

According to optimizers, the neural networks are adjusted by themselves such as weights and biases in order to minimize the losses.

3.4. Optimization Algorithms

Optimizers are commonly used algorithms in order to alter the behavior of the neural network. Due to the importance of reducing loss function, weights and learning rates of the neural network should be adjusted in order to optimize algorithms. The fundamental purpose of the optimizers tries to figure out optimization problems via minimizing the loss function. Furthermore, optimization algorithms are responsible for diminishing the cost function and ensuring as high accurate results as possible. In neural networks, the fundamental method is finding the best optimization algorithm

which ensures the optimal solution. There are four optimization algorithms are mentioned in this section such as Gradient Descent, Stochastic Gradient Descent, Stochastic Gradient Descent with Momentum, and Adaptive Moment Estimation.

3.4.1. Gradient Descent Algorithm

The gradient descent (GD) algorithm is the most fundamental algorithm that relies on the first-order derivative of an objective function. The characteristic of GD is to find how to alter the weights in order to reach a minimum in the objective function. The cost is transmitted from one layer to another layer via the back-propagation algorithm, and the weights which are also known as model parameters are adjusted according to the cost function. Therefore, the cost function can be minimized by using this aspect. The weights are updated until converge to the minimum point. GD algorithm has the useful property is very simple to implement the neural networks. On the other hand, there are some disadvantages to using the GD algorithm. The GD algorithm takes all samples in the dataset in order to compute derivatives of weights for updating them. This means that the neural networks require much more memory because of computing derivatives of gradients at one iteration. The GD algorithm can reach the minima after a long training time but may not reach also. The dangerous point of the algorithm is to stick the saddle point or local minima. That is why the algorithm must be reached a global minimum in order to explore the best solution for the neural network. In Equation (14), the basic principle of weight update is shown as,

$$w_{new} = w_{old} - \alpha(\nabla_w J)w_{old}\,, \tag{14}$$

where α represents the learning rate or step size, J represents the loss function, $\nabla_w J$ denotes the gradients of the loss function, and w_{new} represent the computed new weight via using previous weight w_{old}.

3.4.2. Stochastic Gradient Descent Algorithm

Stochastic Gradient Descent (SGD) algorithm can be seen as a prolongation of GD to figure out the disadvantages of that algorithm. SGD algorithm is a stochastic approximation of GD optimization since it computes gradients randomly selected subset of the whole dataset than computing from the whole dataset. SGD algorithm

derivative is computed for one subset of the entire dataset at a time. Also, the SGD algorithm takes more steps in order to reach the minimum point comparing with the GD algorithm. That is why, SGD algorithm can be observed as a little bit noisier than the GD algorithm, and takes more time to converge towards the global minimum point. This algorithm is also responsible for the convergence towards the optimum point of the objective function in neural networks. So as to solve the optimization problem, the current parameters which are known as weights are computed using the objective function. After that, the gradient computation is done for updating each parameter in the neural network. Also, the parameters will be updated according to against direction of the gradient by multiplying with the learning rate value. These stages are maintained until converge to the optimum point or until a particular epoch number is applied. There are some drawbacks of SGD these are mainly,

Since the SGD algorithm takes more steps according to changing loss, it causes a slow learning effect and long computation time.

If the objective function has a saddle point or local minimum, the SGD algorithm will be stuck in local optimum or on saddle point because of the zero gradients in total for this particular region.

The gradients can be noisy because of the computation on mini-batches.

The equation of SGD in order to update parameters in the neural network using back-propagation so as to calculate the gradient according to Equation (15),

$$\theta = \theta - \alpha \nabla_{\theta} J(\theta; x, y), \tag{15}$$

where θ denotes the parameters that are weights, biases, and activations, α is the learning rate or step size, $\nabla_{\theta} J(\theta; x, y)$ represents the back-propagation computation, ∇ is the gradient, J is the loss function, and $J(\theta; x, y)$ represents the parameter θ as input throughout the training instance x and its corresponding label y.

3.4.3. Stochastic Gradient Descent with Momentum Algorithm

Momentum term that ensures us to develop faster computation in the neural network. This means that the algorithm with the momentum term provides to reaching the optimum point in a faster way. If the objective function is much more complex, the

algorithm may stick in the local minimum. The momentum term can be seen as an additional component into weights, biases, and activations in the neural network, also that is known as a component of time. The time component boosts the momentum. While optimizing the cost function with the SGD algorithm, oscillation can occur along the y-axis and slower progression along the x-axis. The SGD with momentum ensures that fewer oscillations along the y-axis and reach faster to the local minima. Figure 3.11 indicates that the comparison of how fast the algorithm with momentum and oscillation condition.

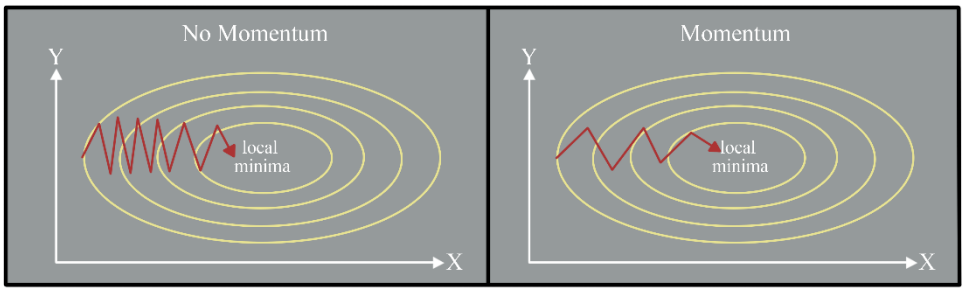

Figure 3.11. Comparison of SGD and SGDM

The momentum term how actually works, it collects more speed in each epoch. Thinking the way of a steady γ, and previous update γv_t to θ. On the other hand, updating θ consist of the second last update to θ and goes on this aspect. It is clearly seen that the computations of the gradients are stored for utilizing the following updates to parameter θ in an updating way. This idea summarized in Equation (16) as,

$$\theta_t = \theta_t - \alpha \nabla J(\theta_t) + \gamma \sum_{\tau=1}^{t} \gamma \nabla J(\theta_\tau)\,, \tag{16}$$

where θ_t denotes the parameters such as weights, biases or activations, ∇ is the gradient operation, α is the learning rate or step size, t denotes the current time step, J is the loss function that trying to minimize, γ is a steady momentum term, and the last update to θ is v_t.

Additionally, the saddle point problem is reduced by using the past gradients via momentum term, even the gradients are zero or close to zero in that particular area.

3.4.4. Adaptive Moment Estimation Algorithm

Adaptive Moment Estimation (ADAM) is one of the most commonly used algorithms that operates the best on averages. Adam utilizes both momentum and adaptive learning rates in order to converge faster to the minimum point. Adaptive learning rates mean that the neural network begins taking big steps and end with the small steps. Due to the learning rate decays, the smaller steps are taken in order to converge faster. Therefore, the neural network does not exceed the local minimum with these big steps.

The Adam algorithm aims to minimize the loss function differentiation with respect to parameter θ, also the algorithm is interested in the expected value of the loss function with respect to parameter θ. Adam optimizer stores both first-order and second-order moments of the gradient, mean and uncentered variance respectively. Exponential moving averages for the first-order moment and the second-order moment are given in Equation (17),

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \end{aligned} \tag{17}$$

where m and v are moving averages, g is a gradient on the current batch, t is the number of iterations and β_1 and β_2 are hyperparameters of the algorithm.

Due to the zero initialization of the moving averages at the beginning of the iteration, bias correction operation will be needed after computation exponential moving averages of the first order and second-order moments. The algorithm updates exponential moving averages of the gradient and squared gradient with the hyperparameters. The hyperparameters check the exponential decays of the moving averages these are also known as mean and variance. The bias correction operation for the first order and second-order moments are given in Equation (18),

$$\begin{aligned} \widehat{m_t} &= \frac{m_t}{1-\beta_1^t} \\ \widehat{v_t} &= \frac{v_t}{1-\beta_2^t} \end{aligned} \tag{18}$$

where $\widehat{m_t}$ and $\widehat{v_t}$ are bias-corrected estimators, m_t and v_t are the first order and second order moving averages of the gradient, β_1 and β_2 are the hyperparameters of the algorithm, and t denotes the iteration number.

The final step of the algorithm is the moving averages to scale the learning rate individually for each parameter according to the weight update function. The weight update function is given in Equation (19),

$$w_t = w_{t-1} - \alpha \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon}, \tag{19}$$

where w is the weight of the model, α is the learning rate or step size, t is the iteration number, and ε is to prevent division by zero and the default value is 10^{-8}, and hyperparameters must satisfy this condition $0 \leq \beta_1, \beta_2 < 1$, and also $0 < \alpha < 1$.

In this thesis, two main optimization algorithms are investigated such as Stochastic Gradient Descent with Momentum and Adaptive Moment Estimation optimizers.

CHAPTER 4
ARRHYTHMIA DETECTION

4.1. Data Set

The MIT-BIH Arrhythmia Database has been supported by Boston's Beth Israel Hospital laboratory and MIT in order to analyze arrhythmia and connected topics in their investigations. The database has been published as test material in order to consider cardiac arrhythmias and has been used for investigations which are related to cardiac dynamics.

The ECG signals analysis has great significance for the early diagnosis of cardiac arrhythmias. Many of the studies about cardiac dynamics are obtained from Boston's Beth Israel Hospital laboratory and MIT-supported long-term ECG records in the literature (Moody and Mark, 2001).

In this thesis, ECG records were acquired from Physionet public website that includes the MIT-BIH Arrhythmia Database in itself (MIT-BIH Arrhythmia Database Directory, 2018). The acquired ECG records were used in order to train, test, and validate the proposed system by using state-of-the-art techniques. The MIT-BIH Arrhythmia Database encompasses the forty-eight records are related to cardiac arrhythmias, and records are sampled at 360 Hz with a precision of 11 bits at 10 mV range. In addition to this information, all of the ECG records have clinical wave structure and complex characteristics.

The ECG records were indicated in accordance with its labels with numbers from 100 to 234 in the MIT-BIH Arrhythmia Database. Table 4.1 indicates that all ECG signals were used according to the record number, number of beats, and patients' age and gender information in this study.

Table 4.1. Metadata of MIT-BIH Arrhythmia Database Records

Record No	Gender	Age	Beat Type	Number of Beats
100	Male	69	Normal	2237
112	Male	54	Normal	2535
115	Female	39	Normal	1952
122	Male	51	Normal	2474
123	Female	63	Normal	1514
106	Female	24	PVC	520
119	Female	51	PVC	444
200	Male	64	PVC	826
203	Male	43	PVC	444
118	Male	69	RBBB	2165
124	Male	77	RBBB	1530
207	Female	89	RBBB	85
212	Female	32	RBBB	1825
109	Male	64	LBBB	2490

4.2. Pan Tompkins Algorithm for QRS Complex Detection

The biomedical signals which are measured from the human body, these signals have crucial importance in order to understand the meaning of biomedical signals in each type of individual. In addition to this, all of the biomedical signals enable medical experts to diagnose the diseases. In particular, ECG signals are widely used in order to detect cardiac abnormalities. That is why the structure of the ECG signals has to be scrutinized carefully by medical experts. On the other hand, the ECG signals must be segmented properly according to the signal structure by engineers. Algorithms have been developed in order to segment and investigate the ECG signals in the literature. In this thesis, the Pan Tompkins algorithm was used for the detection of single QRS complex from real-time ECG signals based on analysis of slope, amplitude, and width of the QRS complex (Pan and Tompkins, 1985). Pan Tompkins algorithm encompasses sequential steps that are low pass filtering, high pass filtering, differentiation operation, squaring function, and moving-window integration.

The Pan Tompkins algorithm comprises four main stages, and these stages are demonstrated in Figure 4.1.

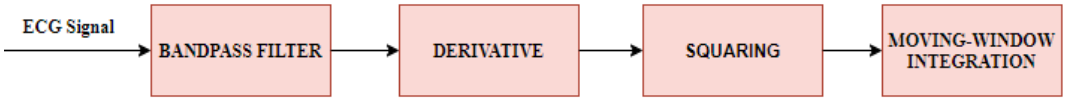

Figure 4.1. Block Diagram of Pan Tompkins Algorithm

The outcoming signals are undergone sequential processes by using three linear filters applied with the aid of the software platform. The first filter is known as the bandpass filter which incorporates low-pass and high-pass filters respectively. The second filter enables us to derivative approach. Then, the squaring function is applied to the digital signal according to its amplitude. At the end of the squaring function, the next step is moving integration. Thus, adaptive thresholds are applied in order to obtain discriminative locations of the QRS complexes.

The bandpass filter diminishes computation complexity and the effect of muscle noises occurred by produced interferences from power lines. ECG signals are highly sensitive because of measurements in mV ranges. This produced interference might be transmitted during measurements of cardiac signals by cables of ECG medical devices. The approach of working in the z-plane provides us to design of filter coefficients as integers. The transfer function of the Low-Pass Filter (LPF) was created as in Equation (20),

$$H_{lp} = \frac{(1 - z^{-6})^2}{(1 - z^{-1})^2}, \tag{20}$$

The convolution operation was applied to the digital signal. The digital signal is convolved with the transfer function of LPF. The convolution property can be written according to Equation (21),

$$y(n) = \sum_{k} x(k)h(n - k + 1), \tag{21}$$

$y(n)$ denotes the output of the convolution operation and h denotes the filter coefficients as vector form that is used for sliding on $x(k)$ input signal, the overlapped area is obtained after convolution. In addition to this, LPF helps to eliminate 60 Hz components of the signal (Rangayyan, 2015).

High-Pass Filter (HPF) design is constructed on this approximation that subtraction the output signal of LPF from allpass filters. The transfer function of the HPF is denoted in Equation (22),

$$H_{hp} = z^{-16} - \frac{(1 - z^{-32})}{(1 - z^{-1})}, \tag{22}$$

The convolution process is described in Equation (21), performed also in the HPF step. The low cut-off frequency of the HPF is adjusted nearly 5 Hz, and also delay is assigned as 16 samples.

The derivative operation is used for compression of the low-frequency elements of both P wave and T wave and enables major gain to the high-frequency elements that are risen from QRS complexes because of the high slopes. According to Equation (23), the derivative operation of the digital signal can be written as,

$$y(n) = \frac{1}{8}[2x(n) + x(n-1) - 2x(n-2) - x(n-3)], \tag{23}$$

where $y(n)$ denotes the output of the digital signal that comprised of convolution operation onto $x(n)$ that represents the input signal with the impulse response.

The main idea of the squaring function is the squared input signal point by point after the derivative operation. The squaring function takes all data points into the positive region and applies non-linear amplification in order to show dominant ECG frequencies. The high-frequency elements of the signal in accordance with QRS complexes are improved.

The main purpose of moving-window integration is removing multiple peaks inside of single QRS complex which is caused by the derivative-based process. The moving-window integration operator can be seen as a filter by a sliding window on the output vector obtained from the squaring function. This step of the Pan Tompkins algorithm performs in order to provide a smoother output signal. The moving-window operation can be written according to Equation (24),

$$y(n) = \frac{1}{N}\{x\big(n - (N-1)\big) + x\big(n - (N-2)\big) + \cdots + x(n)\}, \tag{24}$$

where N denotes the width of the window. The selection of the width of the window has great importance in the evaluation of the output. If the window width is selected too wide, the integration waveform structure will be joined the QRS complexes and T

waves. On the other hand, if the window width is selected too tight, it is caused by the generation of several picks in the integration waveform. That is why it causes false peak detection because of subsequent picks. A window width of 30 samples has been selected proper width for 200 Hz sampling frequency (Pan and Tompkins, 1985; Rangayyan, 2015).

The QRS complexes are consistent with the arising point of the integration waveform. The width of the QRS complex is denoted by the time duration of the arising point. The maximum slope or R wave peak is defined by the arising point of the desired QRS structure.

The set of thresholds are applied to the integration waveform in the Pan Tompkins algorithm. The thresholds are adjusted automatically, and the higher of these thresholds is used in order to analyze the signal in the initial phase. The search-block technique is used if no QRS complex is detected in a related time interval, in order to find the QRS complex by applying a lower threshold. This means that the Pan Tompkins algorithm can detect a peak as R-peak in order to obtain QRS complex or noise as a peak which is irrelevant to the QRS complex (such as T-wave). If a new peak is detected, it has to be classified as a signal peak or a noise peak. If a peak exceeds a threshold, it is classified as a QRS peak.

After that, the Pan Tompkins algorithm computes the average value of R-R intervals. R-R intervals range is adjusted such as both high limit of R-R interval and low limit of R-R interval by using the average of RR intervals. The maximum value between these low and high limits using the search-back process, the located point identified as R peak. The adaptive threshold method provides a dynamic system in order to detect R peaks. In addition, Heart Rate (HR) is derived according to Equation (25),

$$HR = \frac{60}{RR_{average}}, \tag{25}$$

In this thesis, the Pan Tompkins algorithm is used in order to obtain QRS complexes from the raw ECG signal with the aid of the MATLAB 2018b implementation platform. The explanation of the Pan Tompkins algorithm is summarized in some stages as following,

1) The raw ECG signals are obtained from the MIT-BIH Arrhythmia Database. The obtained raw ECG signals are converted into a .mat file as their classes. The raw ECG signal as input is shown in Figure 4.2. (a).
2) DC components of the raw ECG signal are eliminated by computing the mean value of the raw ECG signal and applied subtraction operation by using this mean value from raw ECG signal components.
3) The normalization process is applied in order to examine signals in the same amplitude range by computing the absolute maximum value of the ECG signal. The result of stages 2 and 3 is shown in Figure 4.2. (b).
4) The convolution operation is applied by using coefficients of LPF after stage 3 in order to catch simplicity. The result of this stage is demonstrated in Figure 4.2. (c).
5) The convolution operation is applied by using coefficients of HPF, as same as stage 4. Also, the result of this stage is shown in Figure 4.3. (a).
6) Again, the convolution process is applied by using these coefficients in order to apply the derivation process. Figure 4.3. (b) indicates that the result of stage 6.
7) The squaring function is applied to the ECG signal, impulse response, and squared ECG signal are convolved in this stage. In addition, the result of this stage is shown in Figure 4.4. (c).
8) The convolution operation is applied with window filter coefficients in the moving-window integration process as same as stage 4.
9) According to the adaptive threshold, peak values and locations are detected. The detected Q, R, and S points are indicated in 30 seconds intervals in Figure 4.4. (a) and Figure 4.5. (a). Moreover, Q, R, and S points which are detected are shown in 3 seconds in order to see more clearly and in detail in Figure 4.4. (b).
10) Each ECG signal is 260 samples long, and R-peaks are detected as the maximum value of the peaks by dividing the signal into the left and right sides. In the end, entire QRS complexes are extracted from each record in the MIT-BIH Arrhythmia Database, and the extracted QRS complexes are shown as single beat form that is related to normal sinus rhythm in Figure 4.5. (b).

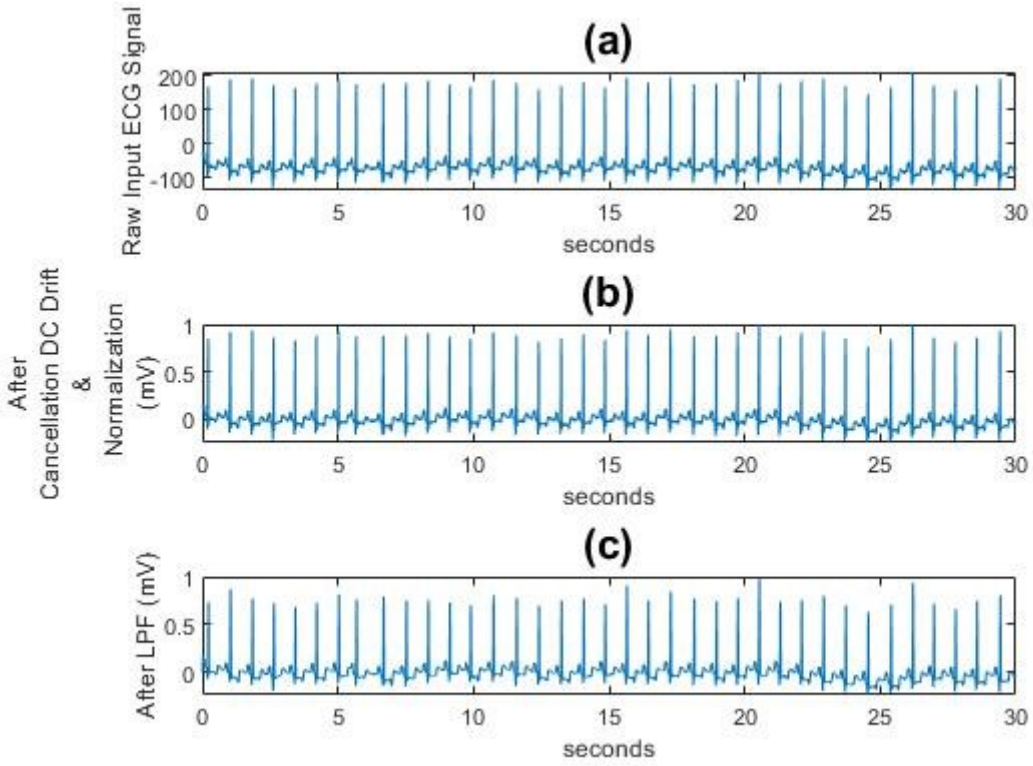

Figure 4.2. Implementation of Pan Tompkins Algorithm (a) Raw ECG Signal as an input, (b) Cancellation DC Components and Normalization in Figure 4.2. (a), (c)Result of LPF application in Figure 4.2. (b).

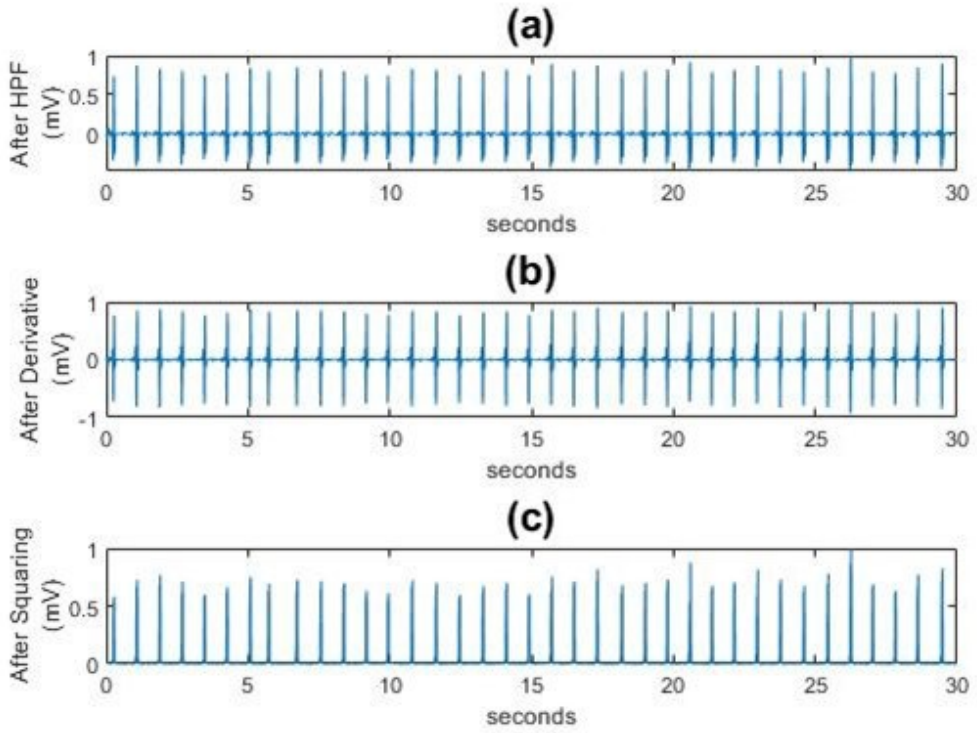

Figure 4.3. Preparation of R-peak Detection (a) Result of HPF application in Figure 4.2. (c), (b) Result of Derivation process in Figure 4.3. (a), (c) Result of the squaring process in Figure 4.3. (b).

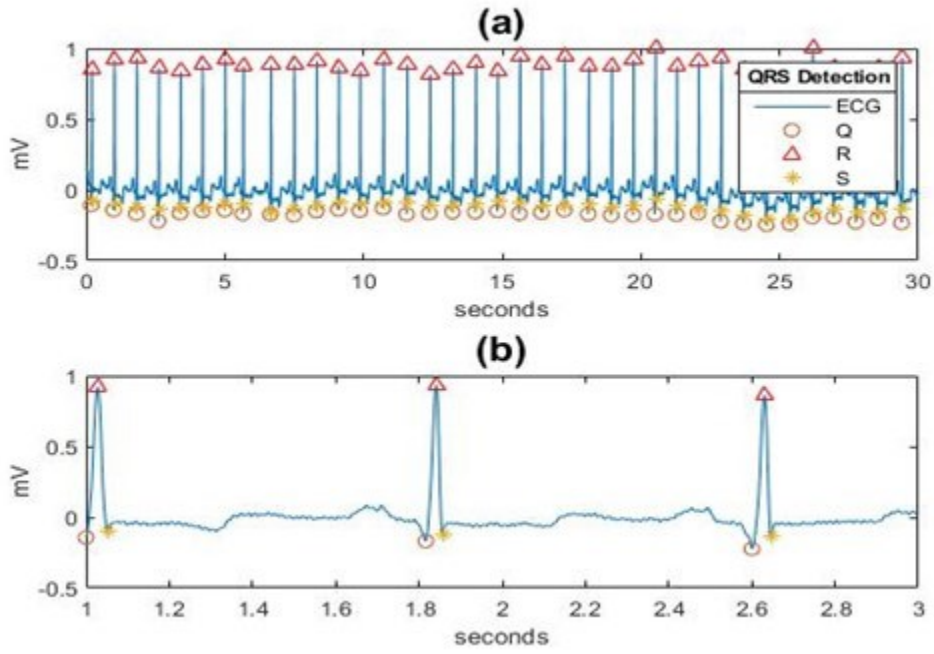

Figure 4.4. Q-R-S Point Detection (a) Located Q, R and S Points in 30 seconds duration, (b) Located Q, R and S Points in 3 seconds duration.

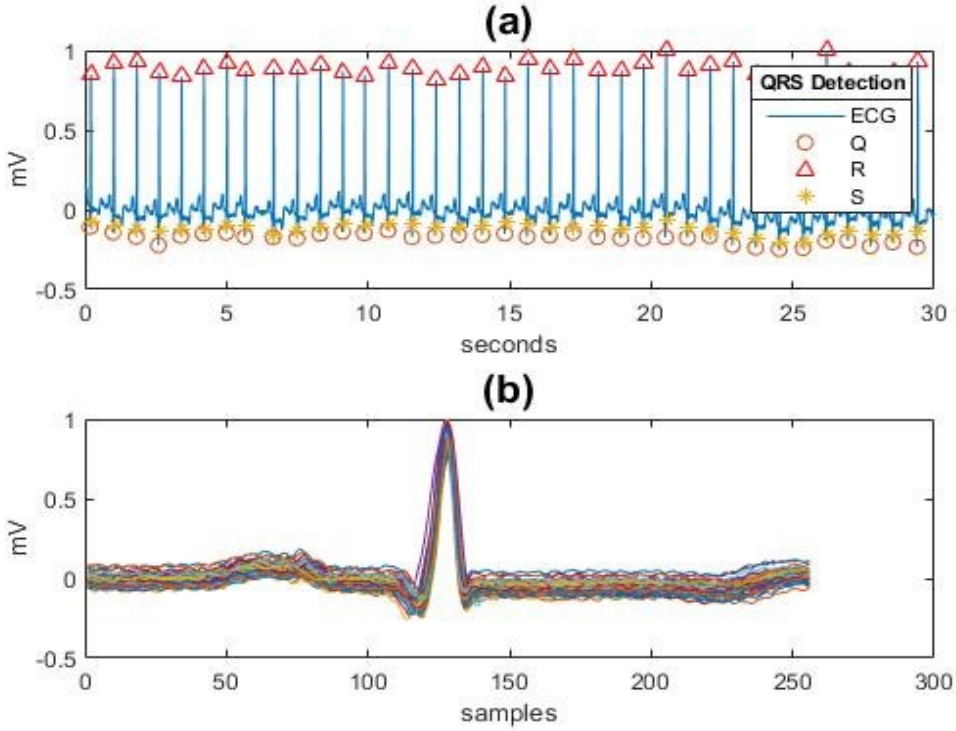

Figure 4.5. Extraction of QRS Complexes (a) Located Q, R and S Points in 30 seconds duration, (b) Bunch of single QRS complexes in one record.

4.3. ECG Arrhythmia Classification with Time-Series and Time-Frequency Approaches

In this proposed study, a large number of data were classified according to arrhythmia types. The first approach was to categorize time-series ECG signals as raw form without applying any feature extraction technique. In order to elaborate further, the same number and the same type of time-series ECG signals are transformed into the time-frequency domain. The main concept of this transformation is to analyze and cope with both time and frequency information of the signals. This kind of biomedical signals (such as ECG signals) have to be analyzed by using suitable state-of-the-art techniques.

In recent years, the major problem is to determine the selection of suitable signal processing techniques compatible with a big amount of data. Moreover, the comparison of both two approaches has great significance in order to reach reasonable conclusions. ECG signals are known as non-stationary with regard to different frequency intervals. According to the characteristics of non-stationary signals,

frequency components of the signals have to be scrutinized in a good sense. The several frequency components can be analyzed in different time intervals by considering the properties of non-stationary signals.

Furthermore, there is a requirement for transferring data according to the particular features of the data. Because of this reason, artificial neural networks, machine learning-based algorithms, and deep neural networks have been developed in order to indicate successful results with regard to data that is used in related studies as mentioned in Chapter 1. The most well-known property of deep neural networks is not required in any feature extraction method. It is stemmed from two main points, selection of the feature extraction method is still challenging according to data and these features can be obtained from special layers of CNN.

4.3.1. ECG Time-Series Signal Classification with CNN

The time-series ECG signal classification approach was used in this study. The examination of time-series ECG signals was done in terms of time-domain properties. The basic concept was to understand how CNNs are accomplished on time-domain signals in this part of the proposed study. Because the CNNs have great accomplishment rates on image classification. By looking at this aspect, in order to see the difference between time-domain signals and time-frequency domain signals via using CNN according to accomplishment rates was performed.

Fundamentally, three types of ECG signals as the time-series were investigated in order to classify heartbeats by using CNN. Figure 4.6 represents the time-series normal sinus rhythm ECG signal, Figure 4.7 indicates the time-series PVC ECG signal, and Figure 4.8 denotes the time-series RBBB ECG signal in one-second duration.

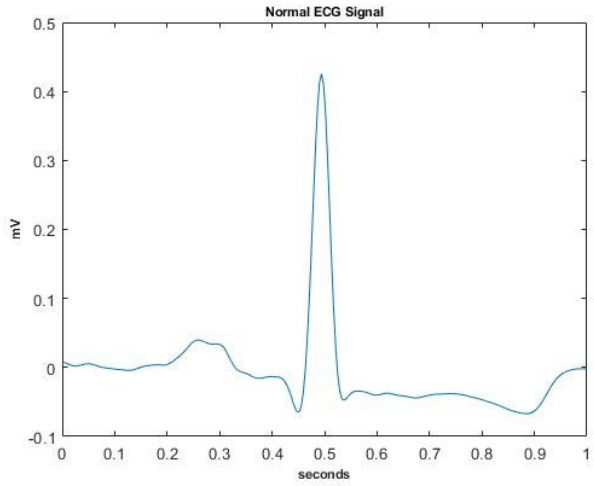

Figure 4.3. Representation of Single Normal Sinus Rhythm in Time Domain

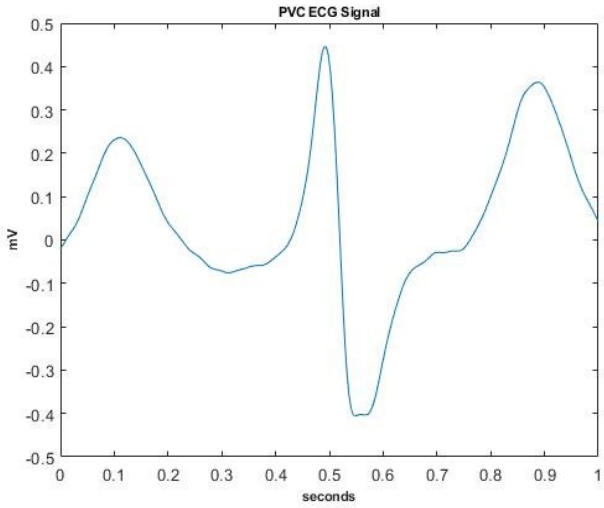

Figure 4.4. Representation of Single PVC Arrhythmia in Time Domain

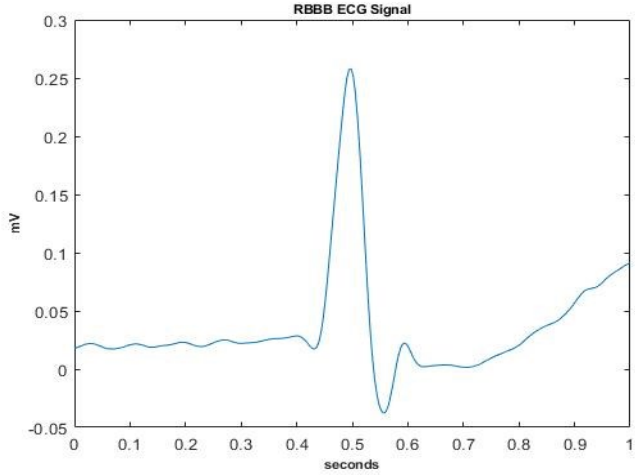

Figure 4.5. Representation of Single RBBB Arrhythmia in Time Domain

In addition, the ECG time-series signals are given in Table 4.2 according to record numbers and the corresponding number of beats that are used in this study.

Table 4.2. MIT-BIH Records and Beats for the Time-Series CNN Approach

Record No	**Type of Beat**	**Number of Beats**
112	Normal	2535
115	Normal	1952
122	Normal	2474
123	Normal	1514
106	PVC	520
119	PVC	444
200	PVC	826
203	PVC	444
118	RBBB	2165
124	RBBB	1530
207	RBBB	85
212	RBBB	1825

The three types of ECG signal shapes were obtained after applying the Pan Tompkins algorithm. The obtained single QRS complexes according to labels, record numbers and construction of beat matrix as time-series ECG signals in compact matrix form are given in Figure 4.9.

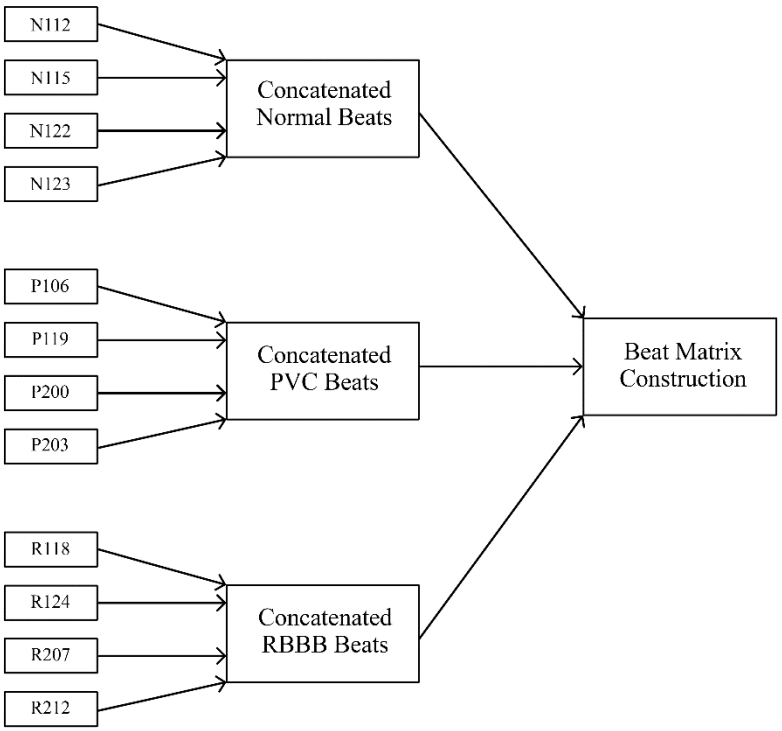

Figure 4.6. Representation of Constructed Beat Matrix with Related Records

After that, the constructed beat matrix has a large volume of data transformed into a four-dimensional array. This process provides us to input data set can behave like an image. Because of the property of CNNs, the image inputs have been used in order to obtain high achievement rates. Figure 4.10 represents the summary of the proposed framework with the large volume of input data and the total number of beats which are used in this study.

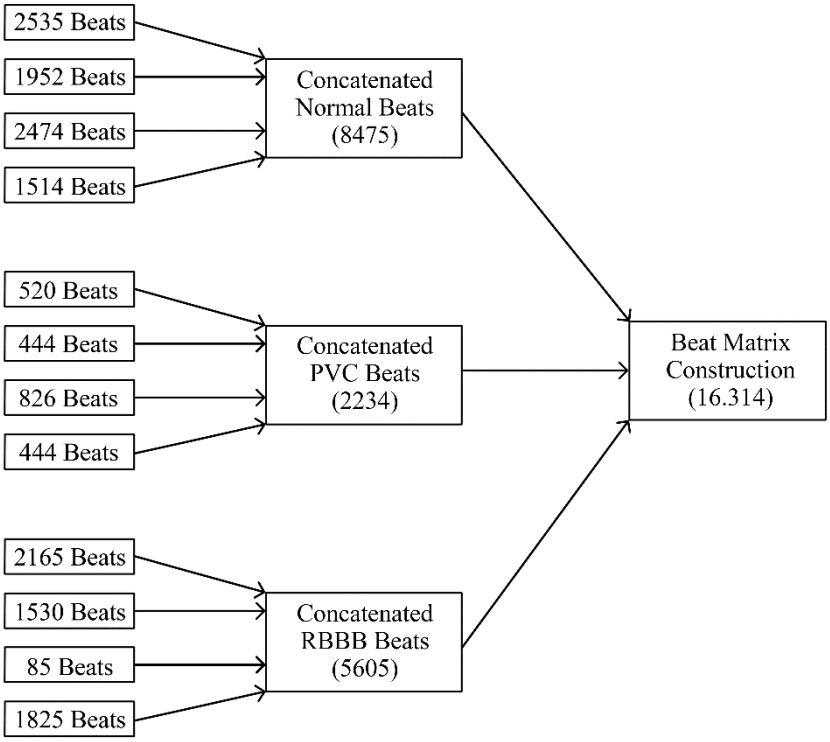

Figure 4.7. Representation of Constructed Beat Matrix with Related Beats

The summary of the time-series ECG signal classification process can be explained in the following stages, and also shown in Figure 4.11.

1) The unprocessed data were taken from the MIT-BIH Arrhythmia Database. Entire data were utilized as single QRS complex form after stage 2. Then, these obtained QRS complexes were recorded as a .mat file.
2) The peak values and locations of the QRS complexes were detected with the aid of the Pan Tompkins algorithm for each type of heart rhythm such as Normal, PVC and RBBB. After the detection of R-peaks, QRS complexes were defined by taking R-peak as a base, 129 samples were taken from the left side of the R-peak and 130 samples were taken from the right side of the R-peak. This means that QRS complexes were defined with 260 samples long. For that reason, the size of the input data was adjusted as a 1x260 vector form.

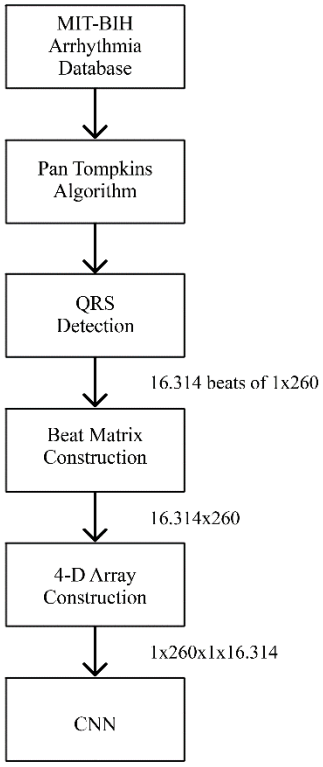

Figure 4.8. The Summary of Proposed Study in Time Domain

3) The obtained QRS complexes were concatenated according to heartbeats which were used in this study. The number of beats was shown as a class-wise in Figure 4.10 until the construction of the beat matrix.
4) The constructed beat matrix has a large volume of data as 16.314 beats. Moreover, 8475 beats were taken from the Normal class, 2234 beats were taken from the PVC class, and 5605 beats were taken from RBBB class. As a result of this stage, the output dimension of the constructed beat matrix was acquired as 16.314x260 matrix form.
5) After the construction of the beat matrix, it was transformed into the four-dimensional array. This stage mainly provides us to use the data in matrix form for CNN. Because the input dimension was 1x260 as a vector form and 1x260x1 was represented with the third dimension as a channel, the 1x260x1x16.314 was exhibited with the fourth dimension as the number of total QRS complexes in order to make it compatible with CNN. Normally, CNN reads the images from a particular folder, but it is not suitable for the

time-series signal classification. This transformation was carried out for this reason.

6) In conclusion, the constructed four-dimensional array was given into CNN as input according to adjusted data split and model architecture.

The constructed beat matrix was obtained after applying the first four stages. This constructed beat matrix was divided into training, testing and validation sets as 70%, 15% and 15%, consecutively and randomly. The summary of the study is indicated as a block diagram in Figure 4.12.

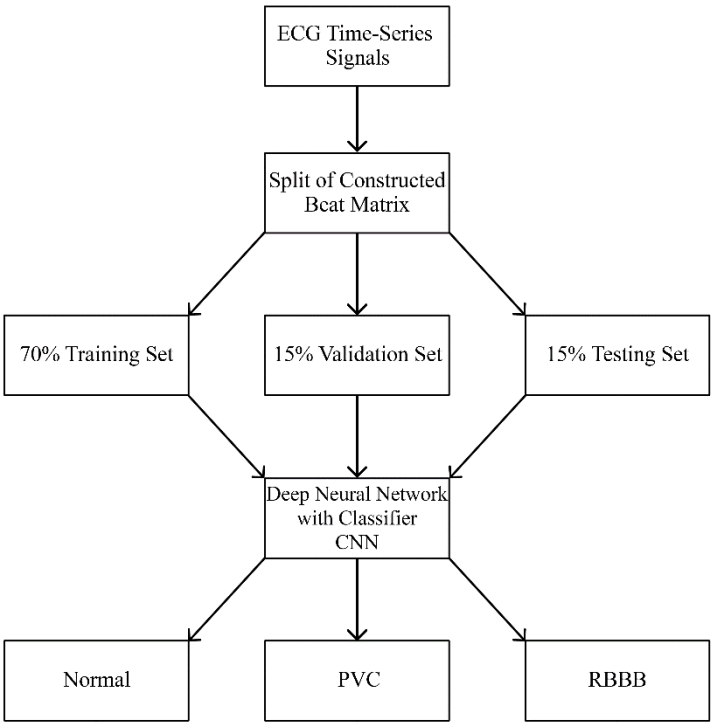

Figure 4.9. The Block Diagram of the Time Domain Classification

This block diagram summarizes the data split according to the most common division percentages. The next step was to determine the classifier model related to the deep neural networks approach after the data split process. Then, CNNs are highly capable of selecting discriminative features. The deep neural network layers and their corresponding functions were explained and the main concept of using CNN was clarified in Chapter 3. Thus, the model architecture was created according to Table 4.3.

Table 4.3. The Proposed CNN Architecture for Time-Series ECG Signal Classification

Layers	Layer Types	Output of Layer	Filter Size	Stride
1	Vector Input	1x260x1	-	-
2	Convolution	1x260x8	[3,3]	[1,1]
3	Batch Normalization	1x260x8	-	-
4	ReLU	1x260x8	-	-
5	Max Pooling	1x87x8	[1,1]	[3,3]
6	Convolution	1x87x16	[3,3]	[1,1]
7	Batch Normalization	1x87x16	-	-
8	ReLU	1x87x16	-	-
9	Max Pooling	1x29x16	[1,1]	[3,3]
10	Convolution	1x29x32	[3,3]	[1,1]
11	Batch Normalization	1x29x32	-	-
12	ReLU	1x29x32	-	-
13	Fully Connected	1x1x3	-	-
14	Softmax	1x1x3	-	-
15	Classification Output	3	-	-

The proposed CNN model was constructed according to model layers and specific training option. The initial learning rate is set to 10^{-5} and the validation frequency is set to 50. This implies that the data will be validated in every 50 samples. This progress is completed in 28 seconds, it is the elapsed time of the network. The training process is completed in 10 epochs and 89 iterations per epoch. It defines the maximum iteration of the network. Epoch term can be seen as a complete pass across all data set. In addition, the iteration term can be defined as an estimation of the gradients and updating of the network parameters in each iteration. Data were shuffled in every epoch. This explains the shuffling of the train data before starting in each epoch and shuffling the validation data before starting in each epoch in order to apply the network validation process. An iteration is one step taken in the gradient descent algorithm towards minimizing the loss function. The solver for the training network is set as the SGDM optimizer, and the momentum term is set to 0.9.

In conclusion, when classification progress finished, it was time to evaluate the performance of the classifier according to evaluation criteria such as accuracy, sensitivity, and specificity which are used in this thesis. These evaluation criteria are used in order to compute the statistical achievements of training, testing, and validation parts of the system in classification problems. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) terms have to be expressed in order to explain evaluation criteria. True Positive (TP) is the number of instances that were classified correctly for the given class, True Negative (TN) is the number of inputs which does not belong to the given class and classified correctly. Likewise, False Positive (FP) is the number of instances that are classified as belonging to the given class incorrectly and False Negative (FN) is the number of instances classified incorrectly as not a member of the given class.

Accuracy term is a representation that the measurement of performance with regard to the classifier in order to predict correctly and known also the closeness of the measurements to the actual value. In particular, accuracy can be defined as the summation of true positive and true negative divided by the summation of entire instances in data set with multiplication hundred in order to obtain percentage rate, also represented in Equation (26),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\ (\%), \tag{26}$$

Sensitivity term is a representation of the true positive rate which defines the number of samples that are labeled correctly for the related cluster. In particular, sensitivity can be defined as a true positive rate divided by the summation of true positive and false negative rates with multiplication hundred in order to obtain a percentage rate. According to Equation (27), sensitivity term can be expressed as,

$$Sensitivity = \frac{TP}{TP + FN} * 100\ (\%)\ , \tag{27}$$

Specificity term is a representation of true negative rate which defines the number of instances that do not belong to the related cluster and classified correctly. Especially, specificity can be expressed as true negative rate divided by the summation of true negative and false positive rates with multiplication hundred in order to obtain a percentage rate. According to Equation (28), sensitivity term can be expressed as,

$$Specificity = \frac{TN}{TN + FP} * 100\ (\%), \tag{28}$$

The confusion matrix is a visual representation of TP, TN, FP and FN according to predicted and actual output cluster. The confusion matrix describes the exact performance of the model. In other words, the confusion matrix is also known as the error matrix. In machine learning statistical classification problems, a confusion matrix is commonly used in order to analyze evaluation criteria that define the performance of the classifier. The misclassified instances can be seen more clearly from the confusion matrix.

As a result of this part of the study, accuracy, sensitivity and specificity rates can reach from the confusion matrix that is shown in Figure 4.13.

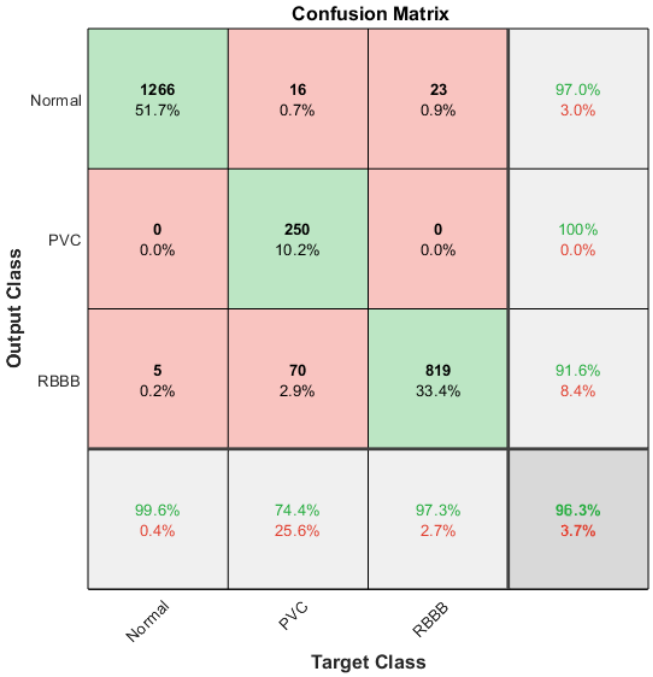

Figure 4.10. The Test Results of the Time-Series CNN Approach with Confusion Matrix

The last right column of the confusion matrix represents the specificity rates according to heartbeat classes as Normal, PVC, and RBBB. Also, the last row of the confusion matrix represents the sensitivity rates for each class. Consequently, the intersection of the last right column and the last row represents the accuracy rate which denotes the performance of the classifier for the related classification problem. In conclusion, the accomplishment rate of the test result is with an accuracy of 96.30%, also sensitivity and specificity rates can be seen from the confusion matrix for each class.

4.3.1.1. Results of ECG Time-Series Signal Classification with CNN

In order to show consistency between the confusion matrix and hand-calculated accomplishment rates by using Equations (26), (27), and (28) are shown in Table 4.4. This table represents the number of instances which are correctly classified and incorrectly classified according to the labels as same as the concept of the confusion matrix.

After computing accuracy, sensitivity, and specificity rates for each class, the average test results were computed for both three evaluation criteria.

Table 4.4. The Test Results of Time-Series Signal Classification with CNN (From Şen and Özkurt, 2019)

	ACTUAL					
PREDICTED	Normal	PVC	RBBB	Accuracy	Specificity	Sensitivity
Normal	**1266**	16	23	96,56%	96,48%	99,60%
PVC	0	**250**	0	96,44%	100%	74,40%
RBBB	5	70	**819**	95,97%	96,52%	97,26%
	Average Results			**96,30%**	**97,66%**	**90,43%**

At the end of the ECG Time-Series Signal Classification by using CNN study, the attained conclusions can be summarized as follows,

- The constructed CNN was a good classifier in accordance with ECG Time-Series Signal Classification.
- The constructed CNN has quite enough discriminative property when one-dimensional input vector is transformed into the four-dimensional array.
- The obtained accuracy, sensitivity and specificity rates were indicated that the classifier can be considered as a good classifier.
- Especially, in the PVC class, the specificity rate was quite high and the sensitivity rate was pretty low. It is desirable to have both sensitivity and specificity rates are high (Şen and Özkurt, 2019).

4.3.2. ECG Spectrogram Images Classification with CNN

The spectrogram images CNN approach was used in this study. The examination of spectrogram images of ECG signals was done in terms of time-frequency domain properties. As mentioned in the previous section, CNNs have great achievement rates on images in many classification problems in accordance with the biomedical subjects. In order to demonstrate the relation, spectrogram images were used in this part of the proposed study. The spectrogram images were obtained from ECG time-series signals by using the STFT technique with a particular window function.

Basically, three types of ECG spectrogram images were investigated in order to classify them by using CNN. Figure 4.14 represents the spectrogram image of normal sinus rhythm, Figure 4.15 indicates the spectrogram image of PVC arrhythmia, and Figure 4.16 denotes the spectrogram image of RBBB arrhythmia in the time-frequency domain.

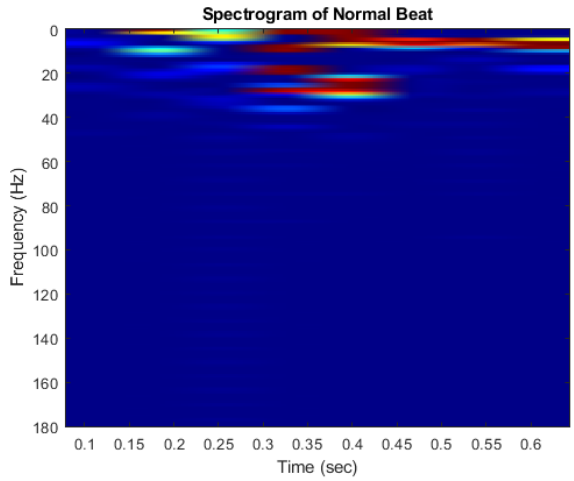

Figure 4.11. Representation of Single Normal Sinus Rhythm in Time-Frequency Domain

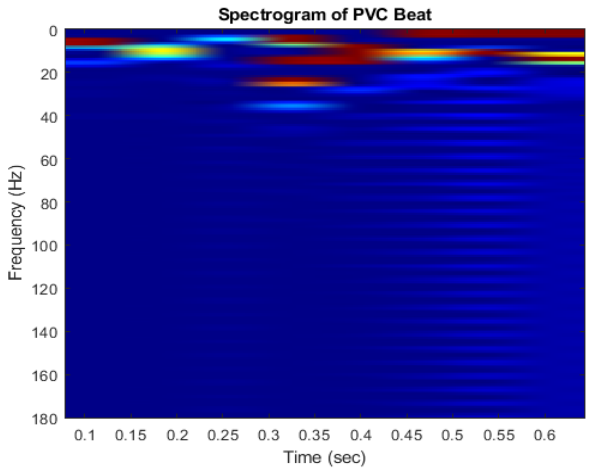

Figure 4.12. Representation of Single PVC Arrhythmia in Time-Frequency Domain

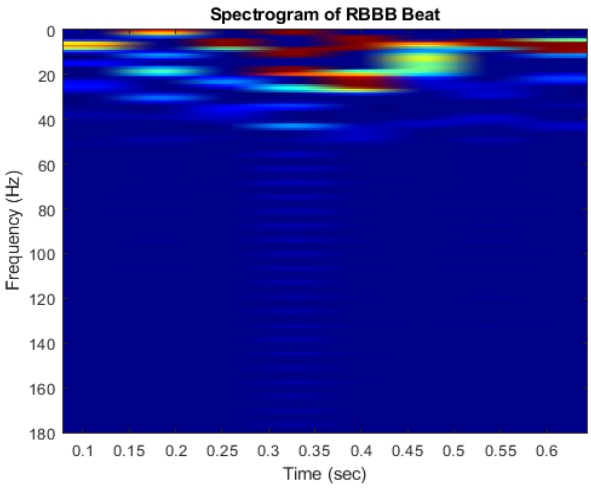

Figure 4.13. Representation of Single RBBB Arrhythmia in Time-Frequency Domain

Moreover, the spectrogram images were obtained from related to the MIT-BIH records and the number of spectrogram images with related classes are given in Table 4.5

Table 4.5. MIT-BIH Records and Beats for the Spectrogram Images CNN Approach

Record No	**Type of Beat**	**Number of Images**
112	Normal	2535
115	Normal	1952
122	Normal	2474
123	Normal	1514
106	PVC	520
119	PVC	444
200	PVC	826
203	PVC	444
118	RBBB	2165
124	RBBB	1530
207	RBBB	85
212	RBBB	1825

The three types of ECG signal shapes were obtained after applying the Pan Tompkins algorithm. The obtained single QRS complexes were transformed into spectrogram images according to labels and records numbers. The construction of each folder concerning their classes and the details of the construction of spectrogram images were given in Figure 4.17.

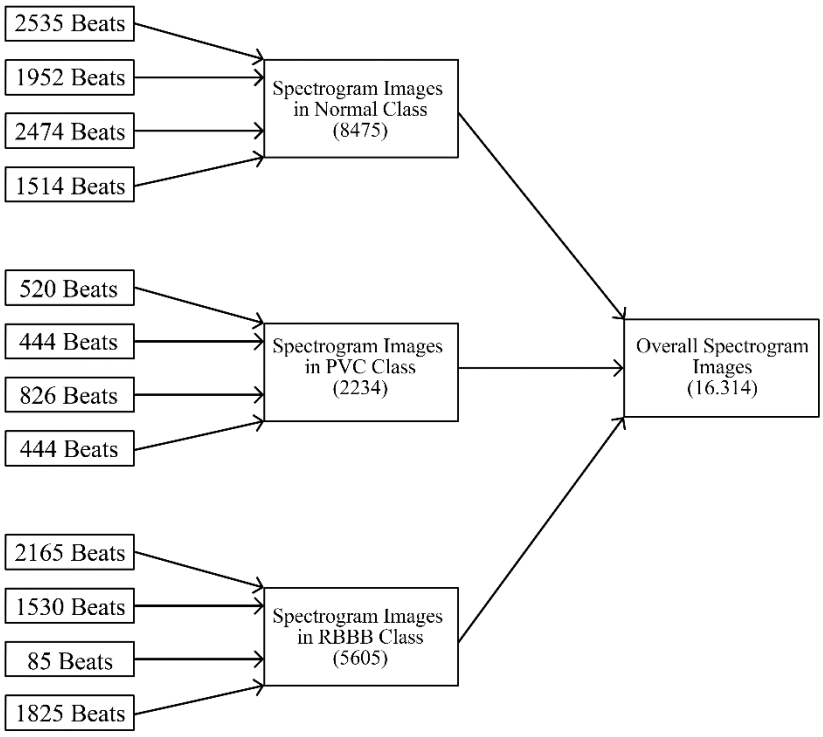

Figure 4.14. Representation of Spectrogram Images with Clusters and Numbers

The summary of the ECG spectrogram images classification process can be described in the following steps and shown in Figure 4.18.

1) The unprocessed data were taken from the MIT-BIH Arrhythmia Database. Entire data were utilized as single QRS complex form after stage 2. Then, the obtained QRS complexes were recorded as a .mat file.
2) The peak values and locations of the QRS complexes were detected with the aid of the Pan Tompkins algorithm for each type of heart rhythm such as Normal, PVC and RBBB. After the detection of R-peaks, QRS complexes were

defined by taking R-peak as a base, 129 samples were taken from the left side of the R-peak and 130 samples were taken from the right side of the R-peak. Hence, QRS complexes were defined with 260 samples long.

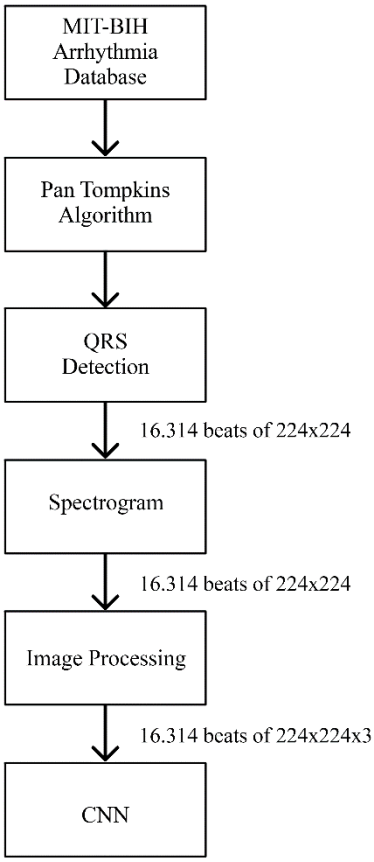

Figure 4.15. The Summary of Proposed Study in Time-Frequency Domain

3) The obtained QRS complexes which are represented in the time domain also were transformed into the time-frequency domain by using the STFT method. In the STFT method, the window function was selected as Hamming with a window size 256.
4) After signal processing techniques were done, the image processing technique was used for resizing to obtain 224x224x3 image input dimension, also jet map command was applied to acquire Red-Green-Blue (RGB) image.

5) The constructed overall data consist of a large volume of data as 16.314 spectrogram images. These constructed overall spectrograms include normal sinus rhythm and cardiac arrhythmias as an image form. Besides, 8475 spectrogram images were acquired from the Normal class, 2234 spectrogram images were acquired from the PVC class, and 5605 spectrogram images were acquired from RBBB class.
6) Consequently, these spectrogram images were given into CNN as input according to adjusted data split and model architecture.

When these stages are finished, overall spectrogram images were separated into training, testing and validation sets as 70%, 15% and 15%, consecutively and randomly. The summary of the study is indicated as a block diagram in Figure 4.19.

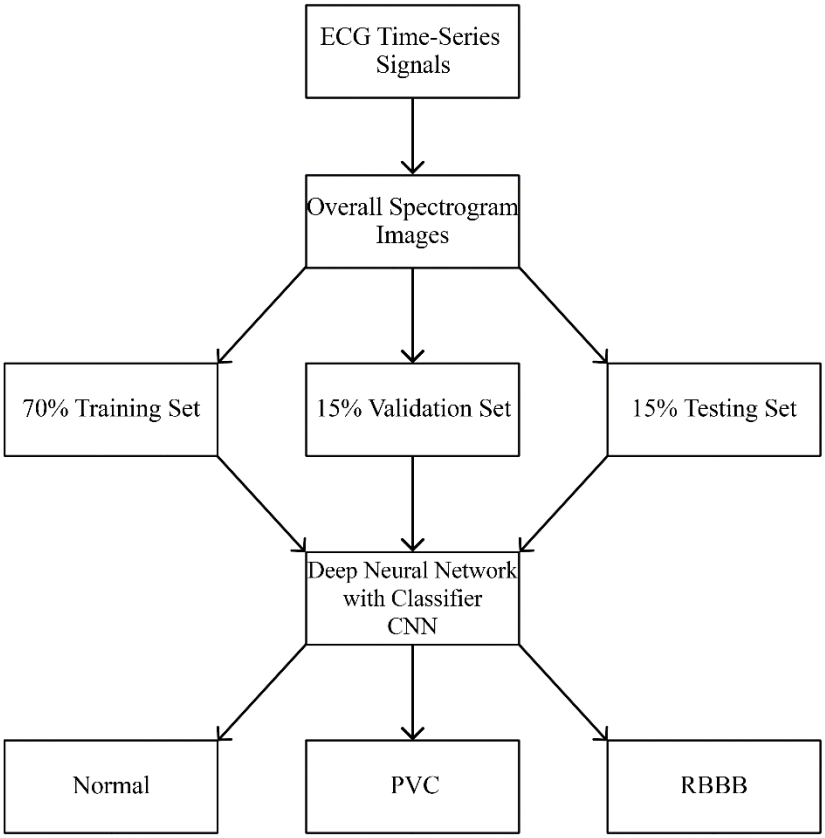

Figure 4.16. The Block Diagram of Time-Frequency Domain Classification

This block diagram explains the data split ratios with common percentages in the literature. According to this block diagram, time-series ECG signals were transformed into spectrogram images to analyze the time-frequency properties of these biomedical signals with the help of STFT. Then, the obtained spectrogram images were divided into training, testing and validation sets for this classification problem. The most well-known property of CNN is the discriminative features can be extracted from data without any traditional feature extraction methods. The distinguished properties were analyzed according to the time-frequency domain. The constructed deep neural network model is indicated in Table 4.6 according to layer types, kernel filter size, stride and layer outputs.

Table 4.6. The Proposed CNN Architecture for ECG Spectrogram Images Classification

Layers	Layer Types	Output of Layer	Filter Size	Stride
1	Image Input	224x224x3	-	-
2	Convolution	224x224x8	[3,3]	[1,1]
3	Batch Normalization	224x224x8	-	-
4	ReLU	224x224x8	-	-
5	Max Pooling	75x75x8	[1,1]	[3,3]
6	Convolution	75x75x16	[3,3]	[1,1]
7	Batch Normalization	75x75x16	-	-
8	ReLU	75x75x16	-	-
9	Max Pooling	25x25x16	[1,1]	[3,3]
10	Convolution	25x25x32	[3,3]	[1,1]
11	Batch Normalization	25x25x32	-	-
12	ReLU	25x25x32	-	-
13	Fully Connected	1x1x3	-	-
14	Softmax	1x1x3	-	-
15	Classification Output	3	-	-

The proposed CNN model was constructed according to model layers and specific training option. The initial learning rate is set to 10^{-5} and the validation frequency is set to 50. In other words, the data will be validated in every 50 samples. This progress is completed in approximately 12 minutes, it is the elapsed time of the network. The training process is completed in 10 epochs and 89 iterations per epoch. It defines the maximum iteration of the network. Epoch term can be seen as a complete pass across all data set. Furthermore, the iteration term can be defined as an estimation of the gradients and updating of the network parameters in each iteration. Data were shuffled in every epoch. This explains the shuffling of the train data before starting in each epoch and shuffling the validation data before starting in each epoch to apply the network validation process. An iteration is one step taken in the gradient descent algorithm towards minimizing the loss function. The solver for the training network is set as SGDM optimizer and the momentum term is set to 0.9.

As a result, the performance of the classifier was scrutinized according to evaluation criteria such as accuracy, sensitivity, and specificity as same as the ECG time-series signal classification approach. These evaluation criteria are used to figure out statistical achievements of training, testing, and validation parts of the system in classification problems. So as to see the accomplishment rates of the test results, the confusion matrix provides a beneficial visualization. The number of instances that are correctly classified or misclassified can be seen from the confusion matrix.

As a result of this part of the study, accuracy, sensitivity and specificity rates can be seen from the confusion matrix that is indicated in Figure 4.20.

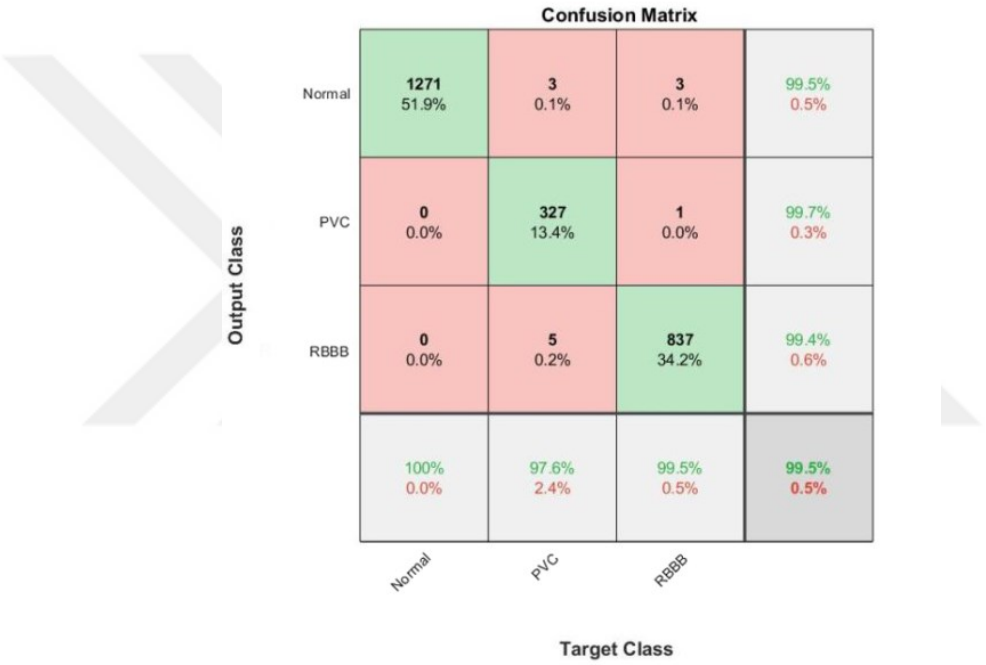

Figure 4.17. The Test Results of the Spectrogram Images CNN Approach with Confusion Matrix

The last right column of the confusion matrix represents the specificity rates according to heartbeat classes as Normal, PVC and RBBB. Also, the last row of the confusion matrix represents the sensitivity rates according to classes. Consequently, the intersection of the last right column and the last row represents the accuracy rate which specifies the performance of the classifier. In conclusion, the accomplishment rate of the test result is with an accuracy of 99.50%, also sensitivity and specificity rates can be seen from the confusion matrix for each class.

4.3.2.1. Results of ECG Spectrogram Images Classification with CNN

The steadiness between achievement rates of the confusion matrix and hand-calculated accomplishment rates by using Equations (26), (27), and (28) can be seen in Table 4.7. The table includes the test result of the classification problem according to the number of instances which were correctly classified or misclassified for each class.
After computing accuracy, sensitivity, and specificity rates according to each class, then average test results were shown for both three evaluation criteria.

Table 4.7. The Test Results of ECG Spectrogram Images Classification with CNN (From Şen and Özkurt, 2019)

	ACTUAL					
PREDICTED	Normal	PVC	RBBB	Accuracy	Specificity	Sensitivity
Normal	**1271**	3	3	99,75%	99,48%	100%
PVC	0	**327**	1	99,63%	99,95%	97,61%
RBBB	0	5	**837**	99,63%	99,68%	99,63%
	Average Results			**99,67%**	**99,70%**	**99,08%**

At the end of ECG Spectrogram Images Classification by using CNN, the reached conclusions can be summarized as follows,

- The constructed CNN provided high accomplishment rates on ECG spectrogram image classification.
- The number of instances that are correctly classified was increased with time-frequency domain analysis.
- Especially, in the Normal class, the sensitivity rate was also quite high, which explains the TP rate increased.
- The sensitivity rate for the PVC class increased by 23% via using the spectrogram method which depends on the STFT approach (Şen and Özkurt, 2019).

4.4. Discussion

The classification of ECG signal types of Normal, PVC, and RBBB was actualized by using two different approaches. Two different approaches are used in this study; the first approach was to classify time-series ECG signals in time, and the second approach was to classify ECG spectrogram images which are obtained after applying STFT. The details of both two approaches were explained in previous sections. On the other hand, these two different types of data were trained with the same model architecture and the same model parameters via using CNN. It was observed that the classification of ECG signals with the spectrogram-based method had more success rates than the classification of ECG time-series signals.

Consequently, the overall accomplishment rates are shown in Table 4.8. According to Table 4.8, accuracy and specificity rates are high, but the sensitivity rate of ECG time-series signal classification is low compared to ECG spectrogram image classification. Generally, it can be said that the time-frequency based classification approach was outperformed the time domain-based classification approach.

Table 4.8. The Average Results of Classification

Method	Accuracy	Specificity	Sensitivity
ECG Time-Series Signal Classification	96.30%	97.66%	90.43%
ECG Spectrogram Images Classification	99.67%	99.70%	99.08%

Table 4.9 shows the test results of all rhythms that were used in this study. Moreover, the accomplishment rates for each class can be seen clearly. The most remarkable change happened in the PVC class. The sensitivity rate of the PVC beat class increased by 23% with the aid of the spectrogram technique which is based on time-frequency domain analysis (Şen et al., 2019).

Table 4.9. The Test Results of Both Approaches for Classification with CNN

Evaluation Criteria	**ECG Time-Series Signal Classification**			**ECG Spectrogram Images Classification**		
	N	**PVC**	**RBBB**	**N**	**PVC**	**RBBB**
Accuracy	96.56%	96.44%	95.97%	99.75%	99.63%	99.63%
Sensitivity	99.60%	74.40%	97.30%	100%	97.61%	99.50%
Specificity	97%	100%	91.60%	99.50%	99.70%	99.40%

CHAPTER 5
HYPERPARAMETER TUNING WITH ADAM OPTIMIZER

In deep learning problems, various optimization techniques have been used in the literature. The most well-known optimization algorithms were examined in Chapter 3. SGD-based algorithms protect single learning rate parameter in order to update weights, and the learning rate parameter does not vary in the training process. Adam algorithm can be seen as an extended interpretation of SG-based optimization algorithms by taking its advantages. By looking at this aspect, the various algorithms are derived from the SGD algorithm such as Adam and RMS Prop. These advantageous properties are the improvements with sparse gradients and can cope with the non-stationary signals by considering performance. ECG signal's characteristics are known as non-stationary. Therefore, the most important point of the Adam algorithm is compatible with the non-stationary ECG signals.

Basically, the Adam algorithm was selected as an optimizer of the network during the training progress in this proposed study. The details of the Adam algorithm were introduced in Chapter 3. In this chapter, the results of the adjusted hyperparameters of the Adam algorithm will be given in detail.

5.1. Data Set

In this proposed study, ECG signals were taken from the MIT-BIH Arrhythmia Database. The details of the MIT-BIH Arrhythmia Database and its corresponding records were explained in Chapter 4.

In this chapter, the number of records and beats were selected differently from the proposed study mentioned in Chapter 4. Table 5.1 indicates that all ECG signals were used according to record numbers, the number of beats, and patients' age and gender information in this proposed study.

Table 5.1. MIT-BIH Arrhythmia Database Record Details

Record No	Gender	Age	Beat Type	Number of Beats
100	Male	69	Normal	2237
109	Male	64	LBBB	2490
118	Male	69	RBBB	2165

The Pan Tompkins algorithm as indicated in Chapter 4 and applied to these taken records according to the mentioned consecutive steps. After the same procedure was applied to ECG time-domain signals, single QRS complexes were obtained. These obtained QRS complexes are shown in Figure 5.1, Figure 5.2, and Figure 5.3 according to corresponding cardiac arrhythmias in one-second duration.

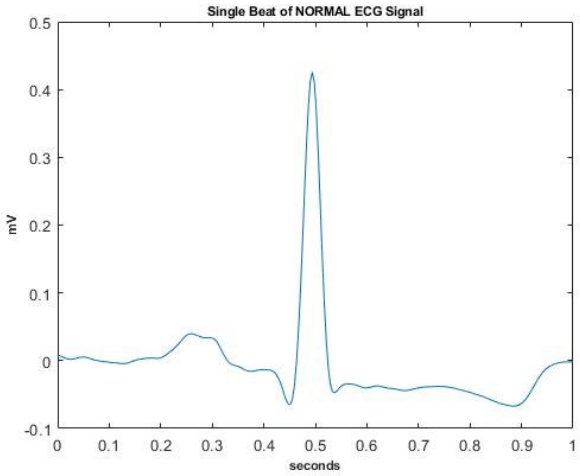

Figure 5.1. Single Normal Sinus Rhythm Time Domain Representation

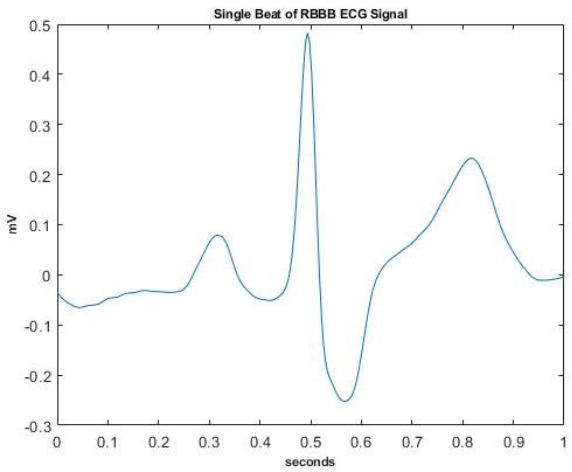

Figure 5.2. Single RBBB Arrhythmia Time Domain Representation

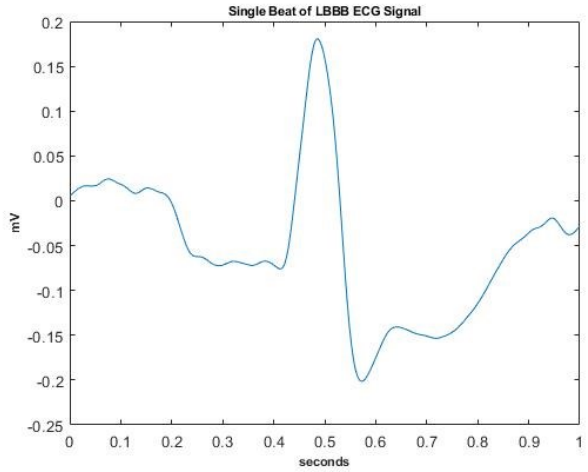

Figure 5.3. Single LBBB Arrhythmia Time Domain Representation

The spectrogram method was applied to these obtained QRS complexes in order to interpret in the time-frequency domain. According to the spectrogram approach, ECG signals can analyze both time and frequency components together. In this study, a particular time-frequency resolution rate was selected to enhance spectrograms. After heartbeats were transformed into a time-frequency domain, spectrograms were bounded by a specific time-frequency resolution rate. In order to search meaningful components of time-series ECG signals, single-sided amplitude spectrum analysis was applied. Figures 5.4, 5.5, and 5.6 represent a single-sided amplitude spectrum of normal sinus rhythm, RBBB arrhythmia, and LBBB arrhythmia, respectively.

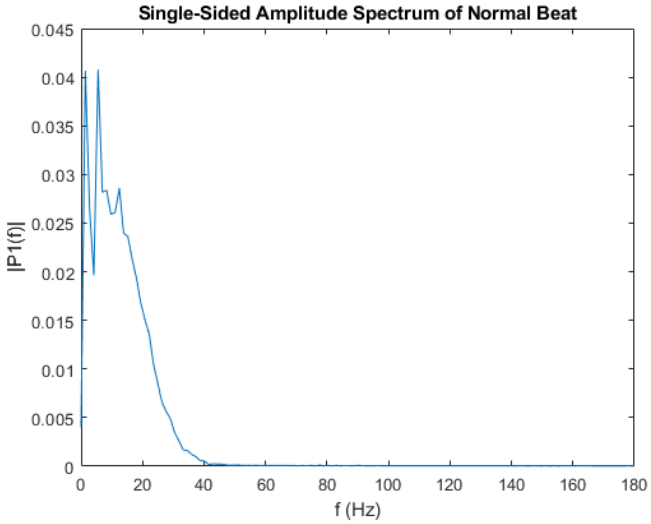

Figure 5.4. Single-Sided Amplitude Spectrum Representation of Normal Beat

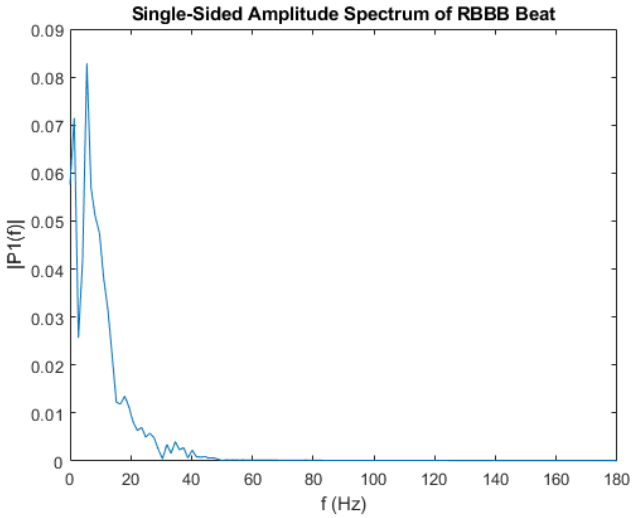

Figure 5.5. Single-Sided Amplitude Spectrum Representation of RBBB Beat

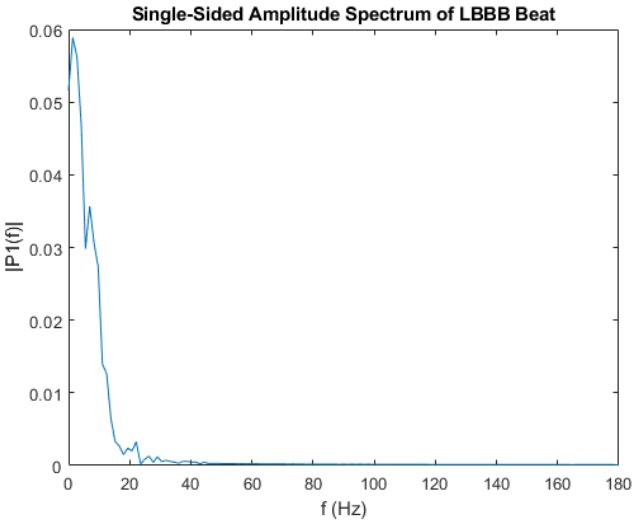

Figure 5.6. Single-Sided Amplitude Spectrum Representation of LBBB Beat

The meaningful frequency components of the signals were extracted by analyzing the single-sided amplitude spectrum. According to these representations, the common frequency value was selected as 40 Hz since all meaningful frequency components of the signals were observed under 40 Hz. It implies that the frequency resolution rate was bounded by 40 Hz. In addition, the time resolution rate was selected 200 ms in a heuristic way. The best suitable time resolution rate was 200 ms. These resolution rates were selected considering no distortion effect on the spectrograms. The bounded time-frequency resolution rate of spectrograms is shown in Figures 5.7, 5.8, and 5.9 with respect to cardiac rhythms such as Normal, RBBB, and LBBB, successively.

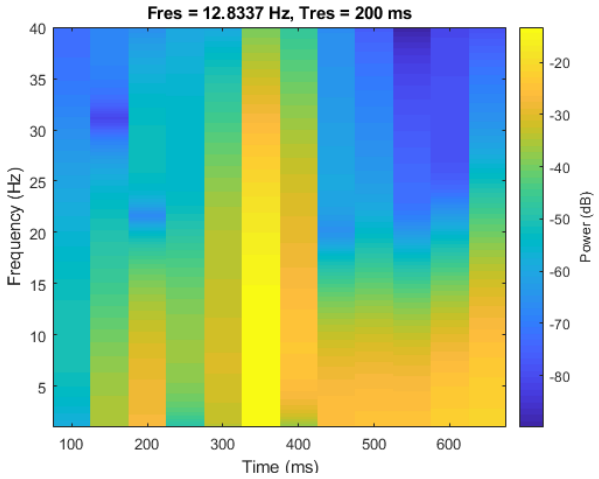

Figure 5.7. Spectrogram of Normal Beat

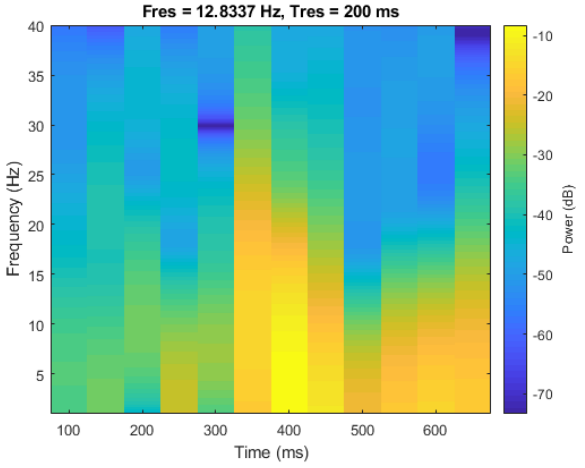

Figure 5.8. Spectrogram of RBBB Beat

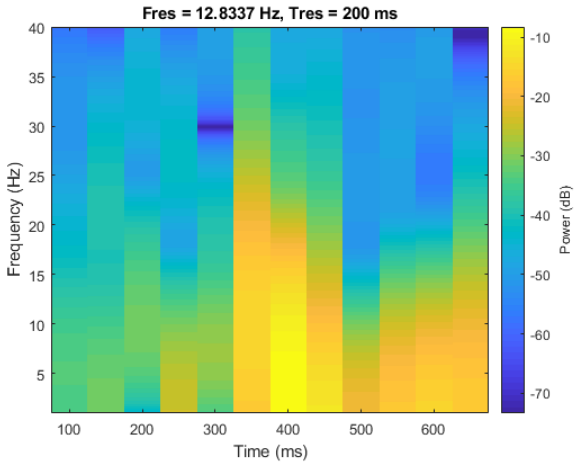

Figure 5.9. Spectrogram of LBBB Beat

The obtained spectrograms were collected in one mutual folder in order to split data as training, validation, and test parts of the study. The total number of spectrograms for each cluster and data split ratios are indicated in Figure 5.10.

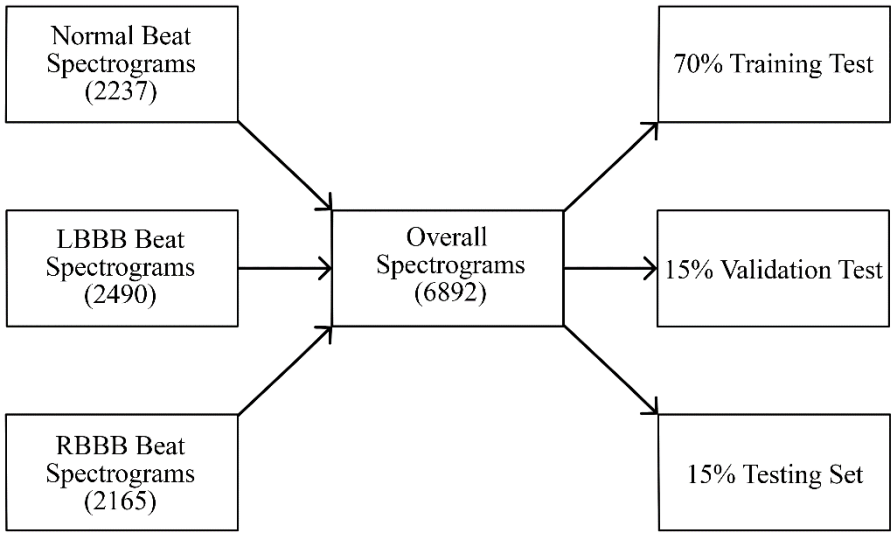

Figure 5.10. Representation of Overall Spectrograms with Split Ratios

After spectrograms were obtained, the entire data divided into training, testing, and validation parts as 70%, 15%, and 15% randomly and respectively. Then, the deep neural network was constructed as a classifier. The deep neural network layers and their functions were explained in Chapter 3. Table 5.2 indicates that the proposed CNN model according to layer information and filter/stride sizes.

Table 5.2. The Proposed CNN Architecture for Spectrogram Classification

Layers	Layer Type	Output of Layer	Filter Size	Stride
1	Image Input	536x607x3	-	-
2	Convolution	534x605x8	[3,3]	[1,1]
3	ReLU	534x605x8	-	-
4	Max Pooling	267x302x8	[2,2]	[2,2]
5	Convolution	265x300x16	[3,3]	[1,1]
6	ReLU	265x300x16	-	-
7	Max Pooling	132x150x16	[2,2]	[2,2]
8	Convolution	130x148x32	[3,3]	[1,1]
9	ReLU	130x148x32	-	-
10	Fully Connected	1x1x3	-	-
11	Softmax	1x1x3	-	-
12	Classification Output	3	-	-

The proposed CNN model was constructed according to given layers and filter/stride sizes. Before starting training progress, weights and biases were initialized in order to obtain better results and to make an easier comparison of the results individually. The training progress was completed approximately 9 hours for eighteen different results. The network was trained for 5 epochs. The hyperparameters were adjusted according to the grid search approach. Basically, the grid search approach tries to find optimal parameters of the model by scanning data; however, the grid search technique takes a long time in computing progress. Therefore, the grid search approach is known as computationally expensive. In the grid search, it iterates every combination of defined parameters and trying them to find the optimal for the model. In training and validation parts, data were shuffled per-epochs. In this proposed study, the adjusted hyperparameters were initial learning rate or step size, gradient decay factor of Adam algorithm which is also known as the first-order moment, and squared gradient decay factor of Adam algorithm which is also defined as the second-order moment. The epsilon value was selected as 10^{-8} in order to prevent division by zero, it was given in the optimization algorithms section as a subtitle of Chapter 3.

The summary of the proposed study can be explained in the following stages, and also shown in Figure 5.11.

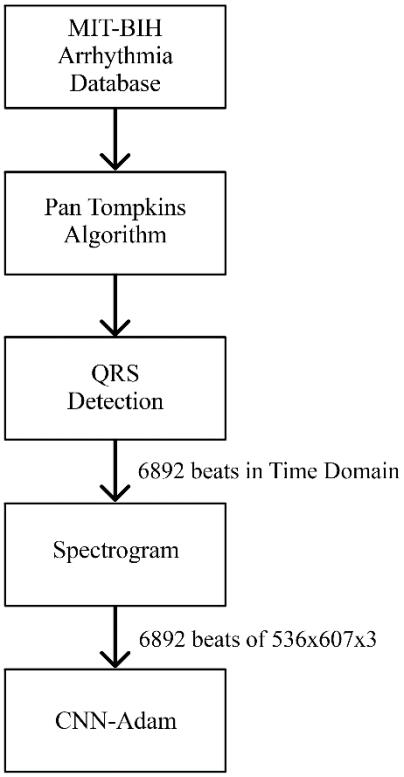

Figure 5.11. Representation of the Summary of the Proposed Study

1) The unprocessed data were taken from the MIT-BIH Arrhythmia Database.
2) These taken ECG signals were processed by using the Pan Tompkins algorithm. The peak values and locations of R-peaks were detected. Therefore, QRS complexes were extracted according to related cardiac rhythms that were used in the proposed study.
3) The obtained QRS complexes were transformed into spectrogram by using STFT with a bounded time-frequency resolution rate.
4) The spectrograms were collected for each class as Normal, RBBB, and LBBB. The total number of instances were indicated in Figure 5.10. The image input dimension was 536x607x3.
5) The total number of instances were divided into training, testing, and validation parts of the study with particular split ratios represented in Figure 5.10.
6) The weights and biases were initialized before starting the training progress of the system.

7) Then, the proposed CNN model was constructed according to specified layer types, filters, and stride sizes.
8) The combination of hyperparameters which are the learning rate, the exponential decay rates for the first moment estimates, and second-moment estimates are represented in six different cases according to grid search. The initial learning rate was evaluated for $10^{-4}, 10^{-3}, 10^{-2}$ values. The first moment estimates (β_1) was evaluated for $0\ and\ 0.9$, also second-moment estimates (β_2) was evaluated for $0.99, 0.999, 0.9999$.
9) In conclusion, the CNN was selected as a classifier according to the adjusted hyperparameters. The results of the study were examined with respect to validation losses for each combination.

5.2. Results of CNN Hyperparameter Tuning with Adam Optimizer

The main concept of the proposed study, observation of the effect of first and second-moment estimations and learning rate in the constructed deep neural network. In order to see the effect of adjusted hyperparameters, the results were evaluated according to validation losses for different learning rates while changing first and second-moment estimations. According to the grid search approach, the initial learning rate is adjusted first, then the gradient decay factor and squared gradient decay factor were adjusted as a combination in itself. The obtained results were evaluated under six cases according to a combination of first-moment estimation and second-moment estimation.

- **Case 1:**

In this case, the learning rate parameter selected as $\alpha = 0.0001$, first-moment estimation selected as $\beta_1 = 0$ and second-moment estimation selected as $\beta_2 = 0.99$ for Option 1, $\beta_2 = 0.999$ for Option 2, $\beta_2 = 0.9999$ for Option 3. In Figure 5.12, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, consecutively.

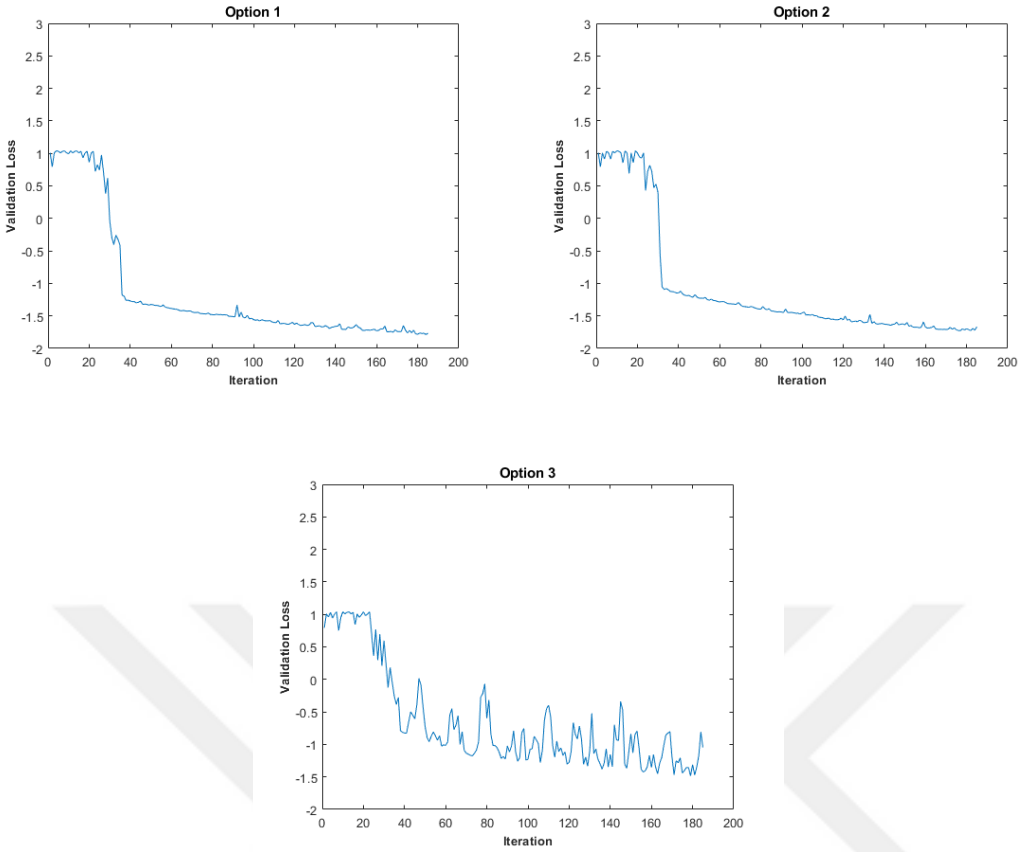

Figure 5.12. Representation of Validation Losses for Case 1

By looking at these three validation loss graphs, the Adam algorithm tried to find the global minimum point but had difficulty because the selected learning rate was too small for the network.

- **Case 2:**

  In this case, the learning rate parameter selected as $\alpha = 0.0001$, gradient decay factor selected as $\beta_1 = 0.9$ and squared gradient decay factor selected as $\beta_2 = 0.99$ for Option 4, $\beta_2 = 0.999$ for Option 5, $\beta_2 = 0.9999$ for Option 6. In Figure 5.13, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, respectively.

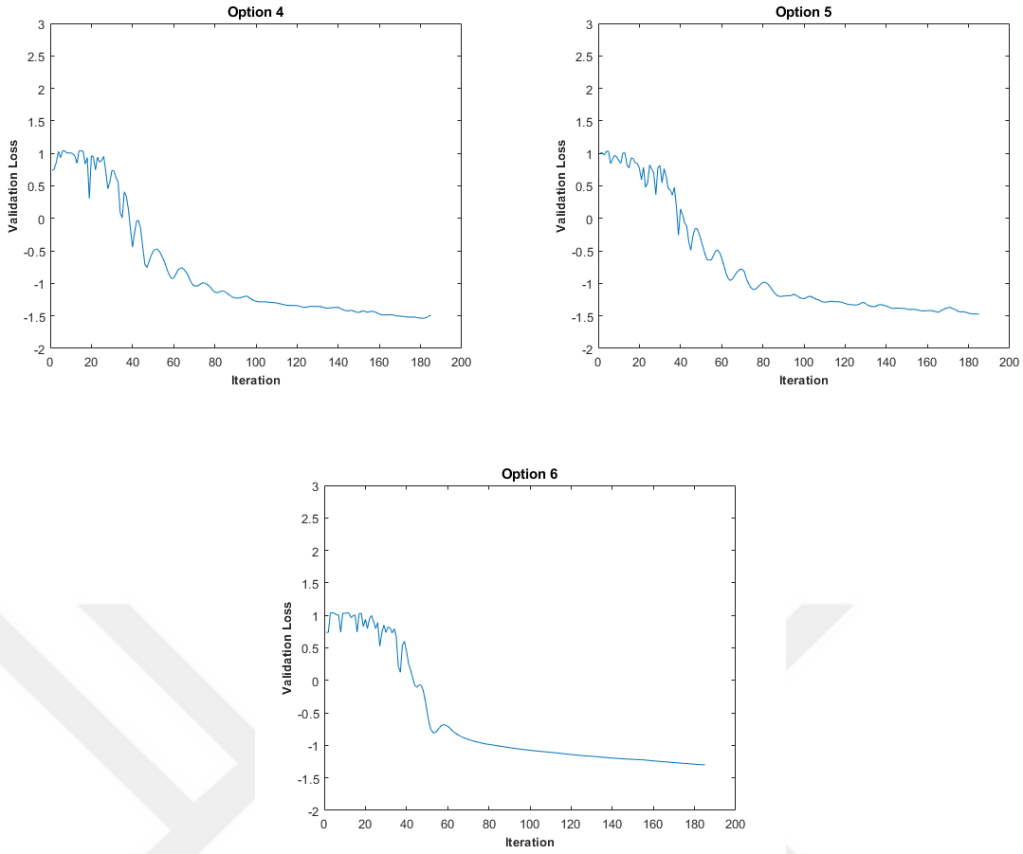

Figure 5.13. Representation of Validation Losses for Case 2

The learning rate parameter was adjusted as a small rate, but it is obvious that the selection of squared gradient decay factor ($\beta_{2)}$ bigger was provided more stable convergence compared with the three validation loss graphs in Case 1.

- **Case 3:**

In this case, the learning rate parameter selected as $\alpha = 0.001$, gradient decay factor selected as $\beta_1 = 0$ and squared gradient decay factor selected as $\beta_2 = 0.99$ for Option 7, $\beta_2 = 0.999$ for Option 8, $\beta_2 = 0.9999$ for Option 9. In Figure 5.14, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, successively.

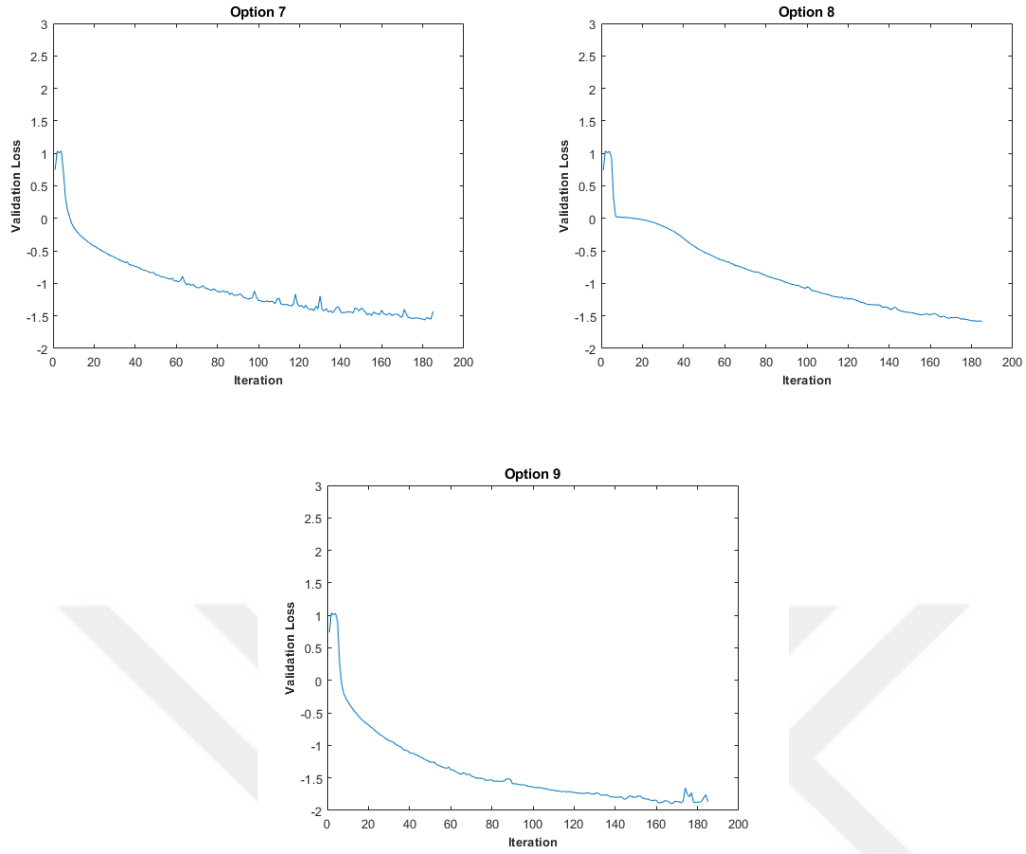

Figure 5.14. Representation of Validation Losses for Case 3

It was obvious that the increase in squared gradient decay factor (β_2) provided a more stable convergence. It is apparent the increase of β_2 provided more stable convergence.

- **Case 4:**

In this case, the learning rate parameter selected as $\alpha = 0.001$, first-moment estimation selected as $\beta_1 = 0.9$ and second-moment estimation selected as $\beta_2 = 0.99$ for Option 10, $\beta_2 = 0.999$ for Option 11, $\beta_2 = 0.9999$ for Option 12. In Figure 5.15, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, consecutively.

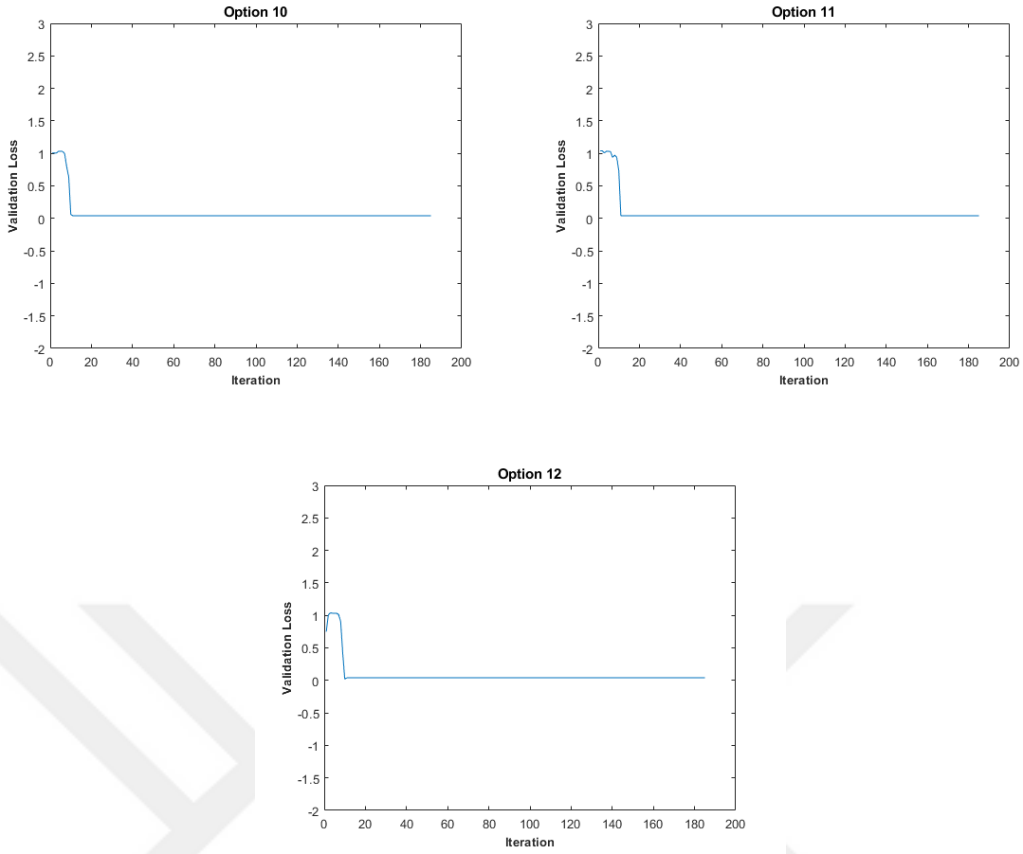

Figure 5.15. Representation of Validation Losses for Case 4

The initial learning rate, gradient decay factor, and squared gradient decay factor which are adjusted hyperparameters were not suitable for the network. All three options indicated that this network model can be seen as a random classifier because the adjusted hyperparameters were too poor.

- **Case 5:**

  In this case, the learning rate parameter selected as $\alpha = 0.01$, gradient decay factor selected as $\beta_1 = 0$ and squared gradient decay factor selected as $\beta_2 = 0.99$ for Option 13, $\beta_2 = 0.999$ for Option 14, $\beta_2 = 0.9999$ for Option 15. In Figure 5.16, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, respectively.

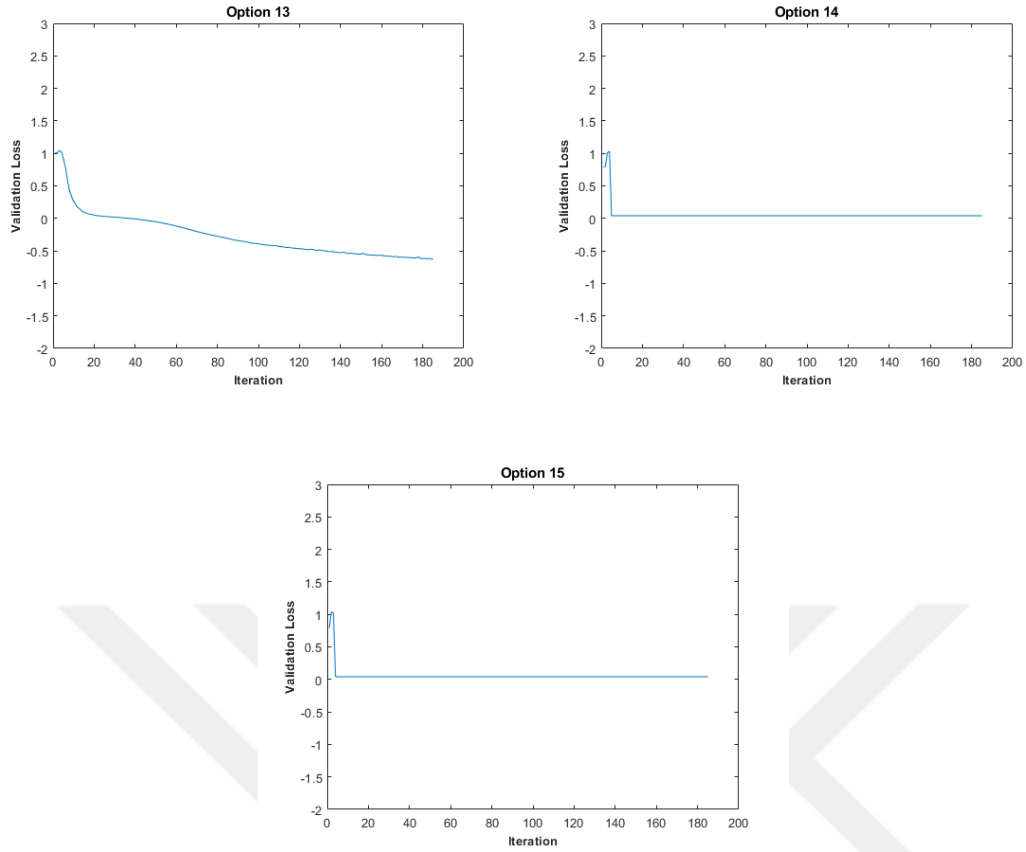

Figure 5.16. Representation of Validation Losses for Case 5

By looking at these three options, each three hyperparameter combinations were not appropriate for the network. According to the results of validation losses, the networks were not distinguished from any random classifier. Moreover, these networks had lost their generalization ability with respect to the adjusted hyperparameters.

- **Case 6:**

In this case, the learning rate parameter selected as $\alpha = 0.01$, gradient decay factor selected as $\beta_1 = 0.9$ and squared gradient decay factor selected as $\beta_2 = 0.99$ for Option 16, $\beta_2 = 0.999$ for Option 17, $\beta_2 = 0.9999$ for Option 18. In Figure 5.17, the results of adjusted hyperparameters are indicated according to validation loss graphs for every three options, successively.

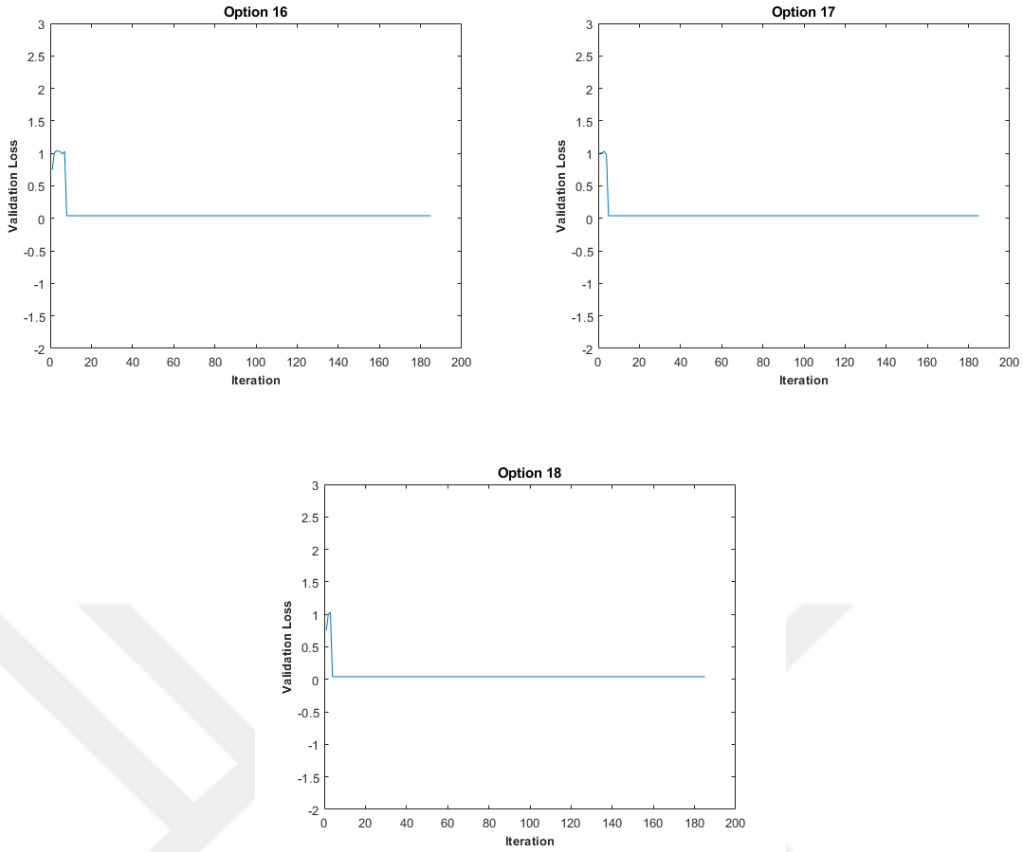

Figure 5.17. Representation of Validation Losses for Case 6

It was obvious that the learning rate parameter was dominating other hyperparameters of the network. Similarly, it occurred because of the learning rate parameter domination in Case 5. According to these three options, it can be said that all three different networks had lost the generalization ability. Therefore, the networks could not reach the global optimum point according to selected hyperparameters (Şen and Özkurt, 2020). Moreover, the validation loss graphs were represented as taking their logarithms in order to see the changes detailly.

Additionally, the performance of the study was evaluated according to the confusion matrix. The confusion matrix represents the accuracy, sensitivity, and specificity rates according to the best hyperparameter selection by considering the test results of the model.

This visual representation makes it easier to understand and analyze the TP, TN, FP, and FN rates also provide us to see the correctly classified instances and misclassified instances according to their labels. Figure 5.18 indicated that the test results according to the best hyperparameter selection.

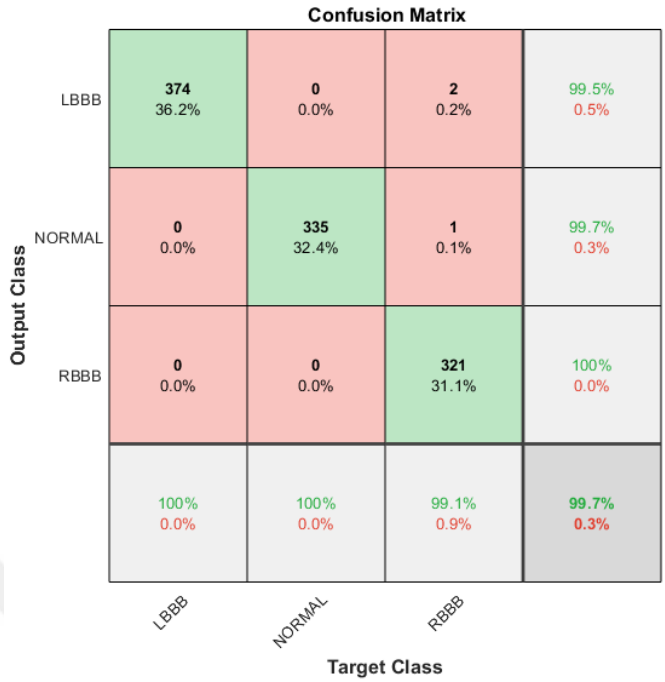

Figure 5.18. The Confusion Matrix for The Test Results of the Best Hyperparameters

The confusion matrix represents the test results of the hyperparameters which is the best validation accuracy/loss results. According to the confusion matrix, the success rate of the network is 99.70%. It is obvious that the test results are highly promising due to just three instances were misclassified. Consequently, the constructed deep neural network is shown a great classification success rate according to the tuned hyperparameters with the Adam optimization algorithm.

5.3. Discussion

When compared to obtained validation losses, the best-tuned hyperparameter selection was Option 9. According to the results of Option 9, training accuracy was 100% and validation accuracy was obtained as 99.80%.

In Option 8, validation loss was a little bit lower than Option 9. But here, there is an important point that has to be evaluated to choose the best solution. In Option 8, there is a tiny difference between training accuracy and validation accuracy. It arises from

some well-known reasons. The first reason, the training loss was measured during each epoch whereas validation loss was measured after each epoch. The second reason, the validation set might be simpler than the training set. Therefore, Option 9 was selected as the best solution for the classification problem by considering these reasons.
All options are underfitted except Options 4, 7 and 9. Therefore, these networks need more training time according to adjusted hyperparameters, and all results are indicated in Table 5.3.

Table 5.3. The Overall Results for Each Hyperparameter Selection (From Şen, 2020)

Option	Base Learn Rate	β_1	β_2	Training Accuracy	Validation Accuracy	Validation Loss
1	0.0001	0	0.99	96.87	99.12	0.0401
2	0.0001	0	0.999	98.43	99.12	0.0294
3	0.0001	0	0.9999	98.43	98.45	0.0390
4	0.0001	0.9	0.99	100	99.32	0.0260
5	0.0001	0.9	0.999	96.87	99.80	0.0227
6	0.0001	0.9	0.9999	97.65	99.32	0.0323
7	0.001	0	0.99	99.21	99.12	0.0393
8	0.001	0	0.999	99.21	99.61	0.0149
9	**0.001**	**0**	**0.9999**	**100**	**99.80**	**0.0168**
10	0.001	0.9	0.99	35.15	36.20	1.0973
11	0.001	0.9	0.999	34.37	36.20	1.0977
12	0.001	0.9	0.9999	46.09	36.20	1.0978
13	0.01	0	0.99	36.71	36.20	1.0967
14	0.01	0	0.999	80.46	84.70	0.4163
15	0.01	0	0.9999	37.50	36.20	1.0967
16	0.01	0.9	0.99	33.59	36.20	1.0967
17	0.01	0.9	0.999	47.65	36.20	1.0967
18	0.01	0.9	0.9999	44.53	36.20	1.0967

According to the adjusted hyperparameters of Option 9 such as gradient decay factor and squared gradient decay factor were not the same as the default values in the literature study. Generally, β_1 is set to 0.9 and β_2 is set to 0.999 in the literature. In conclusion, by looking at the experimental results of the proposed study, the default values of β_1 and β_2 were not suitable for the proposed deep neural network model and data. This showed the dependency of the performance on the data.

In Figure 5.9, the validation losses were represented in logarithmic scale for each option, and validation losses are calculated according to Equation (13). It can be seen from here; Option 9 was the best result with respect to the validation loss criteria which is corresponding Case 3.

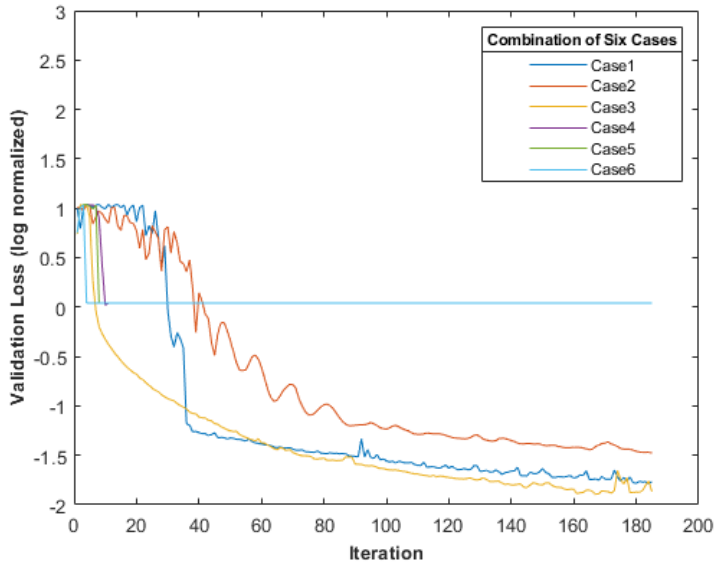

Figure 5.19. The Representation of Validation Losses for Each Case

The test results of the proposed study according to the best hyperparameter selection are indicated in Table 5.4.

Table 5.4. The Test Results of the Best Hyperparameter Selection (Şen, 2020)

	ACTUAL					
PREDICTED	LBBB	Normal	RBBB	Accuracy	Specificity	Sensitivity
LBBB	**374**	0	2	99,70%	99,50%	100%
Normal	0	**335**	1	99,80%	99,70%	100%
RBBB	0	0	**321**	99,70%	100%	99,10%
	Average Results			**99,73%**	**99,73%**	**99,70%**

On the other hand, the test results showed high success rates for each cluster in the classification problem. The most notable point of the test results, sensitivity rate which represents the TP term was obtained as 100% for Normal and LBBB classes. On the other hand, the specificity rate which represents the TN term was obtained as 100% for RBBB class.

CHAPTER 6
CONCLUSIONS AND FUTURE STUDIES

The examination of ECG signals is still a compelling branch in signal processing approaches. Early detection and diagnosis have a great significance in preventing heart diseases. Because heart disorders are restrained people from maintaining a healthy life. One of the compelling points is that the detection and improvement of robust classification methods for cardiac arrhythmias. Many researchers have been figured out the different classification methods to hindering people from probable cardiac disorders. In the analysis of biomedical signals, different signal processing approaches have been used by many investigators such as FT, FFT, spectrogram, or a wavelet transform. Because of the good representation abilities and simplicity spectrogram method were used in this study.

In this thesis, the detection and classification of heart arrhythmias were carried out by extraction of QRS complexes. All of these QRS complexes were extracted with the aid of the Pan Tompkins algorithm that ubiquitous QRS-detection algorithm in the literature studies. The ECG signals were obtained from MIT-BIH Arrhythmia Database via using the STFT approach. Four types of heartbeats are Normal, PVC, LBBB, and RBBB were used for the classification. Pan Tompkins algorithm was utilized to detect and extract of QRS complexes in the data preparation stage.

Two main studies were performed according to the time-frequency-based classification approaches in this thesis.

- ***Time-Series vs. Spectrogram-CNN:*** The first study was aimed to compare the performance of both time and time-frequency domains ECG signal characteristics to classify normal sinus rhythm and cardiac abnormalities.
- ***Spectrogram-CNN-Hyperparameter Tuning with Adam Optimization Algorithm:*** The second study was related to the time-frequency approach for categorizing ECG signals according to their labels and was the aim to adjust hyperparameters of the selected optimization algorithm.

Many classification approaches are based on artificial neural network technique, but CNNs have gained significance with development in deep learning in recent years. For this reason, CNN was selected as a classifier for both of the studies with the help of MATLAB.

In the Time-Series vs. Spectrogram-CNN study, two different methods were used for the categorization of heart rhythms such as Normal, PVC, and RBBB. The objective of the study was to compare the test results which are both time and time-frequency domains. The SGDM was selected as an optimizer for both two parts, also data were separated with a rate of 70% training data, 15% testing data, and 15% validation data. The first method was demonstrated as one dimensional ECG time-series signal classification via using CNN. The accuracy accomplishment rate was 96.30%, also other evaluation criteria gave promising results such as sensitivity was 90.43% and specificity was 97.66%. The sensitivity and specificity test results were quite high in both Normal and RBBB clusters, unlike the PVC class. The second method depends on image classification with the help of CNN. In the beginning, the extracted QRS complexes were transformed into the time-frequency domain via using STFT. After applied Hamming window operation, all of these QRS complexes were resized in RGB image form. The most well-known trait of CNN is successful for image classification. According to test results, the accuracy accomplishment rate was 99.67%. It exhibited highly robust and thriving classification performance with a selected optimization algorithm. The second method which was time-frequency based properties of the data outperformed the first method by comparing achievement rates. Additionally, the sensitivity rate was increased by 23% in the PVC cluster according to the second method. Consequently, the different evaluation results proved that the possession of both time and frequency features was the crucial point in CNN-based classification problems.

In the Spectrogram-CNN-Hyperparameter Tuning with Adam Optimization Algorithm study, the main objective was to classify ECG heartbeats according to tuned hyperparameters with Adam optimizer. The ECG time-series heart signals were transformed into the time-frequency domain via using STFT. Then, the time-frequency resolution rate is bounded by 40 Hz in spectrograms. The first step was to analyze and determine the meaningful components of the ECG heart signals. After the analysis of ECG signals, the meaningful components were detected. These obtained spectrograms were used as input for the CNN. The data which are spectrograms were split with a

rate of 70% training data, 15% testing data and 15% validation data. After the first step was completed, Adam algorithm's hyperparameters were tuned which are gradient decay factor and squared gradient decay factor. Moreover, the initial learning rate parameter was tested on the network to understand the effect of the learning rate. The weights and biases were initialized at the beginning of the training progress in order to see the variation as validation losses according to adjusted hyperparameters. The 18 different combinations of tuned hyperparameters were obtained for comparison individually. In the end, the 9^{th} option gave the best results according to validation accuracy as 99.80%. In this option, the learning rate parameter was adjusted as 0.001, the gradient decay factor was set to 0, and the squared gradient decay factor was set to 0.9999. The test part of the study was accomplished according to the adjusted hyperparameters which gave the best results. The test results were quite superior by considering average accuracy, sensitivity, and specificity rates as 99.70%, 99.70%, and 99.73%, consecutively. Lastly, this study was exhibited great accomplishment rates for each class, also the number of correctly classified instances was pretty remarkable in CNN-based classification.

In future studies, the aim is to investigate different time-frequency domain approaches and different deep learning methods as well as the proposed studies. Additionally, real-time ECG signal processing might be scrutinized by using time-frequency domain techniques in the literature aside from STFT. Then, another objective is to research the patient adaptive arrhythmia classification for smart healthcare.

REFERENCES

Abdeldayem, S. S., & Bourlai, T. (2018). ECG-based Human Authentication using High-level Spectro-temporal Signal Features. *2018 IEEE International Conference on Big Data (Big Data)*, 4984–4993.

Auger, F., Flandrin, P., Gonçalvès, P., & Lemoine, O. (1996). Time-frequency toolbox. CNRS France-Rice University, 46.

Balachandran, A., Ganesan, M., & Sumesh, E. P. (2014). Daubechies algorithm for highly accurate ECG feature extraction. *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, 1–5.

Banerjee, S., & Mitra, M. (2014). A cross wavelet transform based approach for ECG feature extraction and classification without denoising. *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, 162–165.

Benmalek, E., & Elmhamdi, J. (2015). Arrhythmia ECG signal analysis using non parametric time-frequency technique. *2015 International Conference on Electrical and Information Technologies (ICEIT)*, 281–285.

Bibaeva, V. (2018). Using Metaheuristics for Hyper-Parameter Optimization of Convolutional Neural Networks. *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6.

Chen, Y.-J., Liu, C.-L., Tseng, V. S., Hu, Y.-F., & Chen, S.-A. (2019). Large-scale Classification of 12-lead ECG with Deep Learning. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 1–4.

Choudhary, T., Sharma, L. N., & Bhuyan, M. K. (2018). Spectracentrogram: A Time-Frequency Distribution for Signal Processing Applications. *2018 International Conference on Intelligent Informatics and* Biomedical *Sciences (ICIIBMS)*, 1–5.

Cohen, L. (1995). *Time-frequency analysis: Theory and applications*. Prentice-Hall, Inc.

de Albuquerque, V. H. C., Nunes, T. M., Pereira, D. R., Luz, E. J. da S., Menotti, D., Papa, J. P., & Tavares, J. M. R. S. (2018). Robust automated cardiac arrhythmia detection in ECG beat signals. *Neural Computing and Applications*, *29*(3), 679–693.

Fan, X., Yao, Q., Cai, Y., Miao, F., Sun, F., & Li, Y. (2018). Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings. *IEEE Journal of Biomedical and Health Informatics*, *22*(6), 1744–1753.

Gao, R. X., & Yan, R. (2011). *Wavelets*. Springer US.

Ghongade, R., & Ghatol, A. (2008). A robust and reliable ECG pattern classification using QRS morphological features and ANN. *TENCON 2008 - 2008 IEEE Region 10 Conference*, 1–6.

Hall, J. E., & Guyton, A. C. (2010). *Guyton and Hall textbook of medical physiology* (12th ed). Saunders/Elsevier.

Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed). Prentice Hall.

Hsieh, J. C., Tzeng, W. C., Yang, Y. C., & Shieh, S. M. (2005). Detecting ECG characteristic points by novel hybrid wavelet transforms: An evaluation of clinical SCP-ECG database. *Computers in Cardiology, 2005*, 751–754.

Huang, J., Chen, B., Yao, B., & He, W. (2019). ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. *IEEE Access*, *7*, 92871–92880.

Ince, T., Kiranyaz, S., & Gabbouj, M. (2009). A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals. *IEEE Transactions on Biomedical Engineering*, *56*(5), 1415–1426.

Jacobson, M. (2007). Time-frequency analysis of heart rate variability. *2007 9th International Symposium on Signal Processing and Its Applications*, 1–4.

Kamsing, P., Torteeka, P., & Yooyen, S. (2019). Deep Convolutional Neural Networks for plane identification on Satellite imagery by exploiting transfer learning with a different optimizer. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 9788–9791.

Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, *63*(3), 664–675.

Krishna, B. T. (2017). Fetal ECG extraction using time-frequency analysis techniques. *2017 International Conference on Robotics and Automation Sciences (ICRAS)*, 167–171.

Kumar, D., Tu, D., Zhu, N., Shah, R., Hou, D., & Zhang, H. (2017). The Free-Swimming Device Leakage Detection in Plastic Water-filled Pipes through Tuning the Wavelet Transform to the Underwater Acoustic Signals. *Water*, *9*(10), 731.

Liu, Z., Meng, X., Cui, J., Huang, Z., & Wu, J. (2018). Automatic Identification of Abnormalities in 12-Lead ECGs Using Expert Features and Convolutional Neural Networks. *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*, 163–167.

Mehta, D. B., Barot, P. A., & Langhnoja, S. G. (2020). Effect of Different Activation Functions on EEG Signal Classification based on Neural Networks. *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 132–135.

MIT-BIH Arrhythmia Database Directory, https://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm (Access Date: 2019)

Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, *20*(3), 45–50.

Optimizers Explained—Adam, Momentum and Stochastic Gradient Descent. (2019, October 16). Machine Learning From Scratch. https://mlfromscratch.com/optimizers-explained/

Pan, J., & Tompkins, W. J. (1985). A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, *BME-32*(3), 230–236.

Poojary, R., & Pai, A. (2019). Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models. *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 1–4.

Pourbabaee, B., Roshtkhari, M. J., & Khorasani, K. (2018). Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(12), 2095–2104.

Rajkumar, A., Ganesan, M., & Lavanya, R. (2019). Arrhythmia classification on ECG using Deep Learning. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 365–369.

Rangayyan, R. M. (2015). Biomedical signal analysis (Vol. 33). John Wiley & Sons.

Sen, S. Y., & Ozkurt, N. (2019). ECG Arrhythmia Classification By Using Convolutional Neural Network And Spectrogram. *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–6.

Sen, S. Y., & Ozkurt, N. (2020). Convolutional Neural Network Hyperparameter Tuning with Adam Optimizer for ECG Classification. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–6.

Shufni, S. A., & Mashor, Mohd. Y. (2015). ECG signals classification based on discrete wavelet transform, time domain and frequency domain features. *2015 2nd International Conference on Biomedical Engineering (ICoBE)*, 1–6.

Spiegl, A., Steinbigler, P., Schmucking, I., Knez, A., & Haberl, R. (1998). Analysis of beat-to-beat variability of frequency contents in the electrocardiogram using two-dimensional Fourier transforms. *IEEE Transactions on Biomedical Engineering*, *45*(2), 235–241.

Sun, Q., Liu, X., & Bourennane, S. (2019). Optimal Parameter Selection in Hyperspectral Classification Based on Convolutional Neural Network. *2019 5th International Conference on Frontiers of Signal Processing (ICFSP)*, 100–104.

Webster, J. (1998). *Medical instrumentation: application and design*. New York: John Wiley & Sons, Inc.

Wu, Z., Lan, T., Yang, C., & Nie, Z. (2019). A Novel Method to Detect Multiple Arrhythmias Based on Time-Frequency Analysis and Convolutional Neural Networks. *IEEE Access*, *7*, 170820–170830.

Xu, J., Li, T., Chen, Y., & Chen, W. (2018). Personal Identification by Convolutional Neural Network with ECG Signal. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 559–563.

Yu, J., & Ramamoorthi, R. (2019). Robust Video Stabilization by Optimization in CNN Weight Space. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3795–3803.