UNIVERSITY OF TURKISH AERONAUTICAL ASSOCIATION
INSTITUTE OF NATURAL AND APPLIED SCIENCES

LINEAR PREDICTION CODING AND WAVELET BASED MULTI HEART DISEASES CLASSIFICATION VIA SVM

MASTER THESIS

ALAA MAJEED Al-OBAIDI

Electrical and Electronics Engineering Department

Master Thesis Program

AUGUST 2017

UNIVERSITY OF TURKISH AERONAUTICAL ASSOCIATION
INSTITUTE OF NATURAL AND APPLIED SCIENCES

LINEAR PREDICTION CODING AND WAVELET BASED MULTI HEART DISEASES CLASSIFICATION VIA SVM

MASTER THESIS

ALAA MAJEED Al-OBAIDI

1403630042

Electrical and Electronics Engineering Department

Master Thesis Program

Supervisor: Asst. Prof. Dr. Hassan SHARABATY

Alaa Majeed AL-OBAIDI, having student number 1403630042 and enrolled in the Master Program at the Institute of Science and Technology at the University of Turkish Aeronautical Association, after meeting all of the required conditions contained in the related regulations, has successfully accomplished, in front of the jury, the presentation of the thesis prepared with the title of: "Linear Prediction Coding and Wavelet Based Multi Heart Diseases Classification via SVM"

Supervisor **: Assist. Prof. Dr. Hassan SHARABATY**
University of Turkish Aeronautical Association .....

Jury Members **: Assist. Prof. Dr. Abdül Kadir GÖRÜR**
Çankaya University .....

: Assist. Prof. Dr. Yuriy ALYEKSYEYENKOV
University of Turkish Aeronautical Association .....

: Assist. Prof. Dr. Hassan SHARABATY
University of Turkish Aeronautical Association .....

Thesis Defense Date: 15.08.2017

STATEMENT OF NON-PLAGIARISM PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

08.08.2017

Alaa Majeed Al-OBAIDI

DEDICATION

I dedicate my thesis work to my loving parents, who have always loved me unconditionally and who have been a constant source of support and encouragement during the challenges of graduate school and life. A special feeling of gratitude goes to my seven sisters and one brother who have never left my side and are very special to me. I also dedicate this dissertation to my friends who have supported me throughout the process. I will always appreciate all that they have done. I dedicate this work and give special thanks to my two wonderful sons Obeida and Ibraheem for being there for me and as an apology on my dereliction in my motherhood throughout the entire master program. I am truly thankful for having you in my life. Both of you have been my best cheerleaders.

Temmuz 2017 Alaa Majeed Al-OBAIDI

ACKNOWLEDGEMENTS

I would like to extend sincere thanks to my God, who helped and supported me in the execution of this thesis. Special thanks go to the members of my family, and because of their prayers and support, I have reached my goal. Finally, but by no means least, thanks go to mum, dad and sons for almost unbelievable support. They are the most important people in my world.

Foremost, I would like to express my sincere gratitude to my enthusiastic supervisor Dr. HASSAN SHARABATY, who so generously contributed to the work presented in this thesis. I thank him wholeheartedly, for his tremendous academic support, for his countless hours of reflection, reading and encouragement. His guidance, and most of all his patience, helped me throughout the entire process and duration of the research and writing of this thesis.

In addition to my advisor, I would like to thank the remaining thesis committee members for their support, insightful comments: Dr. Yuriy ALYEKSYEYENKOV, and Dr. Abdül Kadir GÖRÜR. I offer my sincere appreciation for the learning opportunities provided by them.

Special mention goes to the University of Turkish Aeronautical Association and its academic stuff for their cooperation. Special thanks and best regards goes to the members of staff.

Finally, I would like to extend my thanks and appreciation to Al-Iraqia University and the Ministry of Higher Education and Scientific Research in Iraq for allowing me to conduct my research and providing financial support.

Alaa Majeed Al-OBAIDI

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ECG : Electrocardiogram
SA : SinoArtial
AV : Atrioventricular
FHR : Fetal heart rate
FECG : Fetal electrocardiogram
OB : Obstetrics
SS : Salt-Sensitive
CWT : Continuous Wavelet Transformation
FFT : Fast Fourier Transform
MMSE : Minimum Mean Square Error
DWT : Discrete Wavelet Transform
XWT : Cross Wavelet Transform
SODS : Second order dynamic system
HRN : Hybrid-recurrent Network
SVM : Support Vector Machine
KNN : k-neural network
LPC : Linear Predictive Coding
AR : Autoregressive
LDA : Levinson Durbin Algorithm
FB : Filter Bank
LS : Lifting Scheme
ICP : Iterative Closest Point
MSE : Mean Square Error
PSNR : Peak Signal to Noise Ratio
DB4 : Debauchee -4
LOOCV : Leave one out Cross Validation

ABSTRACT

LINEAR PREDICTION CODING AND WAVELET BASED MULTI HEART DISEASES CLASSIFICATION VIA SVM

Al-OBAIDI, Alaa Majeed

Master Thesis, Department of Electrical and Electronics Engineering

Thesis Supervisor: Asst. Prof. Dr. Hassan SHARABATY

August 2017, 50 pages

The human heart is considered to be a main part in the human body since it controls the blood in whole parts of the organs. Therefore, diagnosis of heart diseases is considered to be of major importance in biological and clinical systems as these diseases cause more than 15 million deaths per year. The examination of Electrocardiogram (ECG) signals using computerization techniques involves a number of procedures such as pre-processing, feature extraction and post-processing. Most studies focusing on this domain have classified ECGs into two cases (normal and abnormal) for heart cases. Other researches have been interested in the analysis of ECG standard features of normal case (P wave, PR interval, QRS complex, ST segment, T wave, QT interval), and some of these works have applied Artificial Intelligent (AI) methods which require high level operations in processing. In this thesis, we propose an accurate method in order to detect 6 different cases of heart diseases by using Linear Predictive Coding (LPC) as signal regressing and normalizing during the preprocessing phase. Then, during the feature extraction phase, we use Discrete Wavelet Transform (DWT) with 8-level to transfer the ECG signal from time series domain to frequency domain and Wavelet Energy (WE) in order to extract the feature vector. After that, using Support Vector Machine (SVM) in the classification phase to train datasets and classify the test samples and finally finding the closest vector of

testing to train datasets by applying Iterative Closest Point (ICP). Performance of our work was evaluated by using the confusion matrix which gave an accuracy rate of 98.14% for 6 diseases.

Keywords: Abnormal ECG, Multi Heart Diseases classification, Linear Prediction Coding, Discrete Wavelet Transform, Support Vector Machines, Iterative Closest Point.

ÖZET

LİNEER TAHMİNİ KODLAMA VE DALGACIK TABANLI SVM ÜZERİNDEN ÇOKLU KALP HASTALIKLARININ SINIFLANDIRILMASI

Al-OBAIDI, Alaa Majeed
Yüksek Lisans Tezi, Elektrik-Elektronik Anabilim Dalı
Tez Danışmanı: Yrd. Doç. Dr. Hassan SHARABATY
Agustus 2017, 50 sayfa

Organlarda bulunan kanı kontrol ettiği için, insan kalbinin insan vücudunun ana parçalarından birisi olduğu kabul edilmektedir. Kalp hastalıkları yılda 15 milyon'dan fazla insanın ölümüne sebep olmasından dolayı kalp hastalıklarının teşhisi, hem biyolojik hem de klinik sistemlerde büyük önem arz etmektedir.. Bilgisayar teknikleri kullanılarak yapılan Elektrokardiyogram (EKG) sinyallerinin incelenmesi ön-işleme, özellik çıkarımı ve işleme sonrası gibi bir takım işlemleri içermektedir. Bu alanda yapılan çalışmaların çoğu kalp hastalıkları için EKG'leri (normal ve anormal olmak üzere) iki vakaya ayırmaktadır. Diğer araştırmacılar normal vakada EKG standart özelliklerinin analizi ile ilgilenmiş (P dalgası, PR aralığı, QRS kompleksi, ST segmenti, T dalgası, QT aralığı), bu çalışmalardan bazıları ise işlemlerde yüksek seviye operasyonlar gerektiren Yapay Zeka (AI) yöntemlerini uygulamışlardır. Bu tez çalışmasında, ön işleme aşaması sırasında sinyal gerileme ve normalleşme için Lineer Tahmini Kodlamayı (LPC) kullanarak kalp hastalıklarının 6 farklı vakasını tespit etmek için doğru bir yöntem ortaya koymaktayız. Daha sonra, özellik çıkarımı aşamasında, özellik vektörünü çıkarmak için EKG sinyalini zaman serileri alanından frekans alanına transfer etmek için 8-seviyeli kesikli dalgacık dönüşümünü (DWT) ve Dalgacık Enerjisini (WE) kullanmaktayız. Bundan sonra, sınıflandırma aşamasında veri kümelerini çalıştırmak ve test numunelerini sınıflandırmak için Destek Vektör Makinasını (SVM)

kullanmakta ve son olarak Yinelemeli En Yakın Noktayı (ICP) uygulayarak veri kümelerini çalıştırmak için denemenin en yakın vektörünü bulmaktayız.

Çalışmamızın performansı 6 hastalık için %98,14 doğruluk oranı veren karışıklık matrisi kullanılarak değerlendirilmiştir.

Anahtar kelimeler: Anormal EKG, Çoklu Kalp Hastalıkları sınıflandırması, Lineer Tahmini Kodlama, Kesikli Dalgacık Dönüşümü, Destek Vektör Makinaları, Yinelemeli En Yakın Nokta.

CHAPTER ONE

INTRODUCTION

1.1 Background

The human heart is considered to be the most important organ because it controls the blood in different part of the body. One of the most common techniques for monitoring heart functions is reading ECG signals and diagnosing various heart conditions. The analysis of ECGs using computerization techniques requires some processing such as pre-processing, feature extraction, and post-processing. There are many methods that have been used in each process and the main goal of the approaches in their analysis of the ECG signal is to best classify and achieve a better recognition rate.

1.2 Electrocardiogram (ECG)

The electrocardiogram (ECG) is a nearly periodic signal that represents the activity of the heart, and this information helps to diagnose many heart diseases [1]. Instead of that, ECG has waveforms which presents the signals of heart beats, and these waveforms determine the heart rate by calculating the time between beats and between pulses types [2].

The ECG has components and different waves representing ECG signals the components of which are described in Figure 1 below [3].

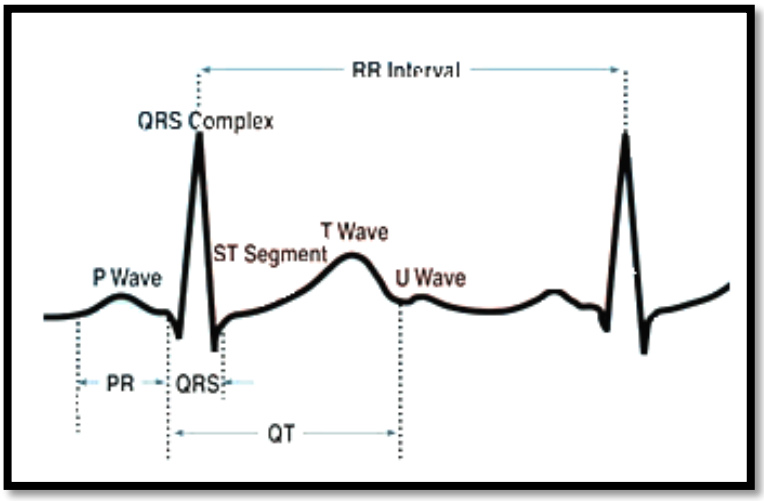


Figure 1: ECG normal condition components.

1.3 The Heart Anatomy

The heart, positioned in the mediastinum, is an essential structure of the cardiovascular system. In addition, it is secured by the skeletal structures of the sternum anteriorly, the backbone column posterior, and the rib cage.

SinoArtial (SA) nodes are the dominant leader of the heart, and are located in the upper part of the right atrium. In addition, it has an intrinsic ratio falling between 60 and 100 bpm.

Atrioventricular (AV) nodes are a part of the AV junction soft tissue. It decelerates transmission, making a small postponement before instincts reaching ventricles. It has a basic rate of 40 to 60 bpm.

Table 1: Electrophysiology.

Action	Effect
Depolarization	Shifting of electrolytes across the cell membrane causes change in the electric charge of the cell resulting in contraction.
Repolarization	Internal negative charge is restored and the cells return to their resting state.

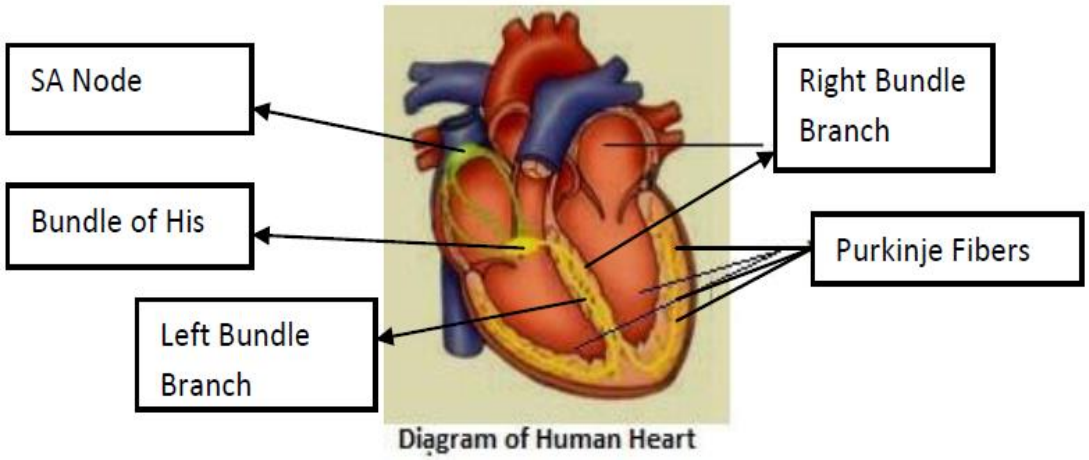


Figure 2: Conduction System structure.

1.4 Lead Systems

Twelve leads usually comprise a diagnostic ECG recording: six limb leads (three bipolar and three unipolar), and six unipolar precordial leads. The instantaneous cardiac scalar voltages resulting from the electrical activity in the heart are measured in each of the 12 leads. Since the cardiac vector varies in magnitude with time over a three-dimensional space, it is important to know its presentation (i.e. appearance or projection) in each of the 12 leads of the ECG [4].

Figure 4 shows the lead placement to acquire the 12-lead ECG. The leads can be categorized into the front leads (I, II, III, aVR, aVL, and aVF), and the crosswise leads (V1, V2, V3, V4, V5, and V6). The front leads determine the projection of the cardiac amount on the front plane of the human body. The front plane is placed in parallel to the ground when lying flat. The transverse, or precordial leads, determine the projection of the cardiac amount on the horizontal flat, (i.e. the plane that is parallel to the floor when standing) [5].

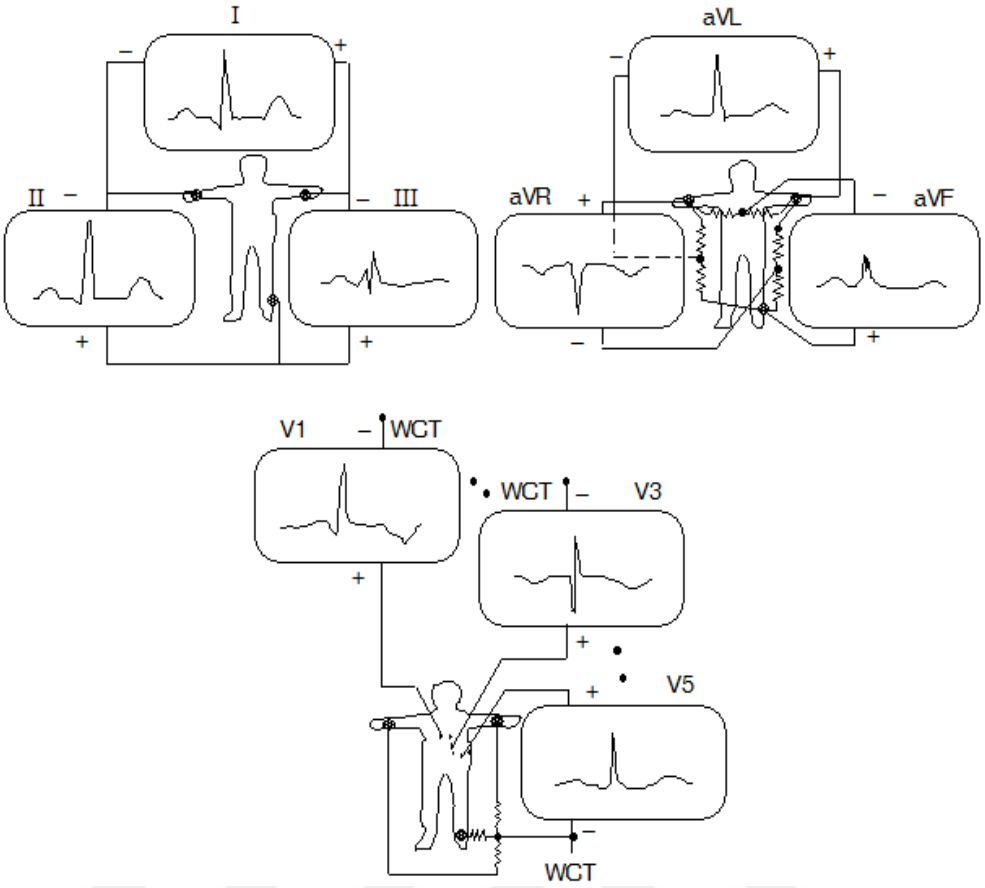


Figure 3: Leads System of ECG.

1.5 Calculating Heart Rate

To find the heart rate, we need to know the following:

a) R-R interval in terms of numbers of large squares

If the R-R intervals are constant 3 large squares, then the heart rate is 300/3=100

The following characteristics should be taken into consideration during paper calibration:

a) 1 inch = 1 second
b) Each inch is divided by dark black lines into 5 large squares.
c) Each large square= 1/5=0.2 sec
d) Each large square is further divided into 5 small squares.
e) 1 small square = 0.2/5= 0.04 sec
f) 1smallsquare = 0.04 sec =1 mm
g) 1 second =25 small squares
h) 60 second =1 min =25 ×60=1500 small squares =300 large squares

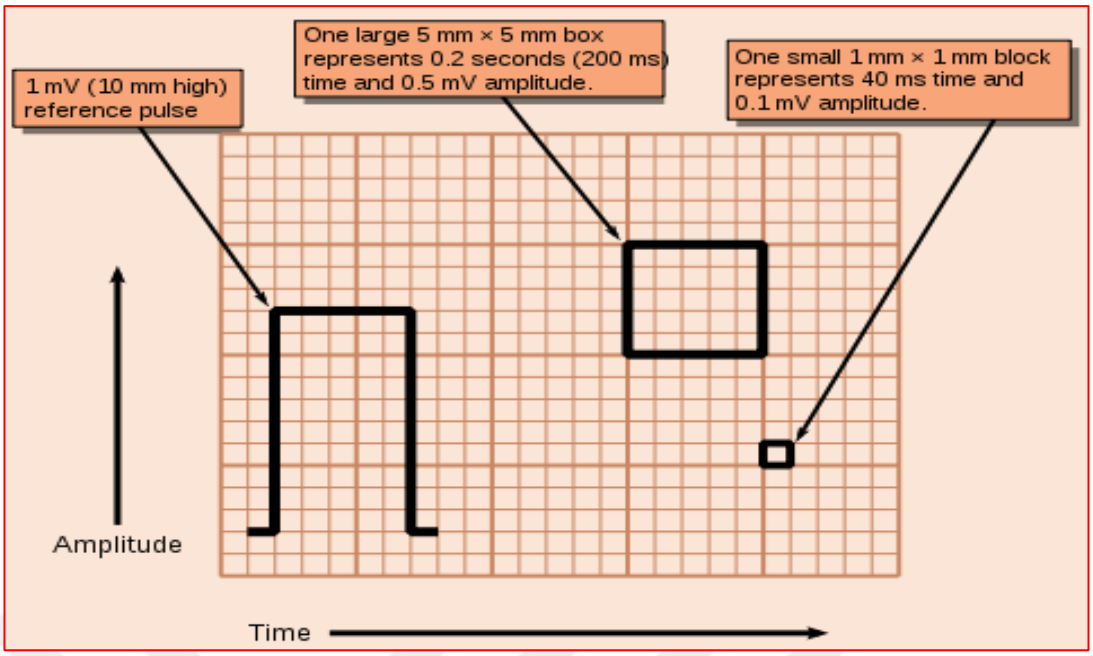


Figure 4: Calculating heart rate.

1.6 Abnormal ECG Description

1.6.1 Abdominal and Direct Fetal

Clinical fetal heart rate (FHR) is an ongoing observation of doctor fetal physiology that comes from the FHR pattern appearing to accept certain characteristics under the effect of numerous hypoxic and non-hypoxic influences. It is important for the clinician to have a simple understanding of the physiology of the fetal respiratory system and the physiological control of the FHR.

Thus, the fetal electrocardiogram (ECG) signal offers the clinician a portion of the electrical pulses of the fetal heart. It is not too essential to signify mechanical activity. The fetal ECG signal is developed a bipolar electrode that enters the skin of the fetal scalp (first pole), and that has a second electrode existing in the secretions of the mother's vagina (second pole). The circuit occurs through the fetal umbilical cord, placenta, and the vagina. The potential variance (voltage) being slow is between the two poles. The original electrode was adapted as a skin clip, but now a spiral electrode is used.

The research material involved in the Abdominal and Direct Fetal Electrocardiogram Database holds multichannel fetal electrocardiogram (FECG) recordings found from five diverse women in labor between 38 and 41 weeks of

development. The recordings were developed at the Department of Obstetrics at the Medical University of Silesia, by income of the KOMPOREL system for the acquisition and examination of fetal electrocardiograms (ITAM Institute, Zabrze, Poland). Each recording includes four difference signals developed from the mother's abdomen and the reference direct fetal electrocardiogram recorded from the fetal head [6].

1.6.2 Ventricular Fibrillation

Ventricular fibrillation is the heart tremors in their place of forcing due to unsystematic electrical motion in the ventricles. The result is a cardiac hold with an absence of notice and no beat. This is tracked by permanent death in the absence of treatment. Ventricular fibrillation is established firstly in around 10% of persons experiencing cardiac arrest.

Ventricular fibrillation can occur without coronary heart illness, valve heart sickness, cardiomyopathy, Bragada syndrome, long QT condition or intracranial hemorrhage. Analysis is performed by an electrocardiogram (ECG) presentation of uneven formless QRS complexes deprived of any strong P waves. A significant difference of the analysis is torsades de pointes.

Ventricular fibrillation usually occurs in unhealthy hearts, and is an appearance of fundamental ischemic heart sickness. Ventricular fibrillation is found with cardiomyopathy, myocarditis, and other heart diseases. Moreover, it is realized with electrolyte instabilities and overdoses of cardiotoxic medications. It is also distinguished by the ventricular fibrillation that occurs when there is no visible heart disease or other obvious reason, namely the so-called idiopathic ventricular fibrillation.

Idiopathic ventricular fibrillation occurs in approximately 1% for all conditions of out-of-care places capture, as well as 3-9% of the conditions of ventricular fibrillations dissimilar to myocardial infarction, while 14% of complete ventricular fibrillation recoveries in sick persons under 40. It tracks because ventricular fibrillation this one is public, idiopathic ventricular fibrillation explanations for a considerable humanity. Recently, defined syndromes, such as the Bragada Syndrome, may provide indications to the fundamental tool of ventricular arrhythmia. In Bragada Syndrome, variations may originate in the inactive ECG with an indication of a right bundle

branch block (RBBB) and ST rise in the chest leads V1-V3, a fundamental tendency to unexpected heart death.

Familial circumstances that predispose persons to evolving ventricular fibrillation and unexpected cardiac death are often the result of gene changes that touch cellular transmembrane ion frequencies.

In occurrences of cardiac failure, fibrillation nearly always comes first by a run of ventricular tachycardia, which finally leads to the fibrillation itself. The onset of fibrillation is very difficult to pinpoint in many cases. Any clinically valuable detector should reply to the runs of tachycardia prior to fibrillation while medical intervention is desired at the earliest opportunity. Thus, any detector responding to the premonitory tachycardia can show a negative "time to alarm" likened to the beginning of fibrillation as recorded in the space annotation files. For this purpose, the database is distinct as a tachyarrhythmia database rather than a fibrillation database [7].

1.6.3 Intercardiac Artial Fibrillation

Artial fibrillation (AF) accounts for approximately 400,000 annual hospital discharges and affects an estimated 2.3 million adults at any given time, more than any other abnormality of heart rhythm. From an epidemiological stand point, AF is predominantly a disease of elderly individuals with cardiovascular disease. As a consequence, two trends affecting industrialized countries are expected to increase the prevalence of AF in the coming decades: the growth of the elderly population and improved survival with cardiovascular conditions such as myocardial infarction and heart failure.

Framingham data indicate that men and women at age 40 have a residual lifetime risk of approximately 1 in 4 of developing AF. This risk remains relatively constant for individuals reaching older ages free of AF, as the shorter lifespan is balanced by the sharp age-related increase in the risk of AF. Thus, the lifetime risk of AF is substantial and higher than the risk of other, non-cardiovascular conditions that affect elderly individuals, such as breast cancer (1 in 9) and hip fracture (1 in 6 for White women, 1 in 20 for White Men).

The regularity of AF in the populace, the numbers with and without cardiac sickness, and its strong correlation with age implies that there are numerous etiologies for these arrhythmias, which may be part of a common pathology and path physiology.

AF is related to practically any type of fundamental heart disease that reasons variations of the Artial myocardium, counting distensions, inflammations, hypertrophies, ischemia, fibroses and infiltrations. Moreover, there are usual age-related variations of the myocardium, with amyloid deposits and fibroses, perhaps accounting for the increased rate of AF in the elderly. Parasympathetic or sympathetic nervous scheme contributions alter Artial electrophysiological belongings and can aggravate AF. Universal infection, pulmonary disease and infection, pulmonary embolisms, hyperthyroidisms, and certain toxins or metabolic abnormality may promote AF smoothing in the nonexistence of underlying Artial sickness. In about 15% of conditions, there is no structural heart disease and no recognizable reason for arrhythmia; this is termed ‘‘lone AF’’ [8].

1.6.4 OB-1

In medicine (obstetrics), the word fetal distress is mentioned in the attendance of signs in a pregnant woman—before or during delivery, which suggests that the fetus may not be healthy. Because of its lack of accuracy, the term is avoided in modern American obstetrics.

In its place of referring to "fetal distress," current references endeavor to look for more exact signs and indications, assess them, and take suitable steps to medicate the situation through the application of intrauterine resuscitation. Usually the diagnosis of "fetal distress" forces the obstetrician to instigate rapid delivery by influential delivery or by caesarean unit if vaginal delivery is not recommended.

The fetal distress stands as being ill-defined, secondhand to definite intrauterine fetal, a result of intrauterine fetal hypoxia. Non-reassuring fetal rank is considered by tachycardia or bradycardia, summary FHR variability, nonappearance of accelerations spontaneous or elicited. It must be underlined that the hypoxia and acidosis issue is the ultimate result of the numerous reasons for intrauterine fetal negotiation [9].

1.6.5 Blood Pressure

Salt-sensitive hypertension is identified to be related to dysfunction of the baro-reflex control system by Dahl salt-sensitive (SS) rats. However, neither the physiological tools nor the genomic regions basic to the baro-reflex dysfunction seen

in this typical rat are ultimately identified. Now, we have assumed an accurate modeling method to examine the physiological and genetic origins of baro-reflex dysfunction by Dahl SS rat.

Blood pressure is spread unceasingly; however, the distribution is twisted to the higher end of the bend. There is a straight and quantitative association between higher blood pressure and humanity. The lack of a described bimodal distribution of blood pressure proposes that it is controlled by a compound group of interrelating genes. This is protected by the lack of a clear-cut manner of legacy; in addition, there is no indication that vital hypertension is a dominant or receding trait. Approximately half of blood pressure variability is supposed to be due to heredity; however, the difference of blood pressure is the outcome the relationship between genetic and environmental aspects. Genetic studies of important hypertension have used two methods: family-based linkage studies and the study of the suggestion of candidate genes in a population based project. Numerous loci have been related to hypertension, and alternatives of many genes have been stated to be related to hypertension, albeit conflictingly. Harrap has opined that, instead of penetrating every allele that controls blood pressure, a determination should be made to explore for molecular clues to the shared physiologic mechanisms fundamental to disease [10].

1.6.6 Gait Neurodegenerative

Neurodegeneration is the advanced loss of construction or function of neurons, with the death of neurons. Many neurodegenerative diseases with amyotrophic side sclerosis, Parkinson's, Alzheimer's, and Huntington's occur as a result of neurodegenerative conditions. Such diseases are incurable, resulting in advanced deterioration and/or death of neuron cells. As study progresses, many comparisons appear that relay these diseases to one addition on a sub-cellular near. Determining these comparisons offers confidence for therapeutic developments that could ameliorate many diseases simultaneously. There are many counterparts between dissimilar neurodegenerative complaints counting atypical protein assemblies as well as encouraged cell death. Neurodegeneration can be created in many dissimilar stages of neuronal circuitry extending from molecular to systemic.

Numerous neurodegenerative diseases are produced by genetic mutations, greatest of which are positioned in completely unrelated genes. In many of the

dissimilar diseases, the mutated gene has a shared piece: a duplication of the CAG nucleotide triplet. CAG converts the amino acid glutamine [11].

1.7 Statement of the Problem

The greatest reason for this research is to study the ECG signal which is considered to be an important matter relating to the human being. Moreover, this research uses various methods in different phases to achieve a high accuracy of classification.

1.8 Research Design

This research focuses on four main subjects: collecting data of recorded ECG signals, preprocessing these data, feature extraction from these data and post-processing and performance of the proposed work.

The data of recorded ECG signals have been gathered from the PhysioNet web service which has a large database containing many datasets of normal and abnormal ECGs.

In the preprocessing phase, a method of regression data is applied to transform the magnitude of the ECG signal to a significant magnitude placed in the range $0 \rightarrow 1$. The wavelet transformation is applied after the regression method to reduce the length of the data based on the number of wavelet levels.

The feature extraction method used in this work to extract of ECG is the Wavelet Energy based on the number of wavelet coefficients.

The last phase of work is post-processing, where two methods are applied: SVM for classification by training and test data, and ICP being applied to find the best closest match of the test sample to the training data.

The following parts of this thesis are prepared as following:

In Chapter 2, we have categorized several approaches into three categories, namely ECG signal preprocessing methods, feature extraction and fusion methods, during ECG classification and feature extraction.

Then in Chapter 3, the fundamentals and principles of the methods and algorithms applied during the phases of proposed work are discussed with mathematical models and flowcharts. In addition, the evaluation metrics used in this

work are described in this chapter.

After that in Chapter 4, the experimental results are represented both numerically and graphically through each phase of the proposed work. In addition, the performance of results is discussed based on the evaluated metrics and accuracy.

Finally, in Chapter 5 both Conclusion and future work have discussed in this chapter.

CHAPTER TWO

LITERATURE SURVEY

In this chapter, we will summarize the previous works done in the 3 phases of ECG analysis, i.e. in the ECG pre-processing phase, in ECG Feature Extraction phase, and finally in the classification phase.

2.1 Literature Studies in ECG Signal Pre-processing

The study [12] focused on ECG signal regression using continuous Wavelet Transformation (CWT) applied to frequency energy on QRS waves distribution. This work was applied on sleep apnea syndrome and the detection results were 98.11%.

However, the basic concept of [13] laid an extract on the fetal electrocardiogram after transforming the original signal of ECG from high dimensional Hilbert space into a linear regression using kernel technology.

[14] focused on high frequencies of ECG considered as a feature for identification systems. The Neural Network was applied to a justification signal and normalization on the amplitude. The classification rate was 99%.

In [15] the researchers focused on ECG preprocessing to eliminate the noise carried with ECG signals. The algorithms used in this study included the wavelet, FFT and nonlinear Bayesian filter.

The study [16] presented a noise reduction method for heart sounds. The method used for noise reduction was the Minimum Mean Squared Error (MMSE) without the need to separate the reference signal; they used an SNR and Wiener filter to gain the amplitude of the signal to increase the noise level.

2.2 Literature Studies in ECG Feature Extraction

DWT was proposed to decompose the R-R intervals representation of the ECG signal in [17]. The features of the DWT coefficients were used as hybrid features for training and testing using the neural network as a classifier. The total classification accuracy was 98.2%.

However, in [18] the researchers presented the cross wavelet transform (XWT) and wavelet coherence (WCOH) methods to reduce the feature extraction. The accuracy was 93.9%.

[19] proposed a second order dynamic system (SODS) technique in the feature extraction of ECG signals as well as the Hybrid-recurrent Network (HRN).

2.3 Literature Studies ECG Classification

A classification method using a hierarchical system based on SVM and the decision rule using full heart beating in a time series by alignment with R-peaks was proposed in [20]. The main goal of this paper was to classify abnormalities in ECG. The total classification accuracy was 97.3%.

In [21] a comparative study focused on ECG classification between ANFIS and SVM to classify normal and abnormal heart beats. The performance of DWT-ANFIS and DWT-SVM in terms of total classification accuracy was 90.0%.

Researchers in [22] presented a method to classify multi-lead ECG signals for normal or abnormal ECGS using CSP as the feature extractor. The work consists of two main stages: CSP-based feature extraction and classification. Three classifiers were employed in the classification stage: LDA, NB, and SVM. The total classification accuracy was 88.2%.

[23] proposed an analysis of ECG signals by applying extreme learning machine to classify ECGs into normal and abnormal classes. The PQRS features and QRS complex were extracted for classification of normal and abnormal beats. The accuracy of proposed method using ELM was 97%.

The study [24] proposed a novel technique for ECG classification based on the Legendre polynomial, simple logistic. KNN was applied to classify the ECG into normal and abnormal. The total classification accuracy was 90.0%.

Three different feature sets are applied separately onto a multilayer perceptron

neural network MLP-NN classifier in [25]. The WT based feature extraction method provides classification performance with an average accuracy of 96.0%, whereas S-transform based features along with temporal features yield classification performance with a 96.9 % average accuracy. On the other hand, the mixture of ST and WT based features along with temporal features shows performance at 97.5% average accuracy.

In Table 2, we summarize all the literature studies on the different ECG classification technique that used MIT-BIH as database.

Table 2: Literature studies on the different ECG classification technique using MIT-BIH database.

Author(s)	Features Extraction	classifiers	Classes	Acc.
Stalin et al. (2015)[23]	WD	- Extreme Learning Machine (ELM) - Support Vector Machine (SVM) - Back Propagation Neural Network (BPN).	2 normal & abnormal	97% 89% 84%
Irshad et al. 2015 [24]	Legendre Polynomials	- simple logistic - K nearest neighbors	2 normal & abnormal	100% 84.4 %
Czarina et al. (2016) [21]	- ANFIS paired with Haar mother WT - SVM paired with mother wavelet DB	- Adaptive Neuro-Fuzzy Inference System ANFIS - SVM using kernel function Radial Basis Function RBF	2	90% 95%
Aljafar et al. (2016) [22]	Common Spatial Pattern CSP	- naïve Bayes (NB) - support vector machine (SVM)	2	84% 100%,
Saminu et al. (2014) [17]	- RT equivalent feature - DWT with statistical	neural network backpropagation algorithm	3 beat	88.22% 99.8%

Table 2 (continuation): Literature studies on the different ECG classification technique using MIT-BIH database.

Manab Kumar Das and Samit Ari (2014) [25]	- WT based features - ST based features - the mixture of ST and WT based features all with along with temporal feature set.	multilayer perceptron neural network (MLPNN)	5 beats	96 % 96.9 % 97.5 %

CHAPTER THREE

PRINCIPLES AND FUNDAMENTALS OF PROPOSED WORK

3.1 Linear Predictive Coding (LPC)

The LPC strictures are a fundamental component of several compression algorithms that are applied in cellular telephony for bandwidth compression and improved privacy. Bandwidth is preserved by decreasing the data rate necessary to characterize the speech signal. This data rate decrease is completed by parameter zing speech in relation to the all-pole filter coefficients and a minor set of excitation strictures.

Linear prediction is a set of mathematical functions where future values of a digital signal are valued as a linear function of earlier samples. The LPC is separated into two parts, namely the linear predictive coefficients and the prediction error signal.

The Linear Predictive Coding is an algorithm that is used to regress the signal. This method of regression is also known as the "All-pole" model or the autoregressive AR model [26]. The predicator function equation can be written as:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] \qquad (\mathbf{3.1})$$

Where $s[n]$ is the need to be predicated and a_k is the coefficient predicator.

There are three steps of LPC, namely autocorrelation, the reflection coefficient, and the Levinson Durbin algorithms, as shown in Figure 5.

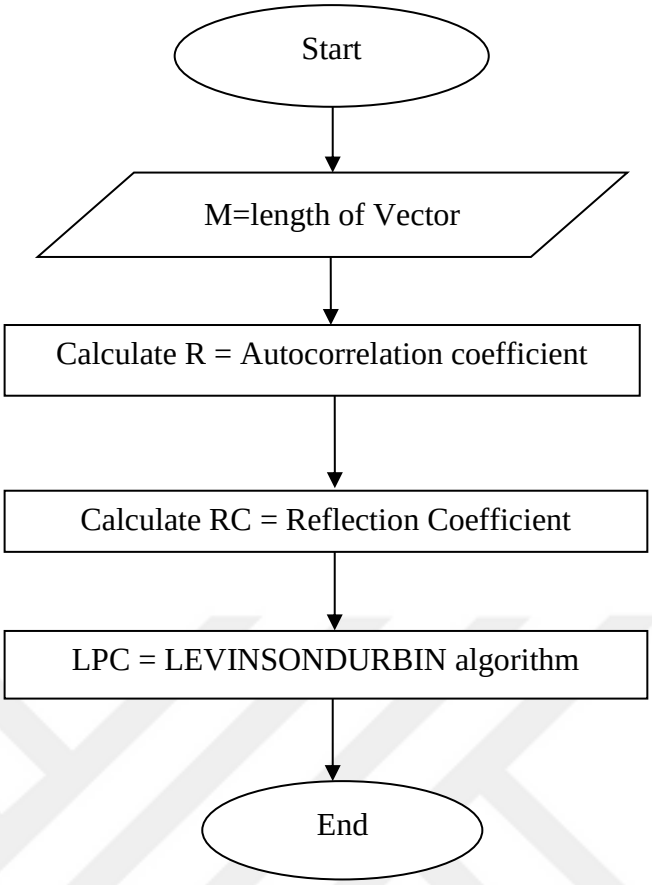


Figure 5: Main steps of LPC.

Below we will discuss each step in detail in both the math model and the flow chart.

3.1.1 Autocorrelation Coefficient

The concept of the autocorrelation coefficient is to measure the sum of the linear correlations between observations of two data intervals. The autocorrelation coefficient has numerical values, which are placed between +1 and –1. The value of +1 means that the observations of data have a strong direction in the positive gradient, whereas a value of –1 means that the observations of the data have a strong direction in the negative slope [27].

The correlation between any two values of some arbitrary sequence such 1,..., n–1, are considered to be an autocorrelation coefficient for this sequence. In addition, the standard autocorrelation coefficient is based on pairs of observations [28].

The equation of autocorrelation coefficient between *x* and *y* is given by [29]:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum(x_i - \bar{x})^2\right]^{\frac{1}{2}}\left[\sum(y_i - \bar{y})^2\right]^{\frac{1}{2}}} \quad (\mathbf{3.2})$$

In the case of a sequence of time series, we will find a difference in the sequence, where the correlation is calculated between one time series and the following one of time units. The first-order autocorrelation coefficient is considered the simple correlation coefficient of the first unit, which is x_t, t=1, 2,..., N–1 and the next unit is x_t, t=2,3,...,N; then the correlation between x_t and x_t+1 is given by:

$$r_1 = \frac{\sum_{t=1}^{N-1}\left(x_t - \bar{x}_{(1)}\right)\left(x_{t+1} - x_{(2)}\right)}{\left[\sum_{t=1}^{N-1}\left(x_t - \bar{x}_{(1)}\right)^2\right]^{\frac{1}{2}}\left[\sum_{t=2}^{N}\left(x_t - \bar{x}_{(2)}\right)^2\right]^{\frac{1}{2}}} \quad (\mathbf{3.3})$$

Where $\bar{x}_{(1)}$ is the first value of N–1 and $\bar{x}_{(2)}$is the last unit of N–1.

The autocorrelation coefficient processing is presented in Figure 6.

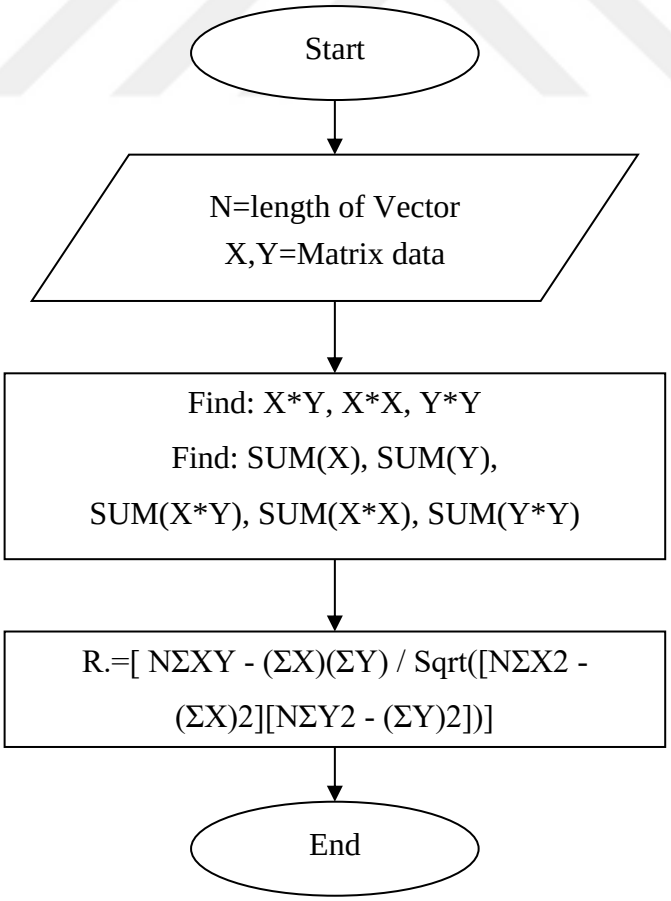


Figure 6: Autocorrelation Coefficient processing.

3.1.2 Reflection Coefficient

The second step in LPC algorithm is to find a reflection coefficient for result of autocorrelation coefficient. The basic definition of reflection coefficient is the difference of amplitude of the reflected wave divided by the sum of v amplitude of the incident wave, as given in equation (3.4).

$$R.c = \frac{amplitudeofthereflectedwave}{amplitudeoftheincidentwave}$$

$$Rc = \frac{V_2 - V_1}{V_2 + V_1} \quad (\mathbf{3.4})$$

Where V_1 represented the limited of the reflected medium, and V_2 represented the limited of incident medium.

The reflection coefficients procedure has illustrated as shown in Figure 7.

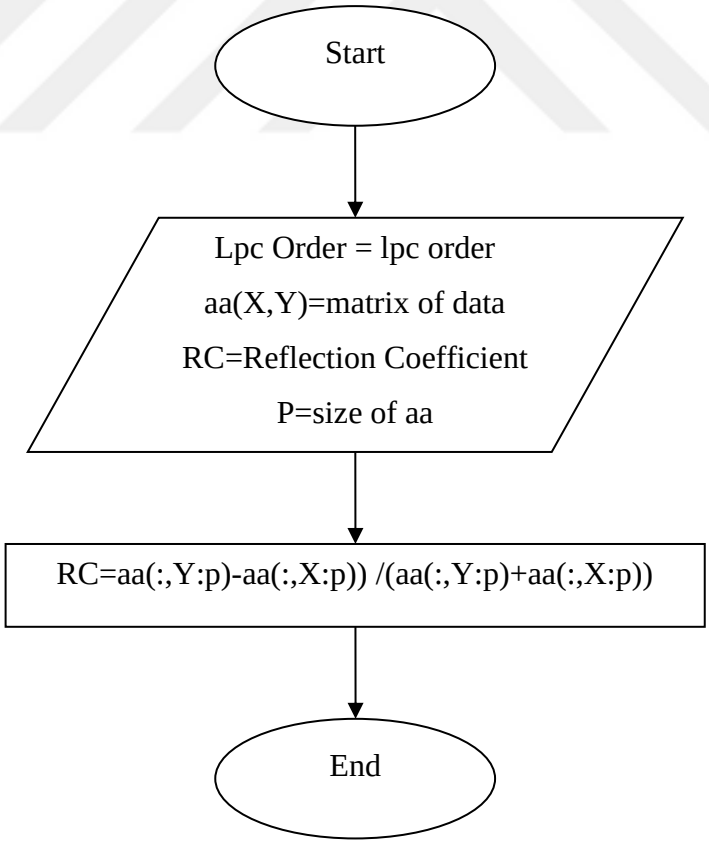


Figure 7: Reflection Coefficient processing.

3.1.3 Levinson Durbin Algorithm

The Levinson Durbin Algorithm (LDA) has used to solve equal dimension of linear matrix, which required number of predictor order, and the recursion keep processed until reach the last predictor order [30,31].

$$k_i = \frac{\left(E(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} E(i-j)\right)}{E^{i-1}} \quad (\mathbf{3.5})$$

$$E^i = (1 - k_i^2) E^{i-1} \quad (\mathbf{3.6})$$

Where E is prediction error, and α is prediction coefficient.

The procedure of Levinson Durbin algorithm has illustrated as shown in Figure 8.

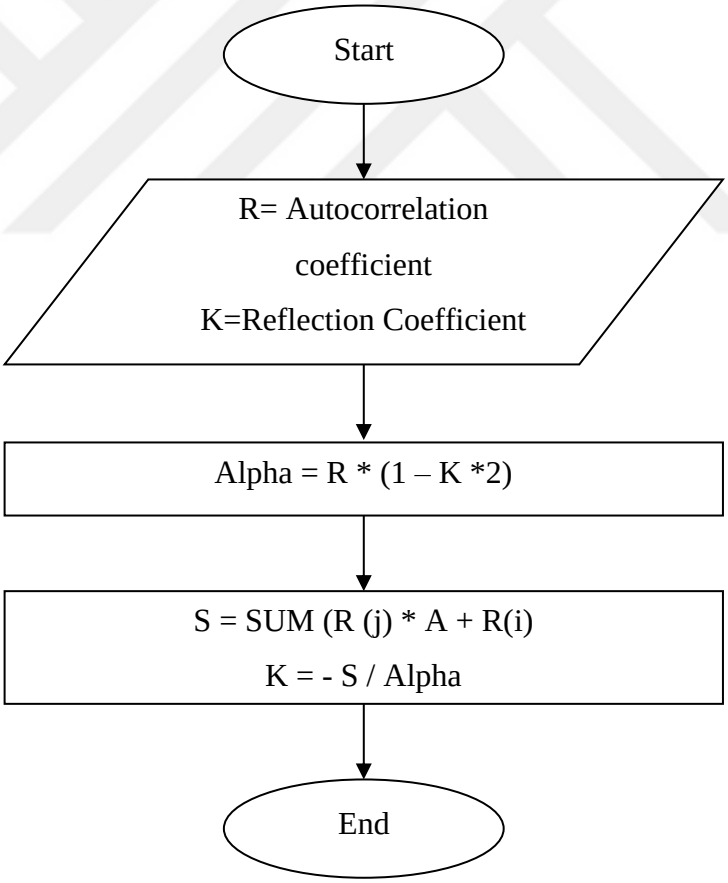


Figure 8: Levinson durbin algorithm.

3.2 Discrete Wavelet Transform

The wavelet transform has working on a time-frequency domain to analysis the signals. There are two principle methods to implement discrete wavelet transforms, both of these methods are constructed on features of time domain or frequency domain. The frequency domain feature is based on method Filter Banks (FB) and the time domain feature is based one method Lifting Scheme (LS). Below I will describe both of these methods.

3.2.1 Filter Bank Method

The FB method, in one level of discrete wavelet decomposition, where the input signal has been divided into two parts of frequency where pass these parts at the same time over a pair of low pass, H(z), and high pass, G(z), filters, as shown in Figure 9. Also, the sub samplings of the filter are output to have output the low pass and high pass outputs (s, d). Now, the FB method achieves the DWT based on convolving filter with number of taps and samples of the input signal. Both low pass H (z) and high pass G (z) filters can be described in these equations (3.7) and (3.8):

$$H(z) = h_0 + h_{1z^{-1}} + h_{2z^{-2}} + \cdots + h_{Nz^{-N}} \quad (\mathbf{3.7})$$

$$G(z) = g_0 + g_{1z^{-1}} + g_{2z^{-2}} + \cdots + g_{Mz^{-M}} \quad (\mathbf{3.8})$$

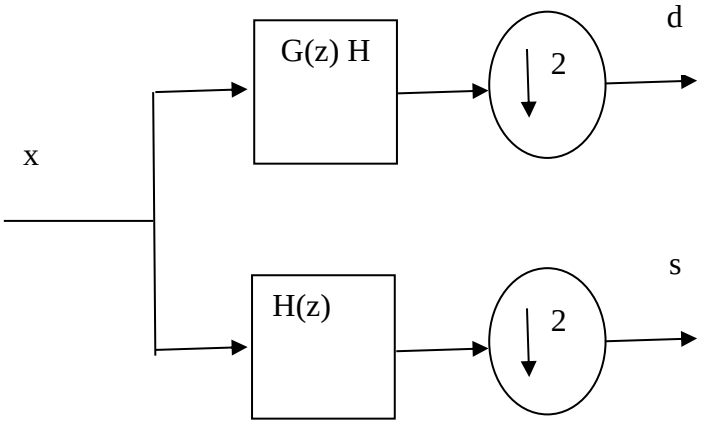


Figure 9: Filter Bank Block Diagram.

3.2.2 Lifting Scheme Method

The define Lifting Scheme (LS) method is a making and execution wavelets based on the time domain [3]. The Figure 10 illustrates the LS structure, where this structure has divided signal of sample into two parts, even and odd samples.

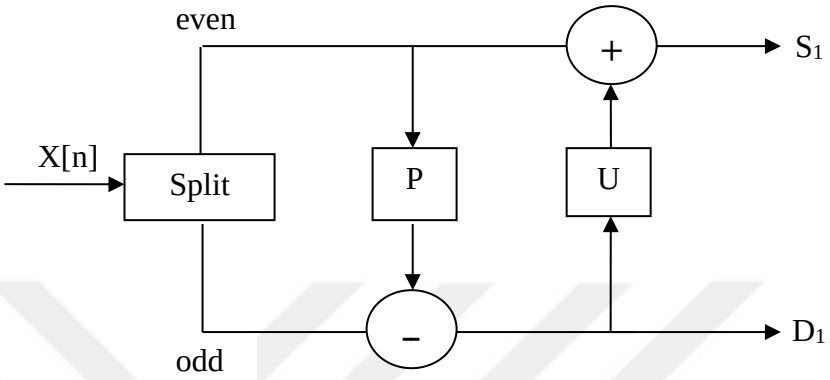


Figure 10: Lifting Scheme method.

Then P is a function, which used on even samples as a prediction function. This function used to predict the odd samples based on even samples. The P function produces the details coefficient (d). The U is a function applied on detail signal and gathering the result of details coefficient (d) with even samples and the output is approximation coefficients (s). The equations of both *d* and *s* function are defined in below.

$$d = X_{odd} - P(X_{even}) \quad (\mathbf{3.9})$$

$$s = X_{even} + U(d) \quad (\mathbf{3.10})$$

3.3 Support Vector Machine (SVM)

The SVM is a well-known capability of being universal approximates of any multivariate function to any chosen degree of exactness. Thus, they have specific attention for modeling the unknown, or partially known, highly nonlinear, complex systems, plants or processes.

The learning case set for SVMs is mentioned as follows: considering there is an unknown and nonlinear dependence (mapping, function) $y = f(x)$ within high-

dimensional space *x* as an input, and the output considered scalar numbers *y*. There is no data about the fundamental combined chance functions. Thus, one must achieve a distribution-allowed learning. The only data obtainable is to train data set D={(xi, yi)∈X×Y},i=1, l, where l attitudes of the number of the training data couples and is consequently like to the size of the trained dataset D. Frequently, yi is meant as di, where d stands for a wanted (goal) value. Henceforth, SVMs have their place in the supervised learning technique.

The Support Vector Machine (SVM) has been generally applied in real-life requests destined for mainly classification determinations. The SVM technique is founded on a physical risk minimization value. The SVM public compensations comprise minimization of experiential risk and existing over appropriate data; to find best margin hyper plane and convex quadratic procedure of problem; find a classifier with support vectors and the kernel function founded on training particularly in the nonlinear settings. The computational difficulty of SVM is the foremost problem as it needs a great amount of memory throughout the training stage. Similarly, lot mode nature of the SVM boundary the usage in real-life applications where training from unknown data is useful. Platt has assumed the one general explanation to large memory use where a progressive minimal optimization algorithm is suggested that uses two variables in iteration. This decreases the memory total for the present procedure in execution and more advancement in SVMs to accelerate the procedures that support less training time.

SVM is a learning system with high efficiency and is very accurately applied in data classification. The approach of SVM is to discover the maximum margin for two classes of data [32]. The procedure of SVM is described below.

- The data input to the SVM was separated into two classes in a two-dimensional plane [33]:

$$X = \{(x_{1i} y_{1i}, i = 1, \dots, N_1)\}, \{(x_{2j} y_{2j}, j = 1, \dots, N_2)\} \qquad (\mathbf{3.11})$$

Where X is the datasets containing two classes, x_{1i}determining the first class, and x_{2j} determining the second class $\underline{y_{1i}}$ defines the labels for the first class with labels 1, and y_{2j} defines the labels for the second class with labels 2.

The output data using a linear classifier for two classes is defined in Equation (3.14) [34].

$$y(x) = sign[w^T x + b] \quad (\mathbf{3.12})$$

Where *w* is the weight and *b* is the bias.

The *y(x)* results in the label for each class and is defined in equation (3.15) [35]:

$$\begin{cases} w^T x + b \geq +1, & if y(x) = +1 \\ w^T x + b \leq -1, & if y(x) = -1 \end{cases} \quad (\mathbf{3.13})$$

3.4 Iterative Closest Point (ICP)

The principle of the ICP algorithm is to search the corresponding closest points between the points found in a model and the points in the data based on an estimation of the transformation by finding and minimizing the distance error [36].

The procedure of the ICP is to take points of the first group (model) defined by *m*, and search the corresponding best closet points in the second group which is defined by *d*, as in Equation (3.11) below**.**

$$M = \{m_1, m, \dots, m_n\}, D = \{d_1, d_2, \dots, d_n\} \quad (\mathbf{3.14})$$

The next step of the ICP is to find the best minimum error for every point in the model to every point in the group. Equation (3.12) calculates the mean square error to find the minimum distance between a point in the model and all points in the data [37]:

$$e = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(d_{ix} - m_{ix})^2 + \left(d_{iy} - m_{iy}\right)^2} \quad (\mathbf{3.15})$$

Where $d_{ix}\ and\ d_{iy}$ are the points located in the data, $m_{ix}\ and\ m_{iy}$the points located in model, and *N* the number of points.

3.5 The Quality Performance Measurements

Measurement metrics were used to measure the quality of performance of most signals for recognition and classification.

We used the confusion matrix method to measure the quality of the proposed work which represents the numbers of total actual classes and predicted classes [38].The values in the confusion matrix table included True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN).

Each value of the confusion matrix is defined as follows:

1. True positive: the number of positive data which are positive found in the actual dataset.
2. True negative: the number of negative data which are negative found in the actual dataset.
3. False positive: the number of positive data which are negative found in the actual dataset.
4. False Negative: the number of negative data which are positive found in the actual dataset.

These values are presented in Table 3.

Table 3: Confusion Matrix.

	Positive predicted	Negative predicted
Positive actual	True positive	False negative
Negative actual	False negative	True negative

From the values of the confusion matrix, we can analyze the accuracy, sensitivity, specificity and precision as defined below:

ACCURACY: The accuracy of a classifier in given test data is the percentage of test data which are correctly classified by the classifier.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (\mathbf{3.16})$$

SENSITIVITY: The proportion of positive data that are correctly identified.

$$sensitivity = \frac{TP}{TP + FN} \quad (\mathbf{3.17})$$

SPECIFICITY: The proportion of negative data that are correctly identified.

$$specificity = \frac{TN}{FP + TN} \quad (\mathbf{3.18})$$

PRECISION: The number of true positives divided by the total number of elements labeled as belonging to the positive class.

$$precision = \frac{TP}{TP + FP} \quad (\mathbf{3.19})$$

CHAPTER FOUR

EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Overview of Proposed Work

The main goal of this research was to achieve high performance of abnormal ECG recorded signal classifications. In this proposed work, we used several methods and algorithms that were prepared during three phases (pre-processing, feature extraction, and post-processing).

Figure 11 shows the illustration of the proposed work.

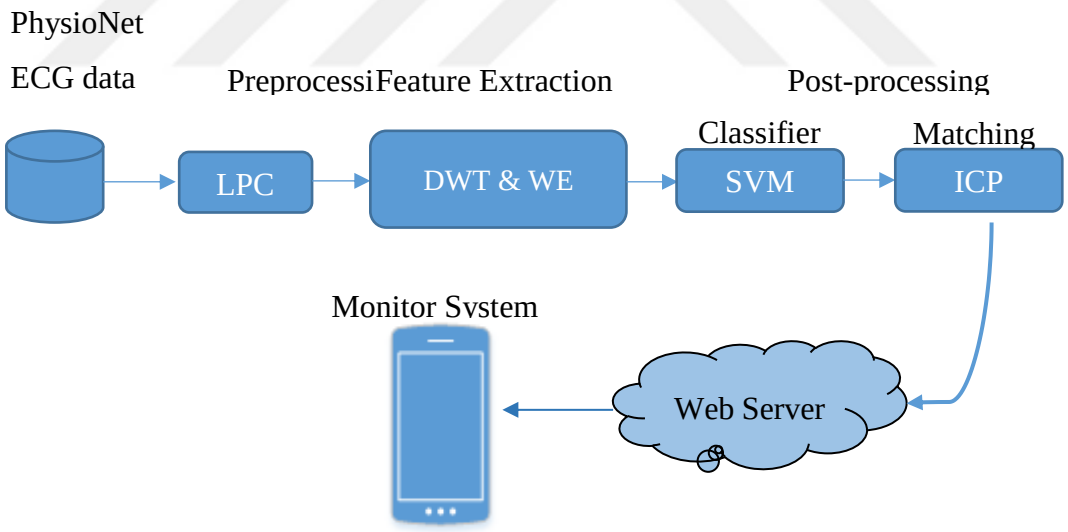


Figure 11: Proposed work phases.

The following paragraph describe in detail all the mentioned phases in the proposed work.

4.2 Preparing ECG Data

The datasets of ECG records used in this work were collected from a larger database found at PhysioNet.org. These datasets are categorized in 6 classes, each of which representing six diseases of heart. The ECG samples collected from PhysioNet are in the time domain and the period length trimmed for each sample is 60 s.

The total number of ECG samples on which we depended in our work is 120 samples, which is divided into 6 classes, each of which containing 20 samples. We selected 6 samples from each class as the test samples, and the remaining 14 samples as the training samples. Table 4 below shown plotted ECG in time-series for 5 select samples of each class.

Table 4: Samples of ECG time domain.

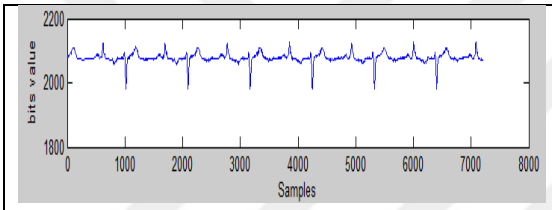
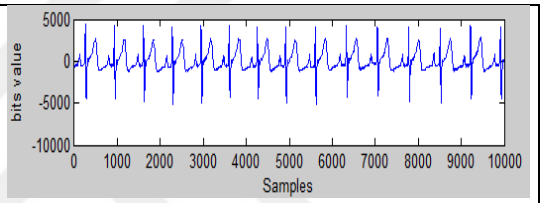
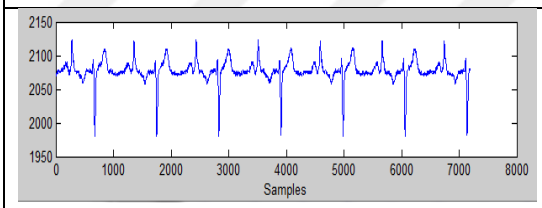
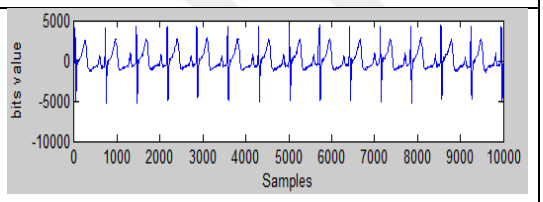
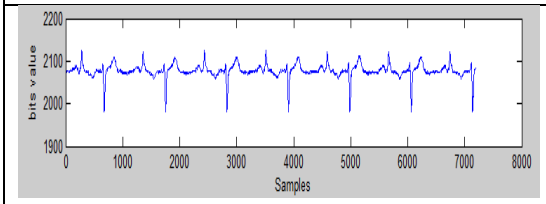
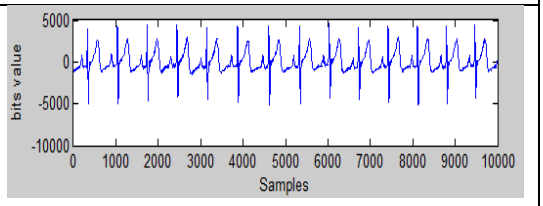
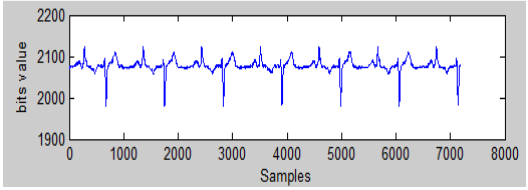
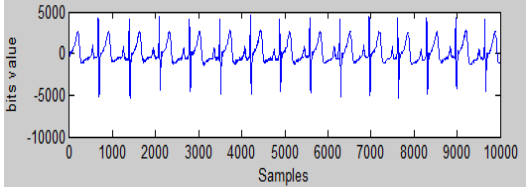
 Class (1) sample (1)	 Class (4) sample (1)
 Class (1) sample (2)	 Class (4) sample (2)
 Class (1) sample (3)	 Class (4) sample (3)
 Class (1) sample (4)	 Class (4) sample (4)

Table (continuation): Samples of ECG time domain.

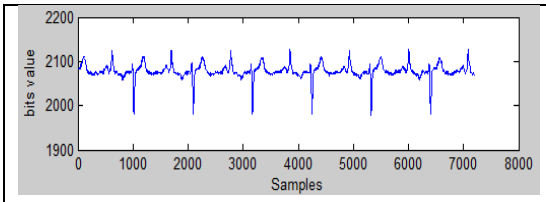
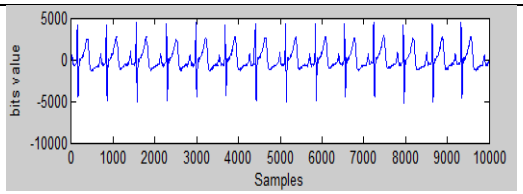
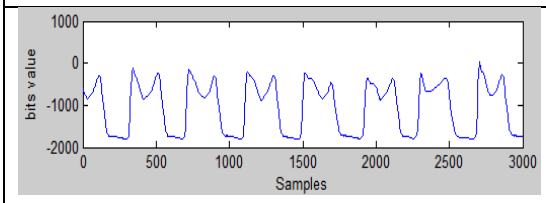
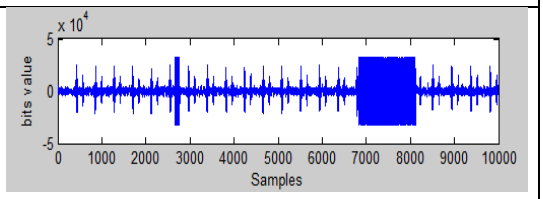
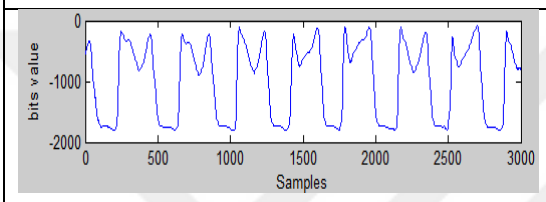
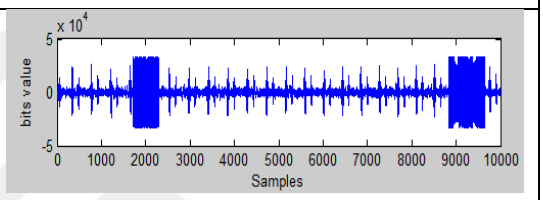
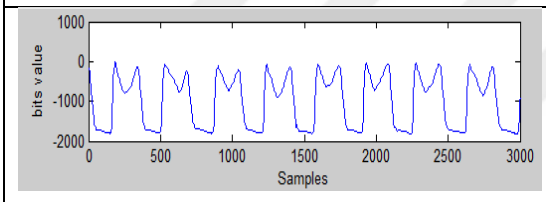
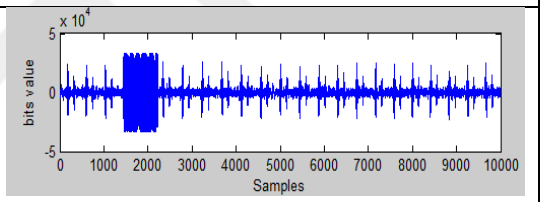
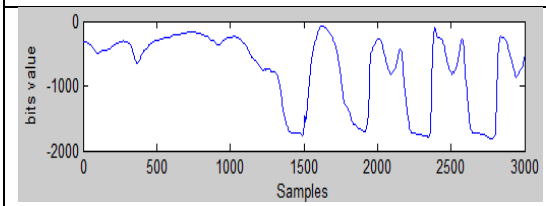
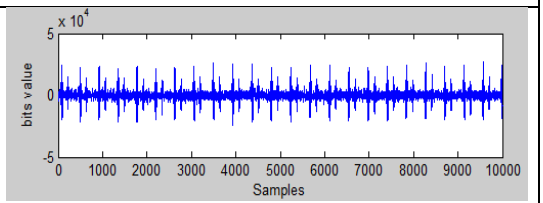
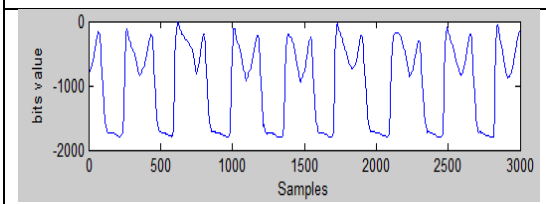
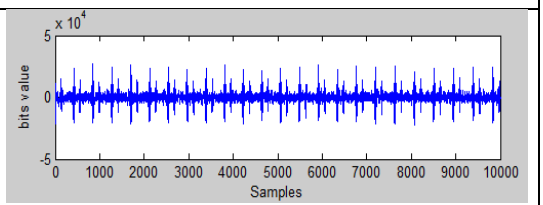
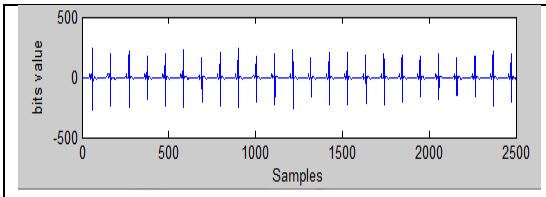
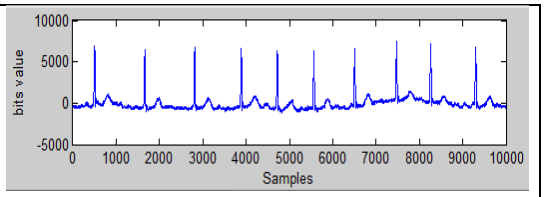
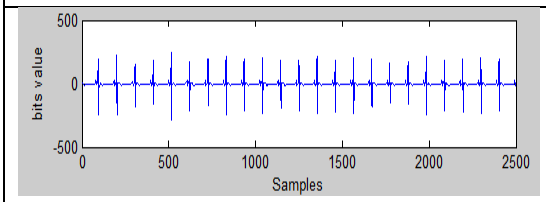
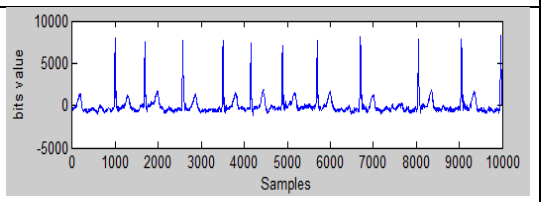
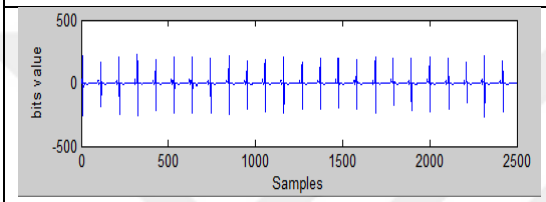
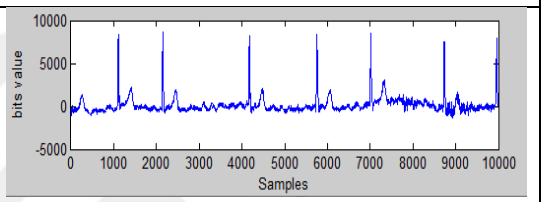
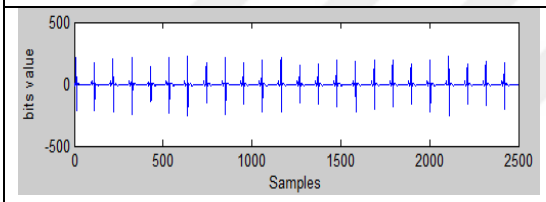
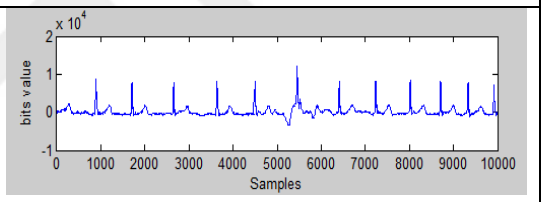
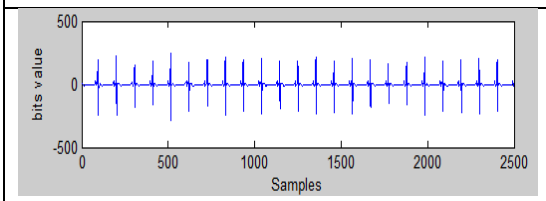
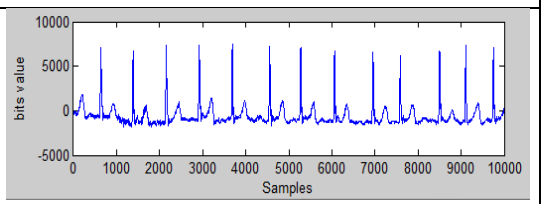
	
Class (1) sample (5)	Class (4) sample (5)
	
Class (2) sample (1)	Class (5) sample (1)
	
Class (2) sample (2)	Class (5) sample (2)
	
Class (2) sample (3)	Class (5) sample (3)
	
Class (2) sample (4)	Class (5) sample (4)
	
Class (2) sample (5)	Class (5) sample (5)

Table (continuation): Samples of ECG time domain.

	
Class (3) sample (1)	Class (6) sample (1)
	
Class (3) sample (2)	Class (6) sample (2)
	
Class (3) sample (3)	Class (6) sample (3)
	
Class (3) sample (4)	Class (6) sample (4)
	
Class (3) sample (5)	Class (6) sample (5)

4.3 Pre-processing the ECG Signal

Most ECG signals contain noise in their signals; additionally, in various cases (normal and abnormal), the magnitude of ECGs can vary. Moreover, ECG signals have various sampling periods which come from different of cases of heartbeats. For these reasons, a preprocessing phase is important to reduce noise and normalize the variability of magnitudes and sampling periods.

In the proposed work, Linear Predict Coding (LPC) was applied in the preprocessing phase to prepare the ECG data for the feature vector, and regression of all samples of the ECG signals of the same length of sampling period with normalization of the magnitude of the signal to be laid between –1 and +1.

Table 5 below shows the plotted samples during preprocessing using LPC and a selection of the same 5 samples in Table 4 as inputs to LPC.

Table 5: Samples Filtered using LPC.

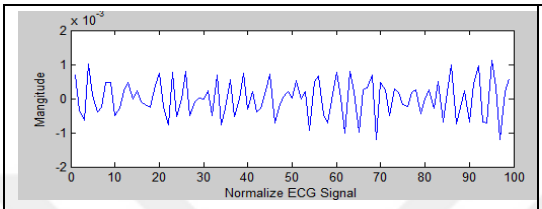
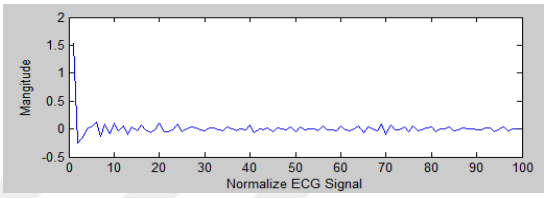
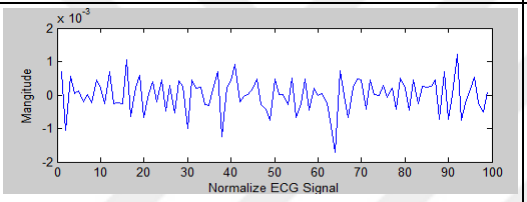
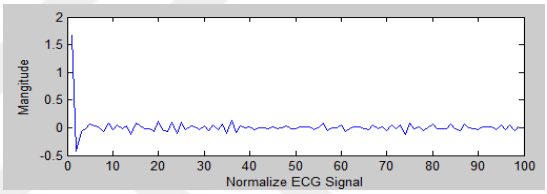
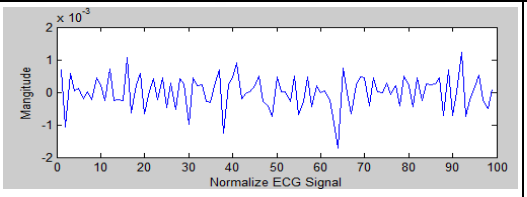
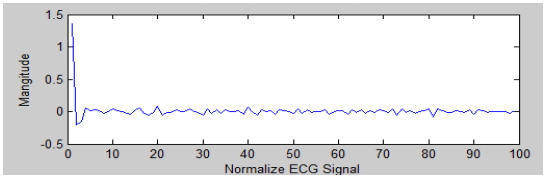
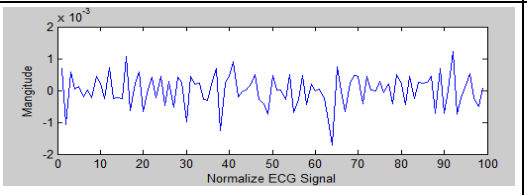
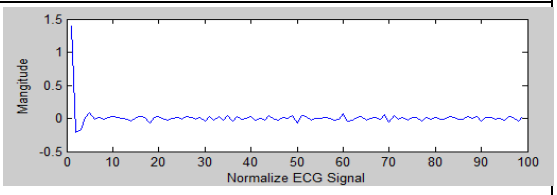
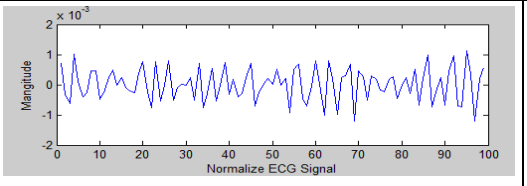
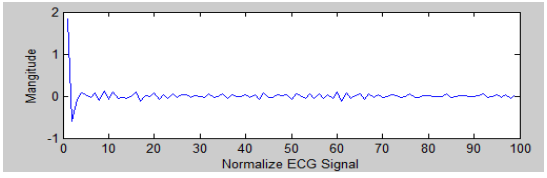
 Class (1) sample (1)	 Class (4) sample (1)
 Class (1) sample (2)	 Class (4) sample (2)
 Class (1) sample (3)	 Class (4) sample (3)
 Class (1) sample (4)	 Class (4) sample (4)
 Class (1) sample (5)	 Class (4) sample (5)

Table 5 (continuation): Samples Filtered using LPC.

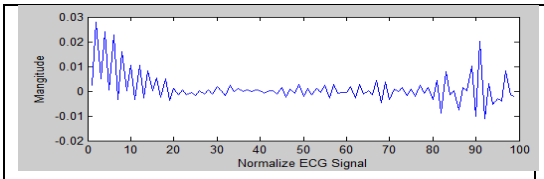
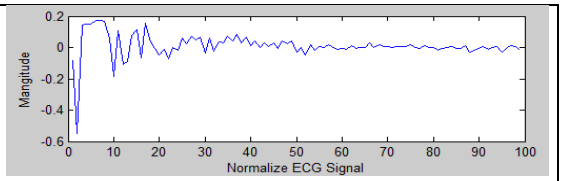
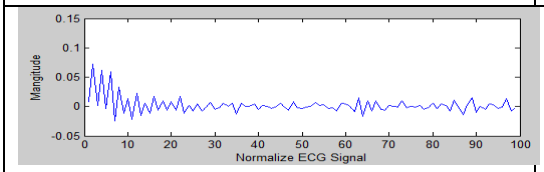
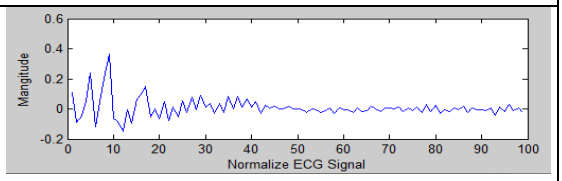
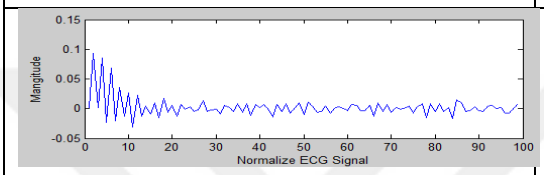
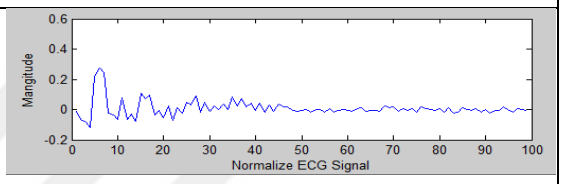
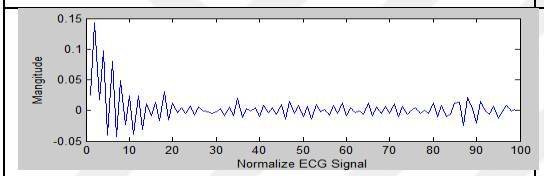
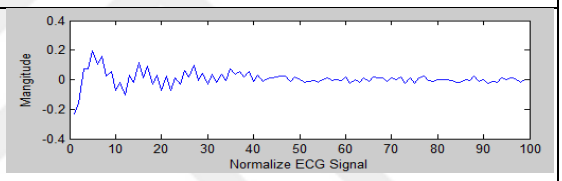
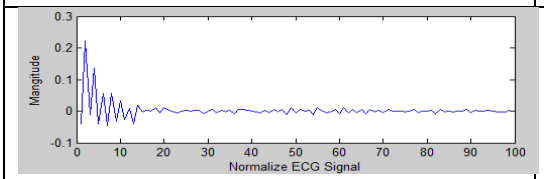
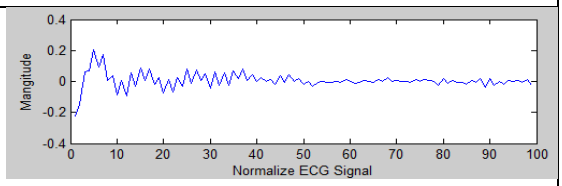
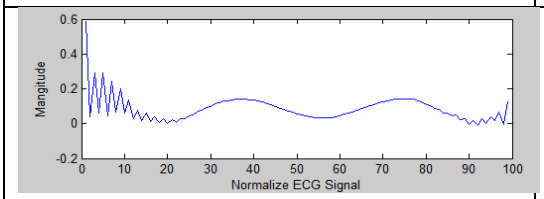
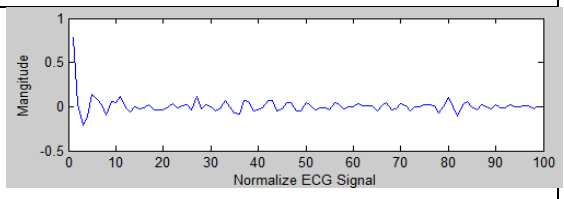
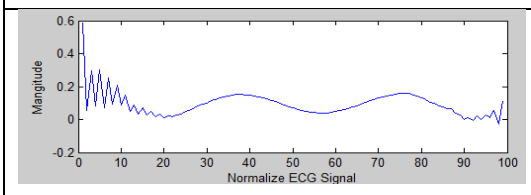
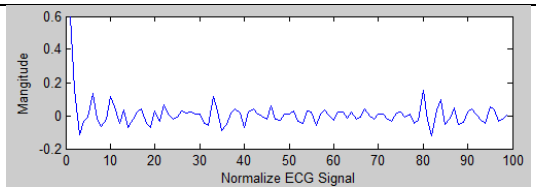
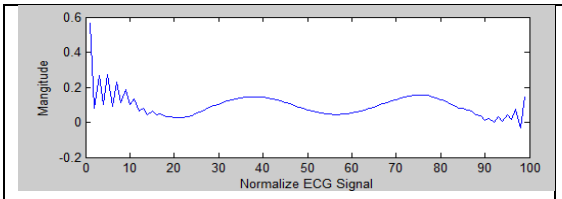
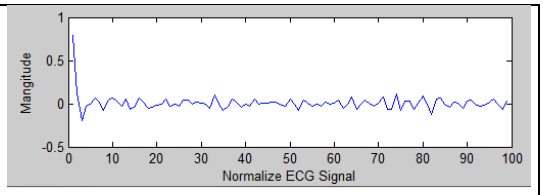
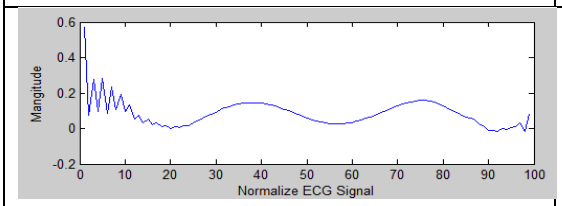
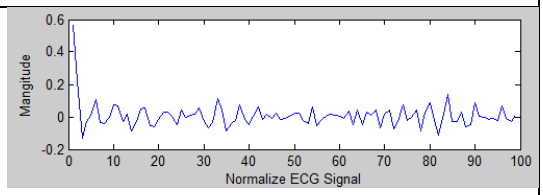
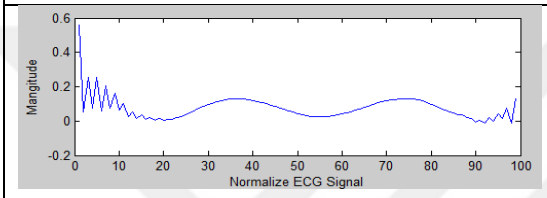
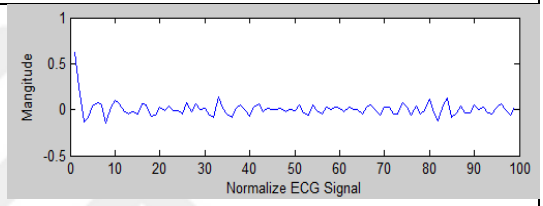
	
Class (2) sample (1)	Class (5) sample (1)
	
Class (2) sample (2)	Class (5) sample (2)
	
Class (2) sample (3)	Class (5) sample (3)
	
Class (2) sample (4)	Class (5) sample (4)
	
Class (2) sample (5)	Class (5) sample (5)
	
Class (3) sample (1)	Class (6) sample (1)
	
Class (3) sample (2)	Class (6) sample (2)

Table 5 (continuation): Samples Filtered using LPC.

 Class (3) sample (3)	 Class (6) sample (3)
 Class (3) sample (4)	 Class (6) sample (4)
 Class (3) sample (5)	 Class (6) sample (5)

4.4 Feature Extraction using DWT and Wavelet Energy

In the feature extraction phase of our proposed work, we applied a one-dimension wavelet transform to reduce the dimension of the samples for better classification in the following phase.

In this work, the DB4 wavelet was chosen with 8 level decomposition and the output of the wavelet was a set of two coefficients, namely approximation and details coefficients. The feature vector in this case was combined by taking the last approximation coefficient and 8 details coefficients.

Table 6 contains the 5 samples of each class, the samples of which represent the approximate and details coefficients.

Table 6: Approximation and details coefficients output of DWT.

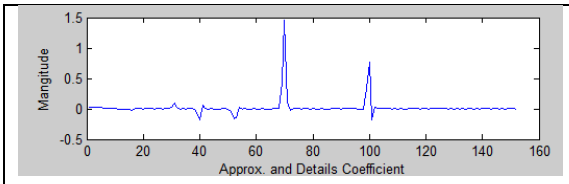
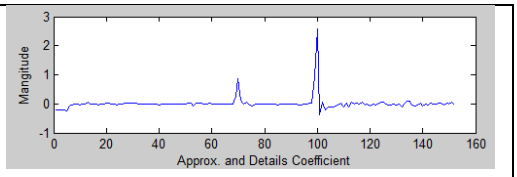
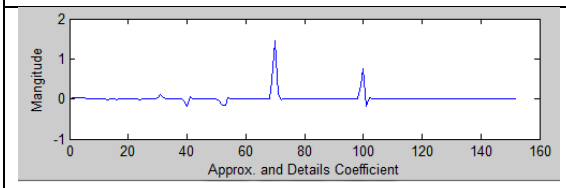
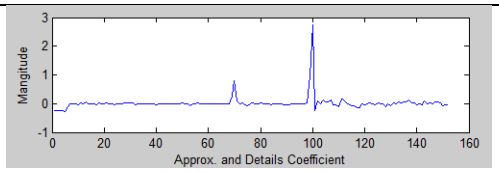
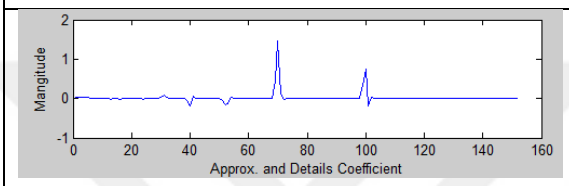
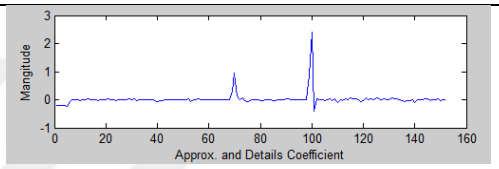
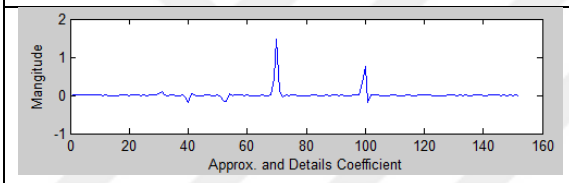
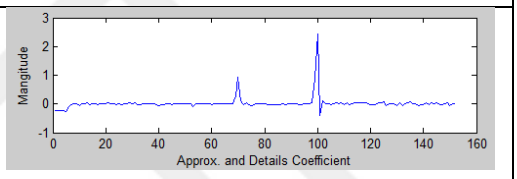
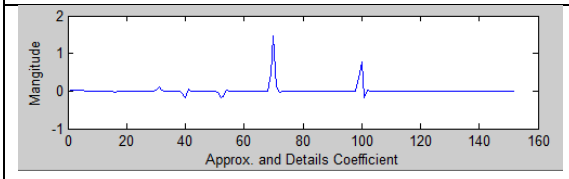
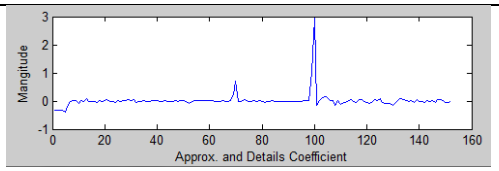
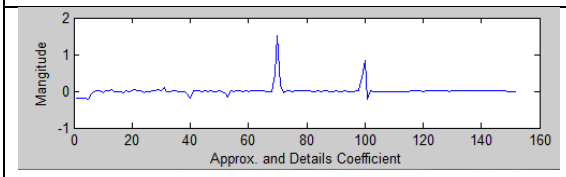
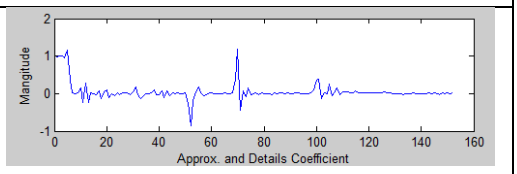
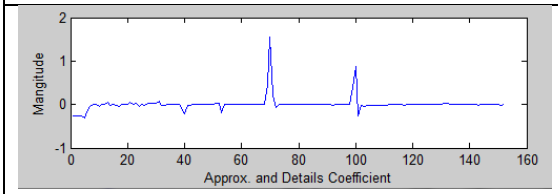
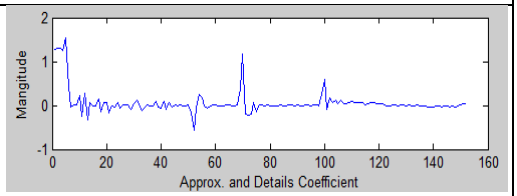
 Class (1) sample (1)	 Class (4) sample (1)
 Class (1) sample (2)	 Class (4) sample (2)
 Class (1) sample (3)	 Class (4) sample (3)
 Class (1) sample (4)	 Class (4) sample (4)
 Class (1) sample (5)	 Class (4) sample (5)
 Class (2) sample (1)	 Class (5) sample (1)
 Class (2) sample (2)	 Class (5) sample (2)

Table 6 (continuation): Approximation and details coefficients output of DWT.

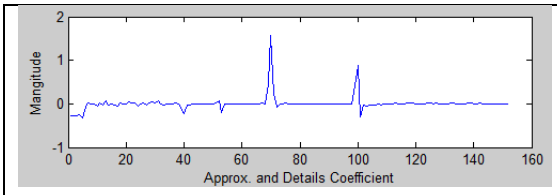
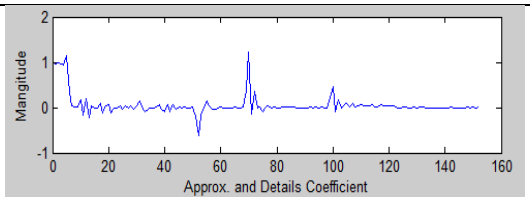
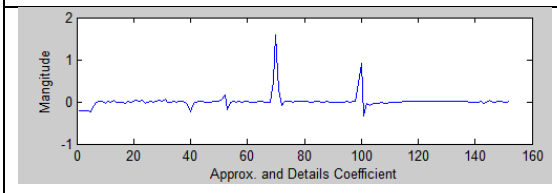
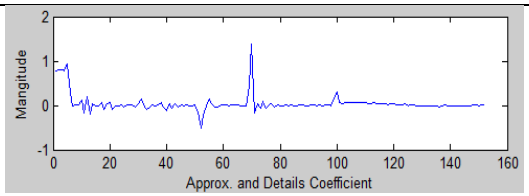
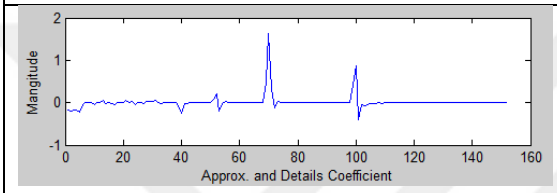
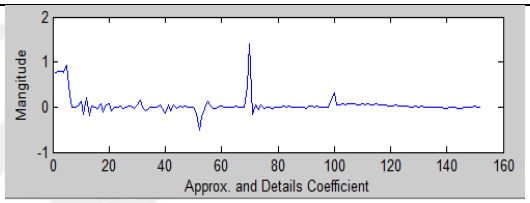
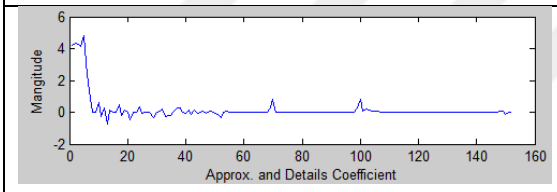
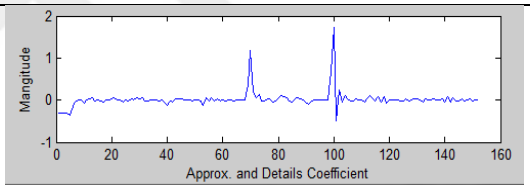
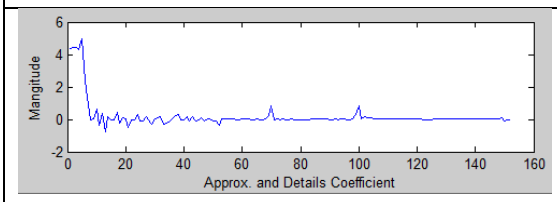
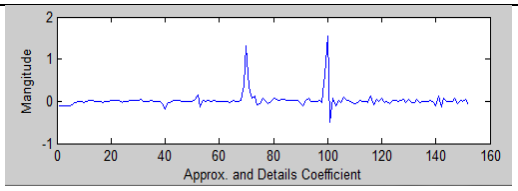
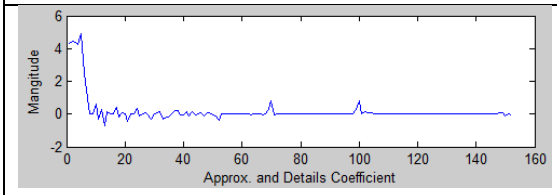
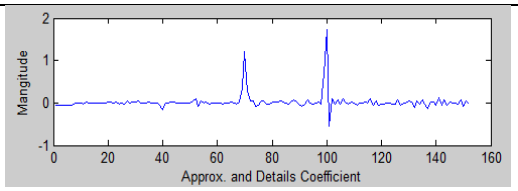
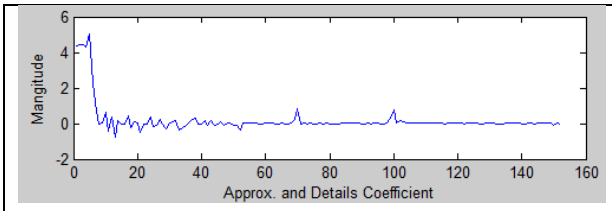
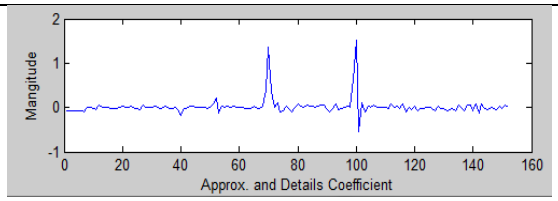
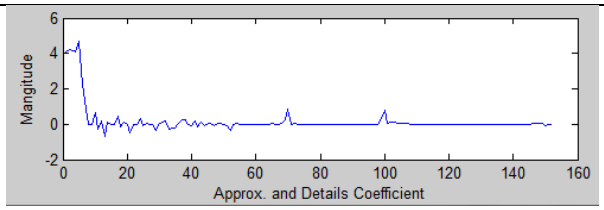
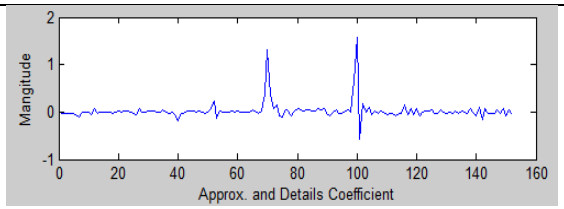
 Class (2) sample (3)	 Class (5) sample (3)
 Class (2) sample (4)	 Class (5) sample (4)
 Class (2) sample (5)	 Class (5) sample (5)
 Class (3) sample (1)	 Class (6) sample (1)
 Class (3) sample (2)	 Class (6) sample (2)
 Class (3) sample (3)	 Class (6) sample (3)

Table 6 (continuation): Approximation and details coefficients output of DWT.

 Class (3) sample (4)	 Class (6) sample (4)
 Class (3) sample (5)	 Class (6) sample (5)

Another method applied during the feature extraction phase is wavelet energy. The basic wavelet is carried out by calculating the energy of each decomposition level. The results of the wavelet energy in our work show fusion in one vector, as presented instable.

Table 7: Feature vector by using wavelet energy.

Class and sample no.	Feature Vector
Class (1) sample (1)	22.8257 73.9413 1.6304 1.1511 0.3499 0.0077 0.0181 0.0055
Class (1) sample (2)	22.8095 73.9567 1.6285 1.1520 0.3496 0.0075 0.0179 0.0061
Class (1) sample (3)	22.8232 73.9302 1.6152 1.1498 0.3483 0.0066 0.0168 0.0052
Class (1) sample (4)	22.8185 73.9446 1.6190 1.1482 0.3453 0.0068 0.0179 0.0061
Class (1) sample (5)	22.8257 73.9413 1.6304 1.1511 0.3499 0.0077 0.0181 0.0055
Class (2) sample (1)	23.4204 68.6347 0.9195 1.1446 0.2368 0.0910 0.0882 0.0903
Class (2) sample (2)	23.9300 64.3867 0.9282 1.1938 0.1796 0.1385 0.1331 0.1487
Class (2) sample (3)	23.8282 63.8785 1.0430 1.2528 0.1654 0.1552 0.1514 0.2354
Class (2) sample (4)	26.4457 64.8592 1.4991 1.3897 0.1162 0.0992 0.0991 0.1024
Class (2) sample (5)	24.3569 67.5610 1.9485 1.5453 0.1054 0.1019 0.0886 0.0911
Class (3) sample (1)	0.8123 0.7265 0.1425 0.0846 0.3269 0.2025 0.4330 1.0079
Class (3) sample (2)	0.7409 0.6806 0.1265 0.0827 0.3017 0.2148 0.4277 1.1191

Table 7 (continuation): Feature vector by using wavelet energy.

Class and sample no.	Feature Vector							
Class (3) sample (3)	0.6902	0.6936	0.1301	0.0813	0.2776	0.1897	0.4053	0.9924
Class (3) sample (4)	0.6872	0.6882	0.1292	0.0891	0.3498	0.2562	0.4776	1.2723
Class (3) sample (5)	0.7792	0.7810	0.1599	0.0907	0.3349	0.2240	0.4622	1.0276
Class (4) sample (1)	87.4512	9.6505	0.0670	0.0436	0.0203	0.0287	0.0301	0.0460
Class (4) sample (2)	89.0501	7.3577	0.0500	0.0235	0.0222	0.0287	0.0364	0.0667
Class (4) sample (3)	84.0618	12.7543	0.1114	0.0879	0.0245	0.0295	0.0361	0.0631
Class (4) sample (4)	89.1912	5.0079	0.1257	0.0125	0.0314	0.0447	0.0543	0.1067
Class (4) sample (5)	84.4096	11.6411	0.0902	0.0696	0.0302	0.0424	0.0411	0.0610
Class (5) sample (1)	4.8080	20.4549	10.0879	0.2921	0.7260	0.0558	0.5406	2.2823
Class (5) sample (2)	4.5140	13.7241	3.5928	0.2291	0.3549	0.1439	0.6733	2.4048
Class (5) sample (3)	4.4774	22.5857	6.0976	0.3347	0.5407	0.0907	0.5171	1.9639
Class (5) sample (4)	2.8352	33.7839	5.5708	0.4824	0.6033	0.0520	0.4871	2.0500
Class (5) sample (5)	3.2612	34.1929	5.5704	0.4973	0.6130	0.0526	0.4784	2.0231
Class (6) sample (1)	62.5995	27.2025	0.4204	0.3481	0.0882	0.1213	0.1086	0.1489
Class (6) sample (2)	58.5613	38.6237	0.8934	0.7022	0.0432	0.0510	0.0283	0.0242
Class (6) sample (3)	68.6307	30.0775	0.4100	0.4486	0.0470	0.0552	0.0145	0.0314
Class (6) sample (4)	57.0713	39.7795	1.2576	0.7730	0.0612	0.1560	0.0333	0.0913
Class (6) sample (5)	1.4505	7.6107	1.2891	0.7640	0.0681	0.2208	0.0337	0.1522

4.5 Disease Classification using SVM

In the proposed work, we have proposed an SVM classifier to classify the classes of ECG to 6 diseases. The basics of SVM are to classify two sets of data by finding the optimal separation line between them. Each of set has a unique label.

Table 8 below presents some of the results of the SVM classifier by using the one-versus-rest method by taking one sample of two classes and finding the best separation line.

Table 8: Some of the results of the SVM classifier.

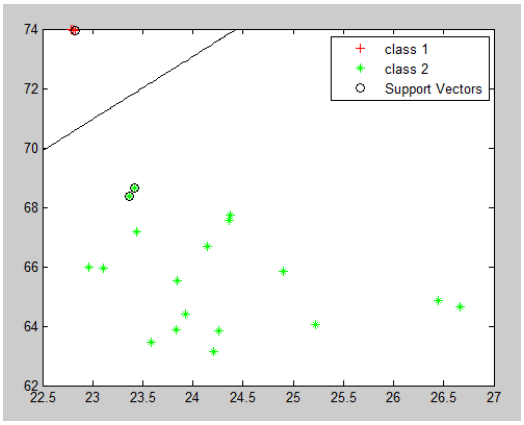
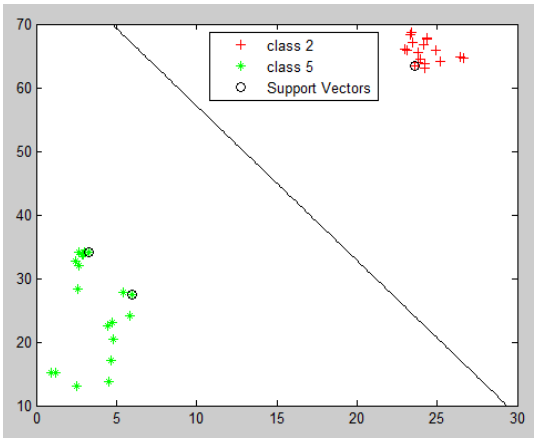
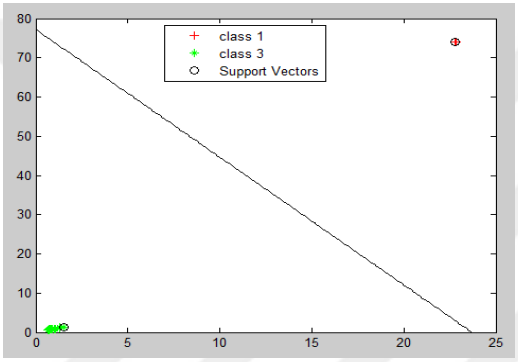
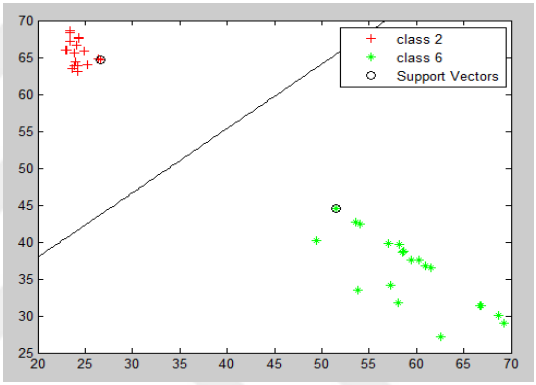
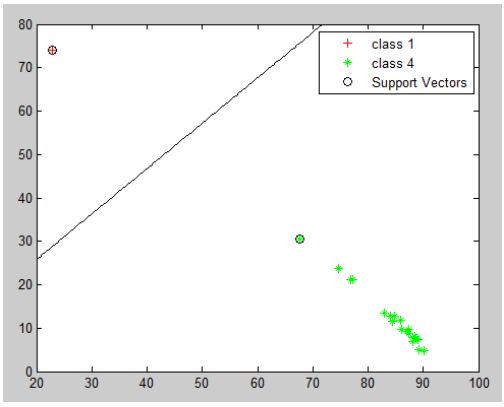
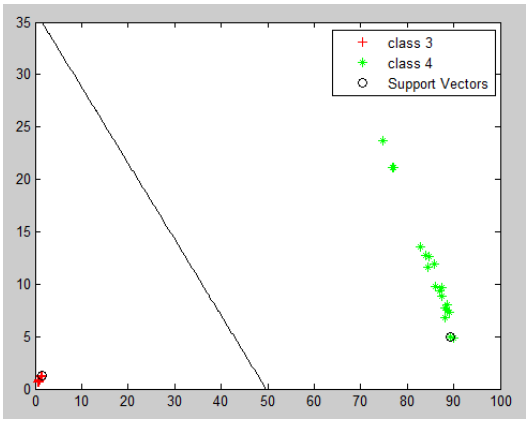
 Class (1) and Class (2)	 Class (2) sample (5)
 Class (1) and Class (3)	 Class (2) and Class (6)
 Class (1) and Class (4)	 Class (3) and Class (4)

Table 8 (continuation): Some of the results of the SVM classifier.

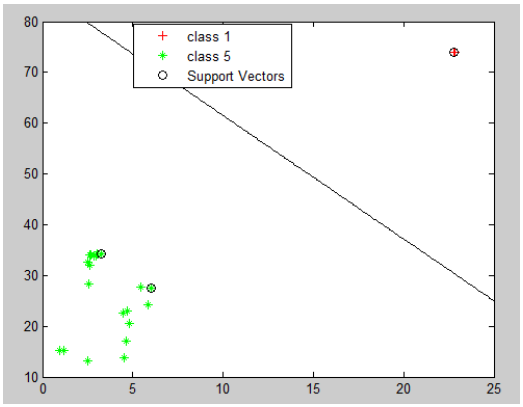
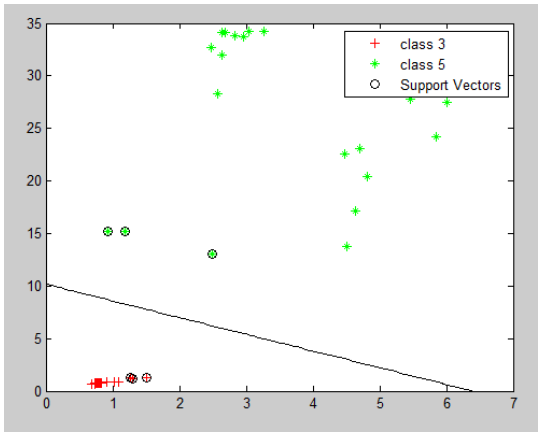
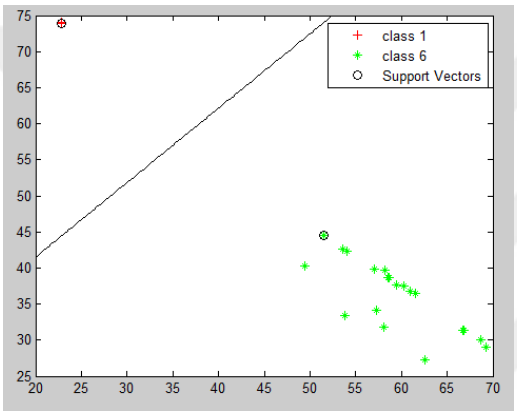
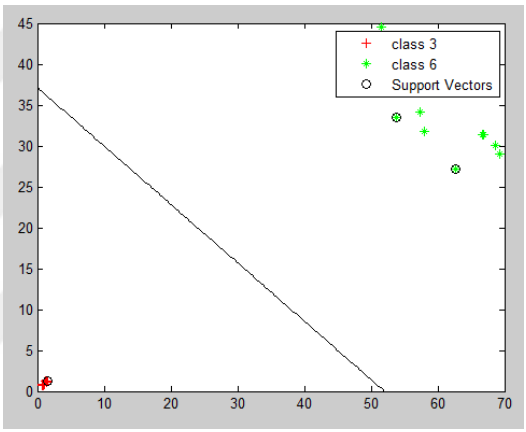
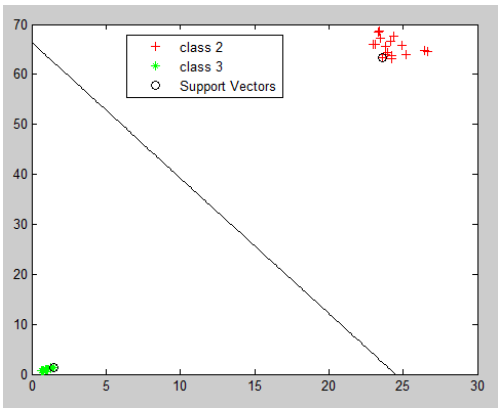
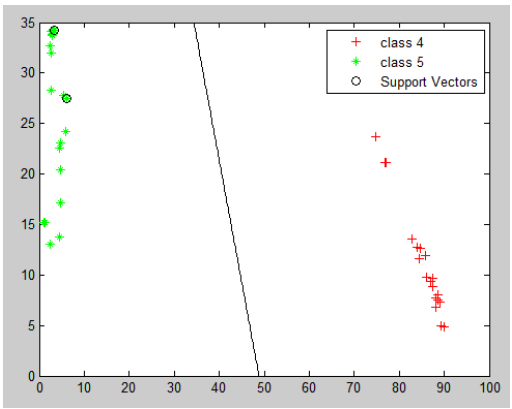
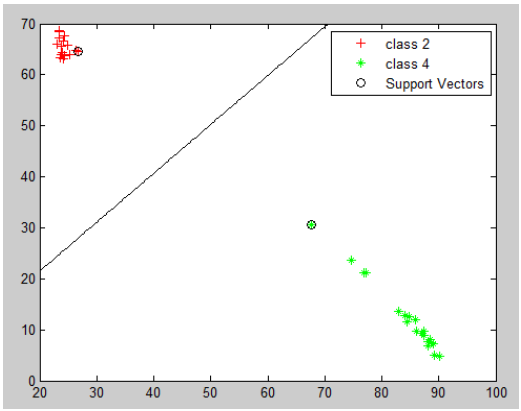
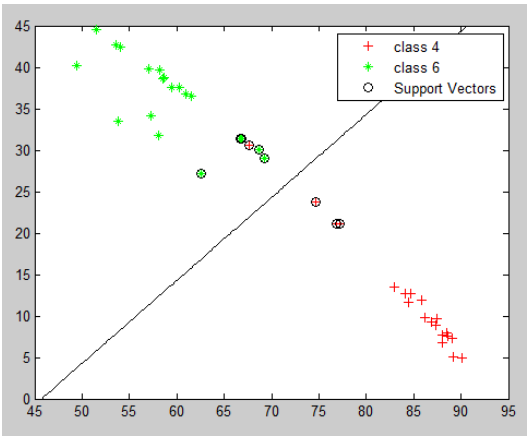
 Class (1) and Class (5)	 Class (3) and Class (5)
 Class (1) and Class (6)	 Class (3) and Class (6)
 Class (2) and Class (3)	 Class (4) and Class (5)

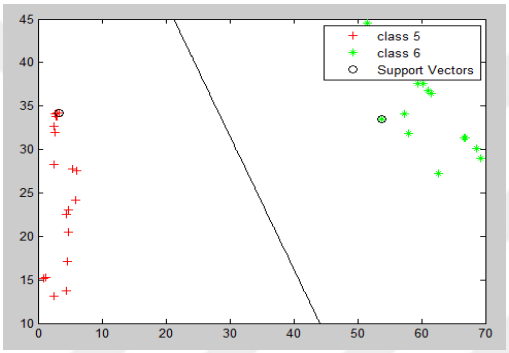
Table 8 (continuation): Some of the results of the SVM classifier.



Class (2) and Class (4)



Class (4) and Class (6)



Class (5) and Class (6)

The other step of the SVM classifier is to find the corresponding training sample matching with test one and in this case, the SVM returns a label belonging to that class. Table 9 presents the matching of test samples to the training datasets.

Table 9: classification results of SVM.

Label of Test Sample	Classes Index Range	Label of Sample matched	Class
1	Class (1) Range (1→14)	1	Class (1)
2		1	Class (1)
3		1	Class (1)
4		1	Class (1)
5		1	Class (1)
6		1	Class (1)

Table 9 (continuation): classification results of SVM.

Label of Test Sample	Classes Index Range	Label of Sample matched	Class
7	Class (2) Range (15→28)	15	Class (2)
8		15	Class (2)
9		15	Class (2)
10		15	Class (2)
11		15	Class (2)
12		15	Class (2)
13	Class (3) Range (29→42)	30	Class (3)
14		29	Class (3)
15		29	Class (3)
16		29	Class (3)
17		29	Class (3)
18		30	Class (3)
19	Class (4) Range (43→56)	43	Class (4)
20		15	Class (2)
21		43	Class (4)
22		43	Class (4)
23		43	Class (4)
24		46	Class (4)
25	Class (5) Range (57→70)	58	Class (5)
26		61	Class (5)
27		1	Class (1)
28		59	Class (5)
29		59	Class (5)
30		59	Class (5)
31	Class (6) Range (71→84)	71	Class (6)
32		71	Class (6)
33		71	Class (6)
34		71	Class (6)
35		71	Class (6)
36		71	Class (6)

4.6 Selecting Best Matching using ICP

In the previous phase, the result of the SVM classifier returns a class name (label) when the feature vector of the test sample is laid in one correspondence space, which represents 6 datasets.

To find which specific sample of training is matched to the sample of the test needed to use the matching method, weave proposed in this phase the Iterative Closest Point (ICP), which returns a minimum error for the best matching of the test corresponding to the training sample.

Table 10 presents the average of each minimum error belonging to each element of the feature vector.

Table 10: Average for each minimum error of feature vector element.

Feature vector element	Average
Element (1)	0.0462
Element (2)	0.1905
Element (3)	0.0375
Element (4)	0.0017
Element (5)	0.0000
Element (6)	0.0013
Element (7)	0.0002
Element (8)	0.0005

4.7 Design Remount Monitor System

The implementation of the diagnostic result to transmits the result to the remote monitor system (smart phone) over web server to reduce the healthcare time.

The application of proposed remote monitor system has designed in Android OS to display the patient's heart disease by considering there has wearable or portable ECG acquisition.

The main goal of design remote monitor system is to alert the closest healthcare center to patient's location in case of detect critical heart condition.

4.8 Evaluation Metrics

The method was used to evaluate the performance of our proposed work by using the confusion matrix to measure sensitivity, specificity, and accuracy.

Table 11 presents evaluation metrics for the same datasets without the implementation of LPC in the preprocessing phase. However, Table 12 presents evaluation metrics depending on the values of the confusion matrix for two selected datasets (3 and 6 diseases) with the implementation of LPC.

Table 11: Average accuracy, sensitivity, and specificity results without implementation of LPC.

	Using DWT &SVM			
	Accuracy	Sensitivity	Specificity	Precision
3 diseases	100%	100%	100%	100%
6 diseases	92.5%	77.7%	95.1%	77.7%

Table 12: Average accuracy, sensitivity, and specificity results with implementation of LPC.

	Using LPC, DWT &SVM			
	Accuracy	Sensitivity	Specificity	Precision
3 diseases	100%	100%	100%	100%
6 diseases	98.14%	94.4%	98.4%	94.4%

In comparison with the performance of classifiers mentioned previously in the literature review of Chapter2, the accuracy we have reached without LPC approach to the accuracy of previous studies but when we used LPC accuracy increased even in case of increased the number of diseases we were able to obtain high accuracy.

CHAPTER FIVE

CONCLUSION AND FUTURE WORKS

The technique used in this work was to convert recorded ECG signals from the frequency domain to the time domain to be processed digitally with several methods to achieve high quality of performance for heart disease classification.

The data of the ECG recorded signals was collected from the PhysioNet database, and the total of the ECG datasets numbered 120 samples for 6 classes, 20 samples for each class these samples were divided into (14×6) as training samples, and (6×6) as test samples.

The preprocessing method used in this work was Linear Predictive Coding (LPC) to normalize all ECG signals to the same amplitude –1 → +1 and the number of the LPC-order selected by experiments was high and equaled the length of the ECG signal in order to capture all the information from the ECG to achieve high quality.

ECG features were extracted using DWT DB4 with 8 levels and Wavelet Energy to calculate 8 vectors for each decomposition level followed by fusion of these vectors to be prepared for the post-processing phase.

After preparing the feature vector for all samples of the ECG, we selected (18) samples from 20 from each class as the learning data used for the training model of the SVM, and the remaining samples as test data were used for the test model. In addition, each training dataset has a unique label which was returned during the classifying step.

The accuracy of the three test samples 3 diseases corresponded to complete training datasets, and by applying the LPC method during the preprocessing phase, it was 100%. Other experiments were performed for six test samples 6 diseases corresponding to complete training datasets and the result gave an accuracy of 98.14% when applying the LPC method, and without the LPC method it was 92.59%. Finally,

the classification result has considered as important information and uploaded to web server to be retrieve by portable monitor (smart phone) based on Android which is an alarm message is sent to an emergency service center to provide immediate assistance to patients.

Future work will be on Real-Time analysis for abnormal ECG classification by implementing our algorithm in a field-programmable gate array (FPGA).

Another possible future work is designing a system for calling the nearest medical care for abnormal heart cases using GPS and maps.

REFERENCES

1. Bayés de Luna A, Fiol-Sala M. The Surface Electrocardiography in Ischemic Heart Disease. Oxford, UK: Blackwell, 2008

2. Colombo J, Arora R, DePace NL, Vinik AI. Clinical Autonomic Dysfunction. Springer press, p.20, 2015.

3. JM David, GM Niall. Oxford Companion to the Body: Electrocardiogram. Yale University Guide to Medical Tests, 2010.

4. Alfonso, V.X.; Wieben, O.; Tompkins, W.J.; Nguyen, T.Q.; "Filter bank-based ECG beat classification"; Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE; 1997.

5. Luna, A. B; “Clinical electrocardiography: a textbook”. New York: Futura Publishing; 1993.

6. Roger K. and Michael P.; "Fetal heart rate monitoring"; Lippincott W. Press, 2012.

7. NHLBI; "Types of Arrhythmia", 2011.

8. Peter K. and Gerald V.; "Artial Fibrillation"; Marcel Dekker press, 2005.

9. Sanjay D. and MD F.; "Obstetric Anesthesia handbook"; Springer press, 2006.

10. David B. and Martin R.; "Molecular and Genetic Basic of Renal Disease, a companion to Brenner and Rector's the kidney"; Saunders Elsever; 2008.

11. Mark P. Mattson; "Pathogenesis of Neurodegenerative Disorders"; Humana Press; 2001.

12. Chinmayee D. and Pradyut K.; "Continuous Wavelet Transformation and Linear Regression"; International Conference on Signal Processing and Communications (SPCOM), 2016.

13. Zheng Wei, Li Xiaolong, Wei Xueyun and Liu Hongxing; "Foetal ECG extraction by support vector regression"; IEEE Journal Vol.52 Issue7, 2016.

14. Futoshi T., Takuya A.; "Individual Identification with High Frequency ECG: Preprocessing and Classification by Neural Network"; 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30-September 3, 2011.

15. Zhe Li, Jun Ni, Xin Gu; "A Denoising Framework for ECG Signal Preprocessing"; IEEE, Sixth International Conference on Internet Computing for Science and Engineering, 2012.

16. Anindya S. Paul, Eric A. Wan, and Alex T. Nelson; "Noise Reduction for Heart Sounds Using a Modified Minimum-Mean Squared Error Estimator with ECG Gating"; IEEE, Proceedings of the 28th IEEEEMBS Annual International Conference New York City, USA, 2006.

17. Sani Saminu, Nalan Özkurt and Ibrahim Abdullahi Karaye; "Wavelet Feature Extraction for ECG Beat Classification"; IEEE 6th International Conference on Adaptive Science & Technology (ICAST); 2014.

18. Swati Banerjee, M. Mitra; "A Cross Wavelet Transform Based Approach for ECG Feature Extraction and Classification without Denoising"; International Conference on Control, Instrumentation, Energy & Communication; IEEE; 2014.

19. Nurul A., Mohd A., Norlaili M.; "ECG Features Extraction using Second-Order Dynamic System and Regeneration using Hybrid Recurrent Network"; IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016.

20. Jagadeeswara Rao Annam and Bapi Raju Surampudi; "Inter-patient Heart-beat Classification using complete ECG beat Time series by alignment of R-peaks using SVM and Decision Rule"; International Conference on Signal and Information Processing (Icon SIP); IEEE; 2016.

21. Czarina Isabelle M. Cruz, Jastine P. Marasigan; "A Comparative Study between DWT-ANFIS and DWT-SVM in ECG Classification"; IEEE Region 10 Conference (TENCON); 2016.

22. Latifah Aljafar, Turky N. Alotaiby, Rand R. Al-Yami1, Saleh A. Alshebeili, Jalila Zouhair; "Classification of ECG Signals of Normal and Abnormal Subjects Using Common Spatial Pattern"; International Conference on Electronic Devices, Systems and Applications (ICEDSA); IEEE; 2016.

23. Stalin Subbiah, Rajkumar Patro, Subbuthai P; "Feature Extraction and Classification for ECG Signal Processing based on Artificial Neural Network and Machine Learning Approach"; International Conference on Inter Disciplinary Research in Engineering and Technology [ICIDRET]; 2015.

24. Irshad Khalil, Adnan Khalil; "Classification of ECG Signals Using Legendre Moments"; International Journal of Bioinformatics and Biomedical Engineering; AIS; 2015.

25. Manab Kumar Das and Samit Ari. ECG Beats Classification Using Mixture of Features. Hindawi Publishing Corporation. International Scholarly Research Notices. Volume 2014, Article ID 178436, 12 pages

26. Mohamed Moustafa Azmy Gad; "Classification of Mental Tasks using Support Vector Machine Based on Linear Predictive Coding and New Mother Wavelet Transform"; IEEE; 2015.

27. Jan W. Gooch; "Encyclopedic Dictionary of Polymers"; Springer Science LLC; 2011.

28. Thomas A. Severini; "Analytic Method in Sports"; CRC Press; 2015.

29. Chris Chtfield; "The Analysis of Time Series, an Introduction"; Sixth Edition; Chapman and Hall CRC; 2016.

30. Lawrence R. and Ronald W.; "Introduction to Digital Speech Processing", NOW published, the essence of knowledge, 2007.

31. Nevio Benvenuto and Giovanni Cherubini; "Algorithms for Communications Systems and their Applications"; Wiley Press; 2003.

32. Shigeo Abe; "Support Vector Machine for Pattern Classification"; Springer Press; 2005.

33. Neha M. and Suendra G.; "Maximal Margin Multi-classifier based on SVM Hyper parameter Turning"; IEEE International Conference on Computer, Communication and Control; 2015.

34. Pinar Y. Tunga G., et al; "Computer and Information Sciences-ISCIS 2005"; 20th International Symposium Istanbul Turkey, Springer press;2005.

35. Johan A, Tony V.; "Least Squares Support Vector Machines"; World Scientific; 2005.

36. Xiaolu Tian and Fang Hua; "An Innovative Localization System of Transformer Substation Automation Robots based on Pattern Recognition and ICP algorithm"; 4th International Conference on Applied Robotics for the Power Industry (CARPI); IEEE; 2016.

37. Mehdi Khosrow; "Geographic Information Systems: Concepts, Methodologies, Tools, and Applications"; IGI Global, 2013.

38- David L. Olson and Dursun Delen, “Advanced Data Mining Techniques” springer.com 2008.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	:	ALAA MAJEED Al-OBAIDI
Nationality	:	IRAQI
Date and Place of Birth	:	06.11.1984 - IRAQ. Al-ANBAR
Phone	:	05377453113
Email	:	alaamajeed1984@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	University of Turkish Aeronautical Association	2017
BS	University of Baghdad	2006
High School	Al-Amiriyah School	2002

FOREIGN LANGUAGES

English, Turkish