DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED

SCIENCES

MULTI-STAGE CLASSIFICATION OF ABNORMAL PATTERNS IN EEG AND ECG USING MODEL-FREE METHODS

by

Yakup KUTLU

April, 2010

İZMİR

MULTI-STAGE CLASSIFICATION OF ABNORMAL PATTERNS IN EEG AND ECG USING MODEL-FREE METHODS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Electrical and Electronics Engineering,
Electrical and Electronics Program

by
Yakup KUTLU

April, 2010
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**MULTI-STAGE CLASSIFICATION OF ABNORMAL PATTERNS IN EEG AND ECG USING MODEL-FREE METHODS**" completed by **YAKUP KUTLU** under supervision of **ASST. PROF. DR. DAMLA KUNTALP** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Asst. Prof. Dr. Damla KUNTALP

Supervisor

Prof. Dr. Cüneyt GÜZELİŞ

Thesis Committee Member

Asst. Prof. Dr. Adil ALPKOÇAK

Thesis Committee Member

Prof. Dr. Musa Hakan ASYALI

Examining Committee Member

Asst. Prof. Dr. Nalan ÖZKURT

Examining Committee Member

Prof.Dr. Mustafa SABUNCU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

MULTI-STAGE CLASSIFICATION OF ABNORMAL PATTERNS IN EEG AND ECG USING MODEL-FREE METHODS

ABSTRACT

In this study, computer based pattern recognition and classification systems are proposed for EEG and ECG patterns which are one dimensional biomedical signals. In the first phase of the study, artificial neural network based automatic recognition system for epileptiform events in EEG is proposed. Recognition process is performed both using single MLP based classifier and using multi-stage classifier. Different methods are used to increase the classification accuracy of the single MLP based system. In the second phase of the study, a novel multi-stage automatic arrhythmia recognition and classification system is proposed. The system performs beat-based classification and classifies 16 different beat types. The first stage of the system classifies five main groups then, in the second stage of the system each main group is classified into subgroups. In both classification stages the best feature set for each main group and subgroup is determined and used in classification process. With this approach, the curse of dimensionality effect is reduced. In addition, selecting and using the most discriminative features for each group increases the classification performance of the system. Furthermore, the third stage is added to the system for classifying beats that are labeled as unclassified beats in the first two classification stages. KNN classifier and raw data as input vector is used in this stage. The performances of the proposed systems are finally evaluated using real EEG and ECG data and results are discussed.

Keywords: Biomedical signal processing, Electrocardiogram, Electroencephalogram, Pattern recognition.

MODELDEN BAĞIMSIZ YÖNTEMLER KULLANILARAK EEG VE EKG İÇİNDEKİ ANORMAL ÖRÜNTÜLERİN ÇOK KATLI SINIFLANDIRILMASI

ÖZ

Bu çalışmada bir boyutlu biyomedikal sinyaller olan EEG ve EKG sinyallerindeki belli örüntüleri otomatik olarak tanıma ve sınıflandırma için bilgisayar destekli örüntü tanıma ve sınıflandırma sistemleri önerilmiştir. Çalışmanın ilk aşamasında EEG işaretinde klinik uygulamaları destekleyen, yapay sinir ağı tabanlı otomatik epileptik örüntü tanıma sistemi önerilmektedir. Tanıma işlemi, hem bir yapay sinir ağ sınıflandırıcı kullanılarak, hem de çok aşamalı bir sınıflandırıcı sistem kullanılarak gerçekleştirilmiştir. Bu sistemde sınıflandırma başarımını arttırmak için farklı yöntemler denenmiş ve sonuçları sunulmuştur. Bunu takip eden çalışmada, yine klinik uygulamaları destekleyen, EKG işareti için çok aşamalı yeni bir otomatik aritmi tanıma sistemi önerilmiştir. Sistem vuru tabanlı olup 16 aritmi tipi sınıflandırabilmektedir. Bu sistemde aritmiler ilk aşamada 5 ana sınıfa gruplanırken ikinci aşamada her bir ana grup alt aritmi gruplarına ayrıştırılmaktadır. Sınıflama işlemi yapılırken, her iki aşamada da her grup ve alt grup için o grubu en iyi tanımlayan öznitelikler belirlenmiş ve sınıflamada bu öznitelikler kullanılmıştır. Bu yaklaşımla hem öznitelik vektörlerinin boyutları düşürülerek, başarım üzerindeki olumsuz etkileri azaltılmış, hem de her bir sınıf için o gruba ait öznitelikler kullanılarak sınıflama başarımı arttırılmıştır. Ayrıca, ilk iki aşamada sınıflanamayan vurular 16 aritmi tipine ayırmak için, sisteme üçüncü bir aşama eklenmiştir. Bu aşamada, sınıflandırıcı olarak k-en yakın komşu ve giriş vektörü olarak da ham EKG verisi kullanılarak ilk iki aşamada sınıflanamayan vurular sınıflandırılmıştır. Sunulan sistemlerin başarımları gerçek EEG ve EKG verileri kullanılarak belirlenmiş ve sonuçları tartışılmıştır.

Anahtar sözcükler: Biyomedikal sinyal işleme, Elektrokardiyogram, Elektroansefalogram, Örüntü tanıma.

CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 Introduction

There has been a new period in medical diagnostic techniques, since the introduction of high technology equipments into health care. Since then, electronics and subsequently computers have become essential components of biomedical signal analysis, performing a variety of tasks such as data acquisition and preprocessing, feature extraction and interpretation. Applications of electronic instrumentation and computers have been widely used in biological and physiological systems and phenomena, such as the electrical activity of the cardiovascular system, the brain, the neuromuscular system, and the gastric system, etc.

Biomedical signal processing focuses on the acquisition of vital signals extracted from biologic and physiologic systems (Haddad & Serdijn, 2009). These signals allow getting information about the state of living systems. Hence, monitoring and interpretation of these signals have significant diagnostic value for clinicians and researchers to obtain information related to human health and diseases. In literature, there are many valuable books about biomedical signal processing and its importance such as (Feng, 2007; Haddad & Serdijn, 2009; Rangaraj, 2002; Sawhney, 2007).

In a signal processing system, obtaining a *measurable* electrical signal is very important. Therefore, sensors and instrumentation must be developed. Then the measured signals from physiological systems can be analyzed. Unfortunately, analyzing such signals is not an easy task for a physician or life-sciences specialist since noise and interferences often mask the clinically relevant information in the signals and it may not be easily comprehensible by the visual or auditory systems of a human observer. Furthermore, the variability of signal from one subject to another, and the inter-observer variability inherent in subjective analysis performed by physicians make consistent understanding or evaluation of any phenomenon difficult. In investigations of physiological systems, these factors created the need not only for improved instrumentation, but also for the development of methods for objective

1

analysis via signal processing algorithms implemented in electronic hardware or on computers.

Until a few years ago, biomedical signal processing was mainly directed toward filtering, spectral analysis and modeling. The filtering is used for removal of noise and power-line interference. The spectral analysis is performed to understand the frequency characteristics of signals. Modeling is utilized for feature representation and parameterization. But new trends in biomedical signal processing have been toward quantitative or objective analysis of systems via signal analysis. The biomedical signal analysis has moved forward to the stage of practical application of signal processing and pattern analysis techniques in order to efficient and improved noninvasive diagnosis (Rangaraj, 2002). The field of engineering aims to apply engineering principles to analyze and solve problems in life sciences and medicine. Techniques developed by engineers are increasingly accepted by practicing clinicians, and the role of engineering in diagnoses and treatment is gaining much deserved respect.

In the application of computers for biomedical signal analysis, the basic strength lies in the ability of signal processing and modeling techniques for quantitative or objective analysis. Observation by human sense is generally perceptual limitation for example, inter-personal variation, errors caused by fatigue, errors caused by the very low rate of incidence of a certain sign of abnormality, environmental distraction, etc. The interpretation of a signal by an expert varies according to the weight of the experience and expertise of the analyst. Such analysis is almost always personal. Computer based analysis has the potential to add objectivity to the interpretation of the expert. Therefore, it is possible to improve the diagnostic confidence and accuracy of even an expert with many years of experience. This approach could be named as computer aided diagnosis.

Automatic recognition helps in the diagnosis and facilitates the expert's work. It is especially useful during long-term monitoring such as electroencephalography (EEG) and electrocardiography (ECG) based monitoring systems. Examination of a record obtained over a period of days or weeks would be much time consuming if it

is done manually. Therefore, an automatic recognition system will intensely reduce the elapsing time.

The rapid development in the field of medicine applies variety of imaging techniques of the human body. The group of biomedical signal measurements includes items as ECG, EEG, electromyography (EMG), magnetoencephalography (MEG), computer tomography (CT), magnetic resonance imaging (MRI), functional MRI etc. EEG reads scalp electrical activity generated by brain structures, and ECG is reads electrical activity of heart. EEG and ECG are completely noninvasive procedures that can be applied repeatedly to patients, normal adults and children with virtually no possible risk or limitation.

Clinical recording of human brain electrical activity is the most important examination method for diagnosis of neurological disorders related to epilepsy. The EEG, which is used to display the electrical activity of the brain, has been a valuable clinical tool for this purpose. It has been accepted for a long time that epileptic spike activity, which is a type of transient waveform that appear in the inter-ictal period, i.e. in between seizures, have a high correlation with seizure occurrence. Therefore, the presence of spikes in the scalp EEG recordings is accepted as a confirmation for the diagnosis of epilepsy (Chatrian et al., 1974; Kiloh, McComas, Osselton, & Upton, 1981; Niedermeyer & Silva, 1993). For this reason, inter-ictal spike detection plays a crucial role in the diagnosis of epilepsy. Unfortunately, these spikes are very similar to and thus can easily be confused with non-spike waveforms produced by other brain disorders.

Similarly, the accurate recognition of the beats from an electrocardiographic (ECG) record has been a very important subject in intensive care units (ICU) and critical care units (CCU). This is due to the fact that the accurate recognition and classification of the various types of arrhythmias is essential for the correct treatment of the patient. Various algorithms for the automatic detection of ECG beats have been developed by different investigators for this purpose. These researchers used different features and classification methods. Despite all these developments, there is

still room for improvement in this area. A major problem challenging today's automatic ECG analysis algorithms is the considerable variations in the morphologies of ECG waveforms among different patients. Therefore, an ECG beat classifier performing well for a given training database could easily fail when confronted with a different patient's ECG signal. Because of this reason, the performance of the arrhythmia classification systems degrades when the number of arrhythmias to be classified is increased. This seems to be a major hurdle that prevents highly reliable, fully automated ECG processing systems to be widely used clinically.

In this thesis, considering the needs and trends in biomedical signal processing field, one dimensional biomedical signals, ECG and EEG, are studied to produce robust solutions for two major clinical problems, namely automatic spike detection and automatic heartbeat classification.

1.2 Organization of Thesis

This thesis consists of six chapters. Chapter 1 states the problems and outlines the motivation and the objectives of the thesis.

Chapter 2 provides background information about the physiological biomedical signal and abnormalities of these signals are also given in detail.

Chapter 3 provides background on pattern recognition methods. This chapter describes main processes of the pattern recognition system. Methods used in the proposed system are given in detail such as preprocessing, feature extraction, visualization of high dimensional features, feature dimension reduction methods and classification.

In Chapter 4, neural network based classification system and a multi-stage classification system are investigated for automatic recognition of epileptiform pattern in EEG signal. Multilayer perceptron networks trained by different training

algorithms are constructed. The training algorithms are compared in terms of their classification performances, and also different transform techniques which are applied the data are compared. A multi-stage classification system is introduced for automatic recognition of epileptiform pattern in EEG signal.

In Chapter 5, multi-stage system is introduced for automatic heartbeat recognition system in ECG records. Different feature extraction techniques are utilized. Feature selection algorithm is performed with sequential floating search and genetic algorithm to determine suboptimal solution. Also artificial neural networks are used for dimension reduction. Ensemble of classifiers system is constructed for both stages of the system. In the first stage, all heartbeats are classified into five main groups, and in the second stage, main groups are then separated into heartbeat classes.

Finally, Chapter 6 gives conclusion and contributions of the thesis and recommendations for future work.

CHAPTER TWO
PHYSIOLOGICAL BACKGROUND

Living organisms consist of many systems. For instance, the human body includes the nervous system, the cardiovascular system, and the musculoskeletal system. Each system consists of several subsystems that carry on many physiological processes.

The physiological processes include nervous or hormonal stimulation and control; inputs and outputs which could be similar to physical material, neurotransmitters, or information; and action that could be mechanical, electrical, or biochemical. Therefore, they are complex phenomena. Most physiological processes are accompanied by or appear themselves as signals that reflect their nature and activities. The signals could be different types, such as biochemical in the form of hormones and neurotransmitters, electrical in the form of potential or current, and physical in the form of pressure or temperature (Haddad & Serdijn, 2009).

When the signal is simple and it appears at the outer surface of the body, the task is not so hard. For instant, a rise in the temperature of the body is caused by most infections. It may be sensed very easily using simple thermometer or via hand. A single temperature is a *scalar,* and it shows the thermal state of the body at a single instant of time t. If the temperature is recorded continuously in some form, signal is obtained as a function of time. The example of body temperature is a rather simple example of a biomedical signal. On the other hand, other diseases such as abnormalities of cardiovascular system, respiratory system cannot be understood by simple observation way (Haddad & Serdijn, 2009).

Figure 2.1 shows a block diagram of medical care system that monitors and analyzes physiological signals from a patient. In data collection stage the physiological signals of patient are measured by sensors and converted to produce electrical signals. The electrical signals are then analyzed by a processor or computer system in data analysis part. The results of analysis are reported. According to the results of signal analysis, the

processor may perform direct therapeutic intervention on a patient or only reports the results of the analysis.
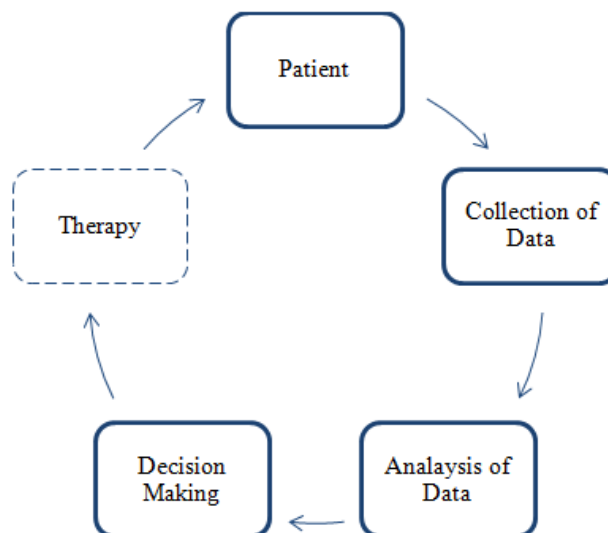
Figure 2.1 Basic elements of a medical care system.

Three basic types of data typically are used in the hospital. These are alphanumeric, medical images, and physiological signals. The patient's name and address, identification number, results of lab tests and physicians' notes are called as alphanumeric data. Medical images include X-rays and scans from computer tomography, magnetic resonance imaging, and ultrasound. Physiological signals are the electrocardiogram (ECG), the electroencephalogram (EEG), and blood pressure tracings.

Physiological signals like ECG, EEG, and EMG, represent an electrical activity. The electrical activity results from the chemical reaction in the cells. Chemical reactions inside and outside the cell provide mobile ions, and a small number of them move through the membrane. The permeability of ions varies for different ions. An imbalance of ions across the membrane of a cell causes voltage level, which changes with the movement of ions.

Table 2.1 shows characteristics of some physiological signals such as frequency band and measurement techniques.

Table 2.1 Medical and physiological parameters (Webster, 1998)

Parameter or Measuring Technique	Principal Measurement Range of Parameter	Signal Frequency Range, (Hz)	Standard Sensor or Method
Electrocardiography	0.5-4mV	0.01-250	Skin electrodes
Electroencephalography	5-300µV	0-150	Scalp electrodes
Electrocorticography and Brain depth	10-5000 µV	0-150	Brain surface or depth electrodes
Electrogastrography	10-1000 µV	0-1	Skin surface electrodes
Electromyography	0.1-5mV	0-10000	Needle electrodes
Electroneurography	0.01-3mV	0-10000	Surface or needle electrodes

The electroneurogram (ENG) is an electrical signal observed as a stimulus and the associated nerve action potential propagate over the length of a nerve. It may be used to measure the velocity of propagation of a stimulus or action potential in a nerve. ENGs may be recorded using concentric needle electrodes or silver-silver-chloride electrodes at the surface of the body (Haddad & Serdijn, 2009).

The electromyogram (EMG) signal indicates the level of activity of a muscle, and may be used to diagnose neuromuscular diseases such as neuropathy and myopathy. EMG signals are recorded using surface electrodes. Skeletal muscle fibers are considered to be twitch fibers because they produce a mechanical twitch response for a single stimulus and generate a propagated action potential (Haddad & Serdijn, 2009).

The electroencephalogram (EEG) signal represents the electrical activity of the brain. It is popularly known as brain waves. In clinical practice, several channels of the EEG are recorded simultaneously from various locations on the scalp for comparative analysis of activities in different regions of the brain (Haddad & Serdijn, 2009).

The electrocardiogram (ECG) is one dimensional signal which indicates the electrical activity of the heart and can be recorded fairly easily with surface electrodes on the limbs or chest. The ECG is perhaps the most commonly known, recognized, and used biomedical signal (Haddad & Serdijn, 2009).

The EEG and the ECG signals are the most commonly used biomedical signals which represent the electrical activity of brain and heart, respectively. In this thesis, these signals are investigated for diagnostic purposes. In the following subsections EEG and ECG signals are examined in detail.

2.1 The electroencephalogram (EEG)

EEG, which is also known as brain waves, represents the electrical activity of the brain and an important clinical tool in diagnosing, monitoring and managing of neurological disorders. It has also been used for investigating brain dynamics in neural engineering. It is comprised of electrical rhythms and transient discharges which are distinguished by location, frequency, amplitude, form, periodicity, and functional properties.

Generated signals by physiological control processes, thought processes, and external stimuli in the corresponding parts of the brain may be recorded at the scalp using surface electrodes. The scalp EEG is an average of the diverse activities of many small zones of the cortical surface beneath the electrode. The 10-20 system of electrode localization for clinical EEG recording has been recommended by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (Haddad & Serdijn, 2009).

The name 10-20 means that the electrodes along the midline are located at 10%, 20%, 20%, 20%, 20%, and 10% of the total nasion - inion distance; the other series of electrodes are also located at similar fractional distances of the corresponding reference distances. The scalp electrode localization is schematically illustrated in Figure 2.2.
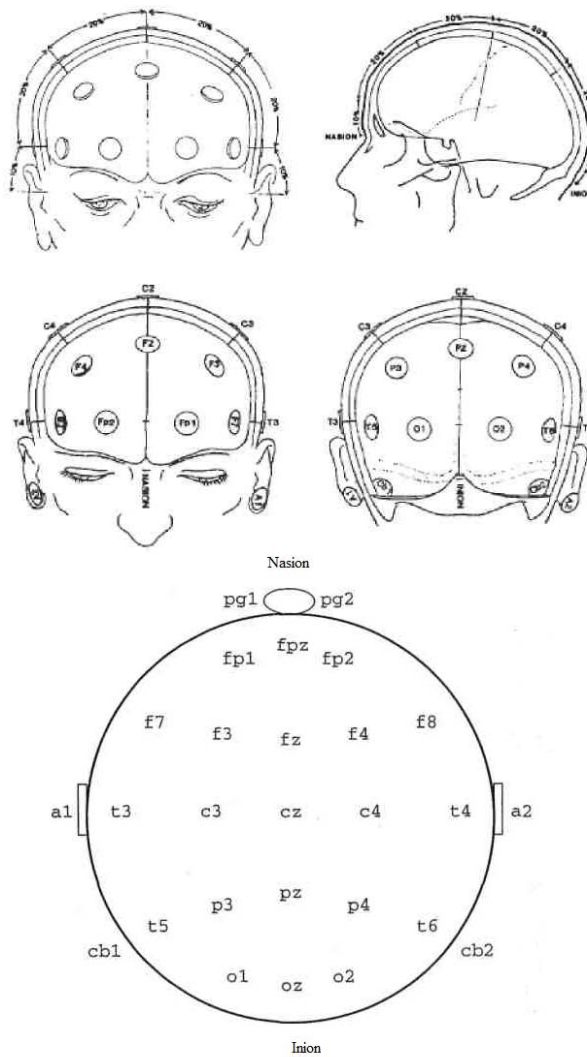
Figure 2.2 The 10-20 electrode placement system
(Acır, 2004).

19 locations are obtained on the scalp according the 10-20 system. Right-sided electrodes are even numbered and left-sided electrodes are odd numbered. Letters preceding the numbers refer to cortical regions. Frontal is 'F', prefrontal is 'Fp' (or frontopolar), parietal is 'P', temporal is 'T', central is 'C' and occipital is 'O'. Electrodes along the midline have no numbers only the letter 'z'. Figure 2.3 shows a sample of 19 channel EEG record.
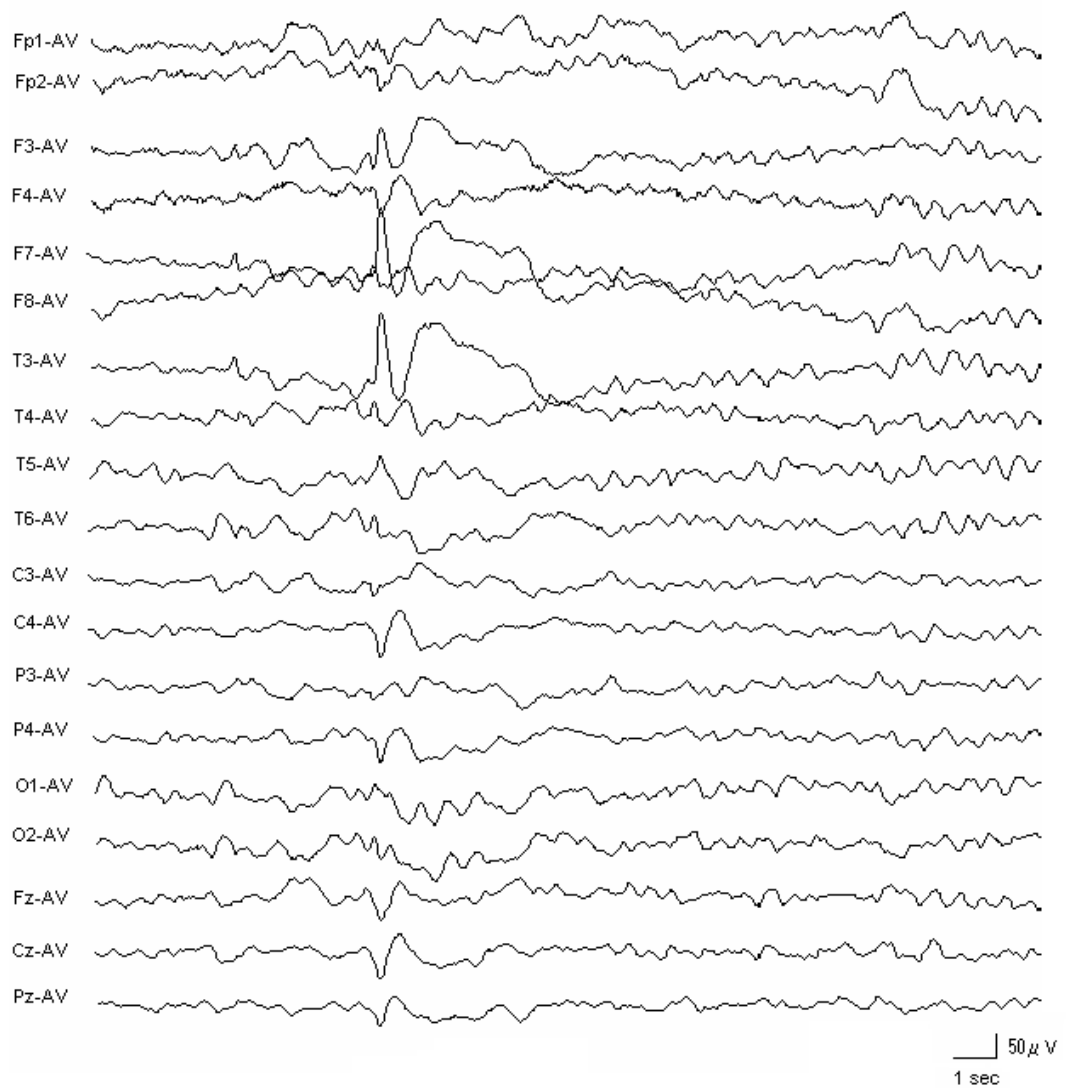
Figure 2.3 A sample of 19 channel EEG record.

EEG signals present several patterns of rhythmic or periodic activity. EEG rhythms are associated with various physiological and mental processes (Rangaraj, 2002). The commonly used terms for EEG frequency bands are:

- Delta: 0.5-4 Hz;
- Theta: 4 -8 Hz;
- Alpha: 8-13 Hz;
- Beta: 13-22 Hz; and
- Gamma: 22-30 Hz.

The delta activities appear at deep stages of sleep. The theta activities appear at the beginning stages of sleep. The amplitude of theta and delta activity is less than 100μV (peak-to-peak). They are strongest over the central region of brain and are indications of sleep. The alpha rhythm is the principal resting rhythm of the brain. The amplitude of alpha activity is usually less than 10μV (peak-to-peak). Auditory and mental arithmetic tasks with the closed eyes cause strong alpha waves and when the eyes are opened it is suppressed. High frequency beta activities appear as background activity in tense and anxious subjects. The amplitude of beta activity is less than 20μV (peak-to-peak). High states of wakefulness and desynchronized alpha patterns generate produce beta activities. The amplitude of gamma activity is less than 2μV (peak-to-peak) and it consists of low amplitude, high-frequency waves that result from attention or sensory stimulation. (Haddad & Serdijn, 2009; Acır, 2004).

2.1.1 Abnormalities in EEG

EEG signals may be used to study the nervous system, monitoring of sleep stages, biofeedback and control, and diagnosis of diseases such as epilepsy. Epilepsy is a very common neurological disorder. It is defined as sudden, excessive and abnormal discharges in brain which may be caused by a variety of pathological processes of genetic or acquired origin. This disorder is often identified by sharp recurrent and transient disturbances of mental function or movements of different body parts (Göksan, 1998). Clinical recording of human brain electrical activity is the most important examination method for diagnosis of neurological disorders related to epilepsy. It relates to a number of diseases associated to the abnormal function of the brain. Episodes of sudden disturbances of consciousness, mental functions, motor, sensory and autonomic activities are called seizures (Fisch, 1991). Sharp transient waveforms are characteristics of the epileptic seizures of focal origin in EEG. They are different from the background and exhibit a paroxysmal or abrupt, high voltage potential. The amplitude and morphologies of them vary from sharp transient to sharp transient. Such epileptiform sharp transients include both spikes with duration between 20 and 70 ms and sharp waves with duration between 70 and 200 ms (Chatrian et al., 1974).

2.2 The electrocardiogram (ECG)

The heart has four chambers (as shown in Figure 2.4) and circulates blood through the body as a pump. The main pumps are the two lower chambers called the ventricles. The upper two chambers, the atria, act as temporary storage for the blood while the ventricles pump blood to the rest of the body. Pumping is a two-phase process consisting of diastole and systole. Diastole is the resting and filling phase. Systole is the contracting and pumping phase. The contractions of both the atria and ventricles are coordinated by electrical activations. These activations propagate through the structure of the heart and cause depolarization and repolarization of cardiac muscle cells.
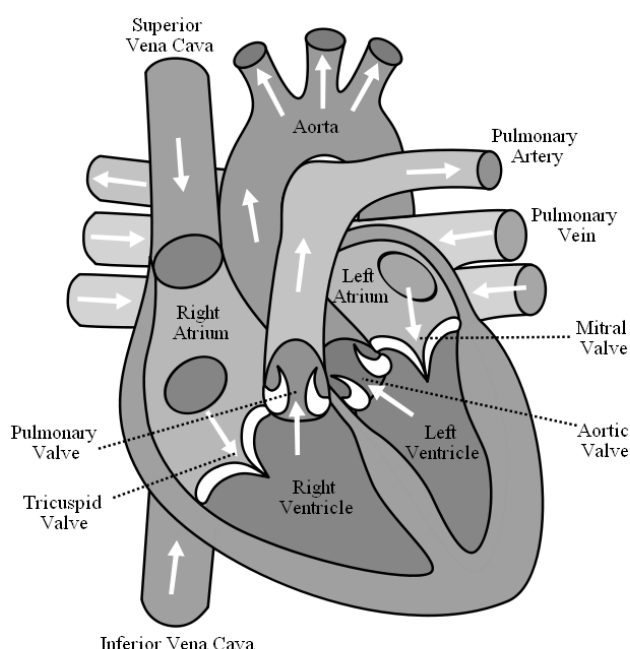
Figure 2.4 Anatomic diagram of the heart (frontal section) (Heart Structure, 2009).

For a normal rhythm activation begins at the sino-atrial (SA) node, also called the pacemaker of the heart. The SA node is located at the right atrium. It controls the rate of heart and rhythm. The atrioventricular (AV) rings prevent conduction between the chambers with the exception of a pathway through the AV node and AV bundle. Conduction continues from the AV node to the ventricles via the rapidly conducting His-Purkinje system. Figure 2.5 illustrates the activation sequence of the electrical activity for sinus rhythm.
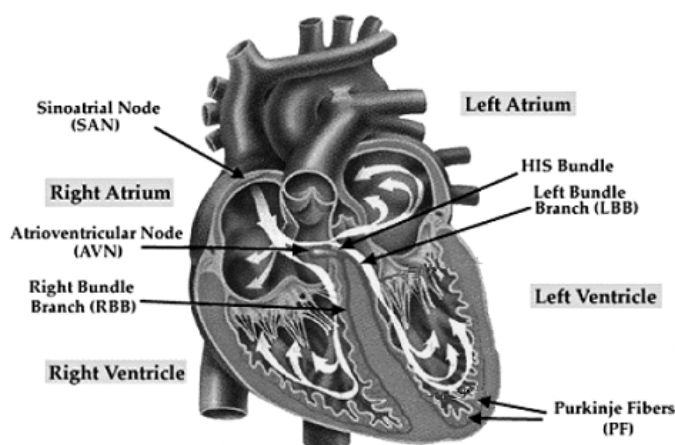
Figure 2.5 Activation sequence of sinus rhythm starting from the sino-atrial node (Activation sequence, 2009).

The electrocardiogram is the prevalent means of non-invasively observing the electrical activity of the heart. The series of activations of the heart result in potential differences that are spatially distributed and vary in time. The ECG can be recorded on the surface of the body; it provides an inexpensive and non-invasive means to monitor the heart's electrical activity.

The ECG signal repeats beat by beat, but the heartbeat rate of a recorded ECG changes with time. The mean and variance of the beat rate vary with time. Therefore, the ECG signal is considered to be quasi-periodic and non-stationary (Rangaraj, 2002).

In order to record ECG, standard twelve lead system is used. A standard twelve lead electrocardiograph uses ten electrodes. Six of these electrodes, which are named the precordial leads, are placed near the heart at anatomically defined positions on the left side of the chest wall as shown in Figure 2.6a. The remaining four electrodes are placed on the left arm (LA), left leg (LL), right arm (RA) and right leg (RL), respectively, as shown in Figure 2.6b. Of these, the right leg electrode is chosen to be the relative ground of the system. Three leads are defined between the electrodes on the arms and legs: lead I, between LA and RA, lead II, between LL and RA, and lead III, between LL and LA. The other three unipolar frontal leads, known as 'aVL',

'aVR', and 'aVF', which are usually called augmented unipolar leads, can be recorded from the same electrodes as the three leads LA, LL, and RA (Figure 2.7). The electrode on the right leg acts as a virtual ground for the system (Webster, 1998). Figure 2.8 shows an example of 12 lead ECG record.

a)   b)
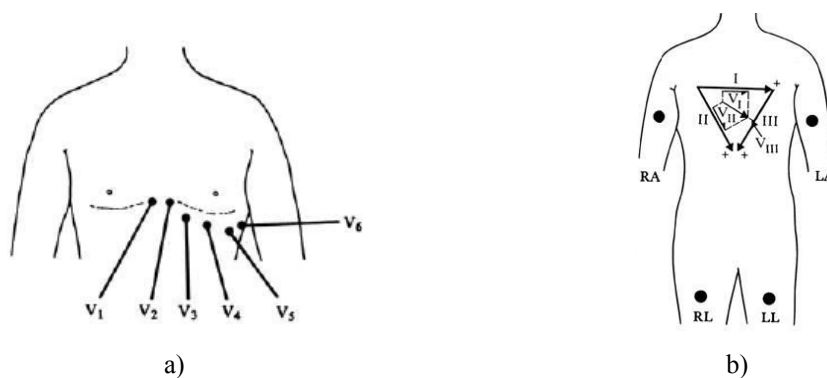
Figure 2.6 Positions of ten electrodes a) precordial leads on the chest wall, b) standard limb lead vectors (Webster, 1993).

a)   b)

c)   d)
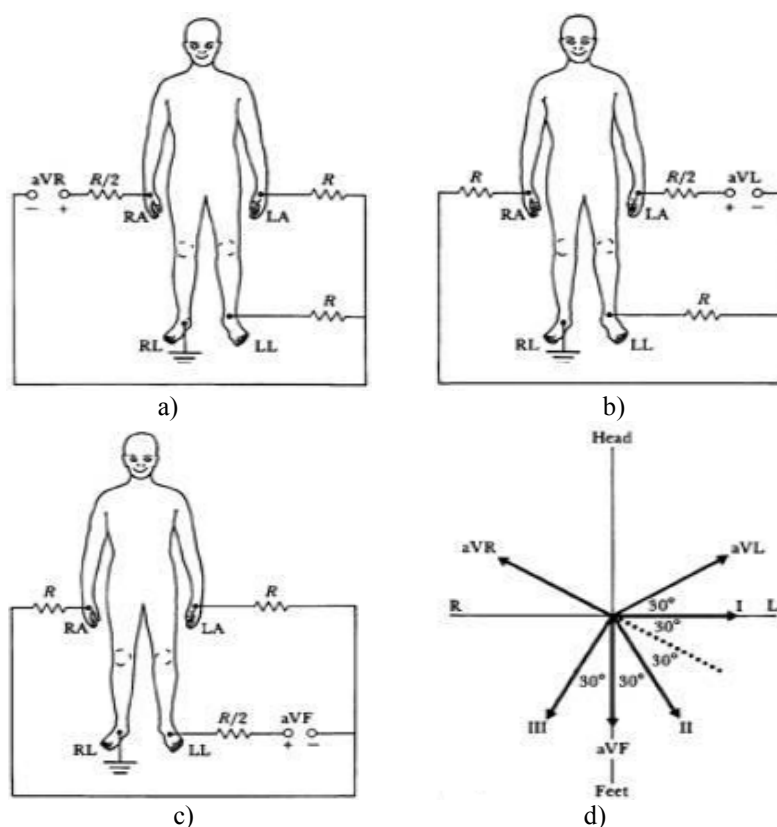
Figure 2.7 (a), (b), (c) Connections of electrodes for the augmented limb leads, (d) Vector diagram showing the directions of limb lead vectors in the frontal plane (Webster, 1998).
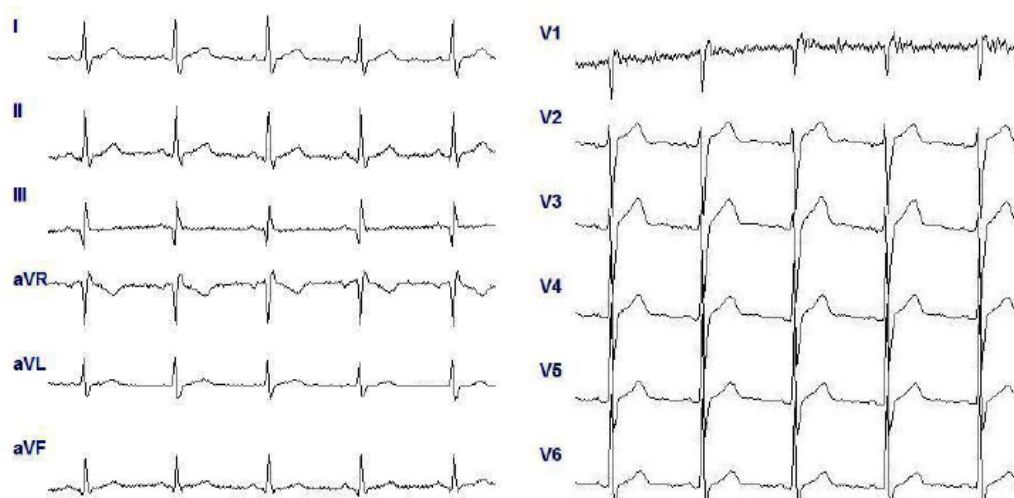
Figure 2.8 An example of 12 lead ECG record, using the BIOPAC MP30 bio-signal recording device.

The potentials arising from the depolarization, and subsequent repolarization, of a large group of heart muscle cells can be recorded by measuring the surface electric potential of the skin. Following is a brief description of how variations in the surface potential are related to the activity of the heart. The sum of these potentials results on the ECG is shown in Figure 2.9.

The electric activation has begun at the SA node as a small electrical activity, called the P wave. The generated action potential is propagated rapidly through the both atria walls. After the depolarization has propagated over the atrial walls, it reaches the AV node. The propagation through the AV junction is very slow. It results in a delay in the progress of activation. This is a desirable pause which allows giving the atria time to contract and empty blood into the ventricles before the ventricles contract. When the electrical activation has reached the ventricles, the propagation continues along the Purkinje fibers to the inner walls of the ventricles. In the next phase, depolarization waves occur on both sides of the septum. The progressive depolarization of the ventricular muscle cells result in the QRS complex on the ECG. This coincides with ventricular muscle contraction, a period known as the systole. Approximately 0.2 seconds after the QRS complex comes the T wave, which represents the repolarization of the ventricular muscle cells.
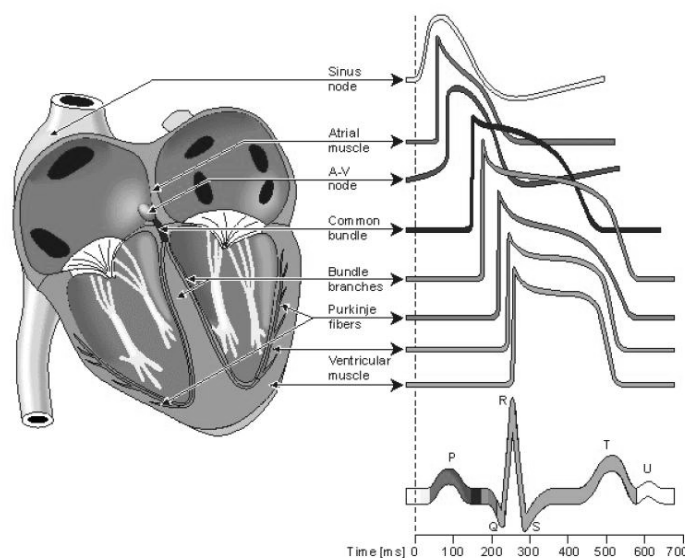
Figure 2.9 Electrophysiology of the heart (Webster, 1993).

In the case of a normal cardiac rhythm, the onset and offset of the QRS complex and the other waves can be readily identified and the shape of the QRS complex is evident. In fact, practicing cardiologists primarily exploit the shape to focus their attention on the ECG features to be studied in detail (Bottoni et al. 1990).

Generally the noise present in ECG recordings is introduced by the electrodes, either by them serving as antenna for electromagnetic radiation or by recording corrupted signals. The most common sources of signal corruption in electrocardiography are power line interference, motion artifacts, skeletal muscle contractions, baseline drift, electrosurgical noise, and electrode contact noise.

2.2.1 Cardiac Arrhythmias

Some of the most distressing types of heart failure occur not as result of abnormal heart muscle but because of abnormal rhythm. Deviation in the heart's rhythm from the normal physiological behavior is called arrhythmia, which is usually associated with abnormal pump function, thus resulting in reduction of life quality, or even death. Arrhythmias can be classified based on their underlying mechanisms into three groups: arrhythmias of abnormal impulse initiation (including automaticity and triggered activity), abnormalities of impulse propagation (including slowed

conduction/block, reentry and unidirectional block, ordered and random reentry), or combined (simultaneous abnormalities of both impulse formation and propagation) (Alpert, 1980; Gertsch, 2003; Webster, 1993; Wagner, 2001; Crawford, 2004).

There are many types of arrhythmias. Arrhythmias are identified by where they occur in the heart (atria or ventricles) and by what happen to the heart's rhythm when they occur. They also are classified as ectopic beats and pattern type arrhythmias.

Ectopic heartbeat is an irregularity of the heart rate and heart rhythm involving extra or skipped heartbeats. Extra heartbeats, called ectopic beats, are very common diseases. They may come either from the atria, the upper chambers of the heart, or the ventricle, the lower chambers. Ectopic beats are not in themselves dangerous and do not damage the heart. Types of ectopic beats and their properties are summarized below.

Supraventricular ectopic beat: It is a heartbeat that is caused by an ectopic impulse that occurs somewhere above the level of the ventricles.

Premature atrial contraction: The heart rate stays normal, but the rhythm becomes irregular due to the premature P wave. This arrhythmia type can cause palpitation, atrial flutter or atrial fibrillation.

Atrial escape beat: They are ectopic atrial beats that emerge after long sinus pauses or sinus arrest. They may be single or multiple; escape beats from a single focus may produce a continuous rhythm (called ectopic atrial rhythm). Heart rate is typically slower, the P wave morphology is typically different, and PR interval is slightly shorter than in sinus rhythm.

Ventricular premature beat (ventricular ectopic beat, premature ventricular contraction): It is an extra heartbeat resulting from abnormal electrical activation originating in the ventricles before a normal heartbeat would occur.

Premature ventricular contraction: Heart rate is variable. P wave is usually obscured by the QRS, PST or T wave of the premature ventricular contraction. The wideness of the QRS complex is more than 0.12 seconds and its morphology is unusual with the ST segment and the T wave opposite in polarity. QRS complex may be multi-focal and exhibit different morphologies.

Ventricular escape beat: It is an ectopic beat that occurs after an extended pause in a rhythm, indicating either the failure of the SA node to initiate a beat or the failure of the conduction of this beat to the AV node.

Premature junctional beat: it originates near the AV node junction. In general, they do not require treatment.

Left bundle branch block: activation of the left ventricle is delayed, which results in the left ventricle contracting later than the right ventricle. The duration is caused expansion of QRS complex.

Right bundle branch block: During a right bundle branch block, the right ventricle is not directly activated by impulses traveling through the right bundle branch. However, the left ventricle is still normally activated by the left bundle branch and these impulses travel through the left ventricle's myocardium to the right ventricle and activate the right ventricle. The duration is caused expansion of QRS complex.

Junctional escape beat: It is a delayed heartbeat produced from an ectopic focus somewhere in the AV junction. When the rate of depolarization of the SA node falls below the rate of the AV node, it occurs. This dysrhythmia may also occur when the electrical impulses from the SA node could not reach the AV node because of SA or AV block.

The other kinds of arrhythmias are pattern type arrhythmias. These types of arrhythmias are identified by the characteristic of consecutive beats, and grouped as supraventricular or ventricular.

Supraventricular arrhythmias occur in two upper chambers of heart (atrium). Types of supraventricular arrhythmias include atrial fibrillation (AF), atrial flutter, paroxysmal supraventricular tachycardia (PSVT). Ventricular arrhythmias occur in two lower chambers of heart (ventricles). Types of ventricular arrhythmias include ventricular fibrillation (AF), ventricular flutter, and ventricular tachycardia. The most dangerous types of arrhythmias are ventricular arrhythmias, since they may cause death.

Atrail fibrillation: It is an electrical rhythm disturbance. Abnormal electrical impulses in the atria cause the muscle to contract erratically and pump blood inefficiently. Hence, the atrial chambers are not able to completely empty blood into the ventricles. Pooling of blood in the atria can cause red blood cells to stick together and form a clot. The most worrisome complication of atrial fibrillation is dislodgement of a clot and embolism of the clot material to one of the major organs of the body (e.g., the brain) (Crawford, 2004).

Ventricular fibrillation: Ventricular fibrillation occurs when parts of the ventricles depolarize repeatedly in an erratic, uncoordinated manner. The ECG in ventricular fibrillation shows random, apparently unrelated waves. Ventricular fibrillation is almost invariably fatal because the uncoordinated contractions of ventricular myocardium result in ineffective pumping and little or no blood flow to the body. There is lack of a pulse and pulse pressure and patient lose consciousness rapidly. When the patient has no pulse and respiration, he/she is said to be in cardiac arrest.

Ventricular flutter: This is especially dangerous when the heart rate exceeds 250 beats per minute. The chambers of the heart contract so quickly that there is hardly any time for the blood to flow into and fill the chambers. In this situation, the heart transports only a little blood into the circulation. The person who is experiencing this is close to unconsciousness.

Ventricular tachycardia: Ventricular tachycardia is a rapid heartbeat initiated within the ventricles, characterized by 3 or more consecutive premature ventricular beats.

CHAPTER THREE
PATTERN RECOGNITION METHODS

3.1 Introduction

The main aim of pattern recognition is the classification of some patterns. Basic pattern recognition system consists of the following parts: preprocessing, feature extraction/selection, and classification as shown Figure 3.1.
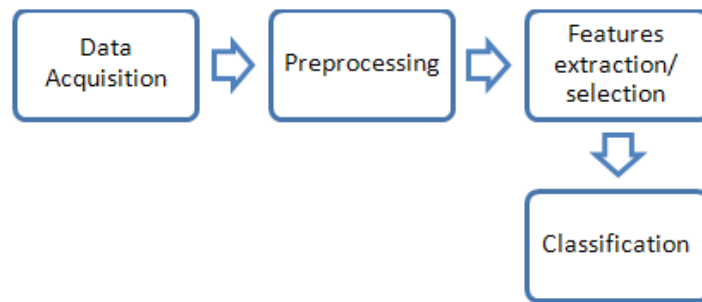
Figure 3.1 Basic process of pattern recognition system.

Data acquisition, noise removal, signal enhancement, and preparing data for feature extraction are the main operations of pre-processing. Feature extraction and selection are very important and crucial steps in pattern recognition. Feature extraction is the determination of a feature or a feature vector from a pattern. The feature vector is comprised of the set of all features which describe a pattern. The feature vector is reduced in size at the feature selection step. The classification step will be the final stage in automatic pattern recognition system. It makes a classification decision according to the input feature vector representing the sample data.

3.2 Pre-processing

In data acquisition step, data almost always be affected and corrupted by the environment. Other then the desired signal, interference, artifact, or simply noise are always present in the acquisition data. The sources of noise can be physiological, the used instrumentation, or the environment of the experiment. It is especially a big

problem for biomedical signals. All biomedical applications require an accurate analysis of the signal. Thus, noise in the signal must be removed in pre-processing stage.

Preparing signal for feature extraction stage is also operation of preprocessing stage such as peak detection, determining region of interest etc...

3.3 Feature Extraction

Feature extraction is an important step in pattern recognition. It is the process of information extraction which represents the characteristics of the pattern. The set of extracted information or features is called feature vector.

Various methods can be used for feature extraction to obtain information from the signal. Each feature can independently represent the original data, but none of them completely represents the all data for practical recognition applications. Furthermore, there seems to be no simple way to measure relevance of the features for a pattern classification task (Bhaskar, Hoyle, & Singh, 2006; Jain, Duin, & Mao, 2000; Duda, Hart, & Stock, 2001). In this case, diverse set of features often need to be used in order to achieve robust performance. The rapidly growing technology has also facilitated the use of detailed and diverse methods for data analysis and classification. Hence, the set of features will be selected from a large pool of candidate features including morphological, temporal, spectral, time-frequency, and higher-order statistical ones.

3.3.1 Raw Data

A specific window is determined and amplitude values of data in the window are used as a feature vector. It is a simple feature extraction method. Furthermore, it is not required additional computational process. The window size is a parameter that may be investigated to achieve good performance.

3.3.2 Higher Order Statistics

In signal processing, the first and second order statistics are widely used tools for signal representation. But they are not always sufficient for representing some signals. Higher order statistical methods are used, when the signals can't be examined properly by second order statistical methods. while the first and second order statistics contain mean and variance, higher order statistics contain higher order moments (m_3, m_4, …) and non linear combinations of higher order moments which are known as cumulants (c_1, c_2, c_3 …). Cumulants are blind to any kind of a Gaussian process.   Therefore, cumulant-based methods boost signal-to-noise ratio when signals are corrupted by Gaussian measurement noise (Mendel, 1991).

For zero mean discrete time signal moments and cumulants are defined as:

$$m_2(i) = E\{X(n).X(n+i)\} \tag{3.1}$$

$$m_3(i,j) = E\{X(n).X(n+i).X(n+j)\} \tag{3.2}$$

$$m_4(i,j,k) = E\{X(n).X(n+i).X(n+j).X(n+k)\} \tag{3.3}$$

where $E(.)$ is defined as the expectation operation, and X is the random process.

$$c_2(i) = m_2(i) \tag{3.4}$$

$$c_3(i,j) = m_3(i,j) \tag{3.5}$$

$$c_4(i,j,k) = m_4(i,j,k) - m_2(i)m_2(j-k) -$$
$$-m_2(j)m_2(k-i) - m_2(k)m_2(i-j) \tag{3.6}$$

Higher-order statistics are applicable for non-Gaussian processes. Many applications in real world are truly non-Gaussian.

In addition to representing the signals in time domain, we can also compute the spectra of the random signal, which is called the power spectrum. Power spectrum is given as the discrete Fourier Transform (DFT) of the second order moment c_2.

$$P_2(f) = DFT(c_2(m)) = \sum_{m=-\infty}^{\infty} c_2(m)e^{-j2\pi mf} \qquad (3.7)$$

Similarly, the spectrum of the 3rd order cumulant, the bispectrum, is given as:

$$B(f_1, f_2) == \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_3(m,n)e^{-j2\pi(mf_1+nf_2)} \qquad (3.8)$$

The bispectrum is a function of two frequencies and carries information about the phase. The power spectrum does not carry any information about the phase.

3.3.3 Frequency Domain Measures

Fourier transform (FT) is often called the frequency domain representation of the original signal. It describes which frequencies are present in the original signal so it is important tool for the digital signal processing. Implementation of algorithm of FT can be found in many popular digital signals processing book such as (Ingle & Proakis, 2000)

Discrete Fourier Transform (DFT) of an N-point evenly-spaced sequence is

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \qquad k=0,1...N-1 \qquad (3.9)$$

where, X_k is the DFT of x_n.

The energy spectral density describes how the energy of a signal is distributed with frequency and given as

$$\Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} f_n e^{-iwn} \right|^2 = \frac{F(\omega)F^*(\omega)}{2\pi} \qquad (3.10)$$

where, $F(\omega)$ is DFT of f_n.

3.3.4 Time-Frequency Domain Measures

The original signal or function can be represented in terms of wavelet expansions. The wavelet expansions are coefficients in a linear combination of the wavelet functions and the corresponding wavelet coefficients can be used in practice as features to represent the signal. Wavelet analysis has found wide area applications, since wavelet analysis can be applied to both stationary signals and non-stationary signals.

Wavelets are functions that satisfy certain mathematical requirements. The wavelet analysis procedure consists of determining a wavelet prototype function, and calculating the correlation between the signal and the dilated and shifted wavelet prototype function. The wavelet prototype function is called mother wavelet denoted as *Ψ(t)*. A set of basic functions used in wavelet transform are the scaled and translated versions of the *Ψ(t)*.

$$\Psi_{a,\tau}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x - \tau}{a}\right) \qquad \tau \in R \qquad (3.11)$$

where, τ is a shift position, a is a positive scaling factor, $a > 1$ corresponds to a dilation, while $0 < a < 1$ to a contraction of *Ψ(t),* and R denotes the set of real numbers.

Equation 3.11 shows that wavelets are used with different scaling factor a. This preserves the same shape and changes the size. Such a dilation or contraction property is used to represent a non- stationary function through wavelet transform (Meyer, 1993).

The continuous wavelet transform (CWT) of a real valued function *x(t)* is given as

$$C(a, \tau) = \int x(t) \frac{1}{\sqrt{a}} \Psi\left(\frac{t - \tau}{a}\right) dt \qquad (3.12)$$

where, $\Psi(t)$ is the mother wavelet and $x(t)$ is the original signal, $C(a, \tau)$ is called wavelet coefficients, which represent the correlation between the signal and the chosen wavelet at different scales.

For a given shift τ, the CWT is the result of the local analysis of the signal $x(t)$ at the given position τ with wavelet function whose width depends on the scale factor a. The amplitude of the coefficients is reaches the maximum at a position τ where the scaled prototype best matches the original function.

The mother wavelet $\Psi(t)$, must be band-limited in the frequency domain, must be a zero mean function, and must be a function with finite energy.

The discrete wavelet transform (DWT) has been presented in order to reduce the redundancy of the continuous wavelet transform. The algorithm to implement the DWT through multi-resolution analysis using filter banks is described by (Mallat, 1989). The general procedure of this DWT algorithm is to decompose the discrete signal into an approximation signal H_i and a detail signal G_i. Where, i represents scale level in the multi-resolution analysis. While the approximation signal is the low-passed signal, the detailed signal is the high-passed signal. Both of these signals have been down sampled after each scale.

Figure 3.2 shows the filter bank scheme of decomposing a signal. The implementation procedure of the multi-resolution decomposition of the signal by filter banks H and G is shown. The signal is decomposed into detail part by G and approximation by H, then down-sampled by 2, respectively. The decomposition and down-sampling for approximation are repeated again and again until a chosen scale is met or only one sample is left in the resulting approximation.
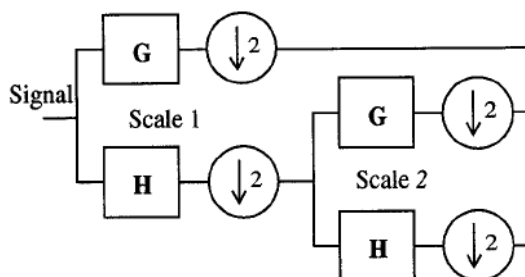
Figure 3.2 The filter bank scheme of decomposing a
signal.

DWT are commonly used in biomedical pattern recognition problems for feature extraction. The wavelet packet decomposition (WPD) method is an expansion of the classical DWT (Daubechies, 1992). The DWT only decomposes the low frequency components. Not only does the WPD utilizes the low frequency components but also the high frequency components (details) (Daubechies, 1990; Learned & Willsky, 1995; Misiti et al., 2004; Unser & Aldroubi, 1996). Figure 3.3 shows the wavelet decomposition trees of DWT and WPD. In Figure 3.3a, the signals are split into high frequency components (Details: D) and low frequency components (Approximations: A). The approximation achieved from the first level is split into new detail and approximation components and then this process is repeated. Therefore, it may miss important information which is located in higher frequency components. The original signal S is split as shown in Figure 3.3b for the 3-level decomposition. The top level of the WPD is the time representation of the signal. But the bottom level has better frequency resolution (Learned & Willsky, 1995). Thus, using WPD, a better frequency resolution can be achieved for the decomposed signal.

a)

b)
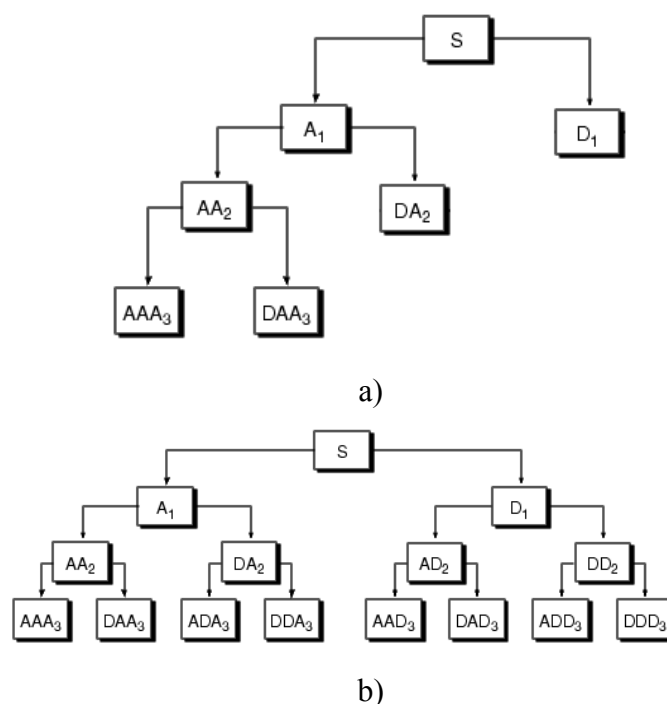
Figure 3.3 Decomposition trees (a) discrete wavelet transform and (b) wavelet packet analysis.

The advantage of the wavelet packet analysis is that it is possible to combine the different levels of decomposition in order to construct the original signal.

3.3.5 Morphological Representation

Morphological feature extraction method is one of the classical feature extraction methods. This approach is based on peak point in the signals. Morphological properties such as amplitude of peak, width of peak, slope of peak etc. are used as features.

3.4 Feature Transformation

Each data set is further transformed by using different transform methods such as normalization, nonlinear transformation, principal component analysis (PCA), and whitening transformation. These transformation methods are described in detail below.

- **Normalization:** The process of transforming the data from its original value into the range of -1 and 1 is called as normalization. There are several ways to normalize a data. The approach used in this study works by dividing the actual value by the absolute maximum value of each sample vector (Bishop, 1995; Duda, Hart, & Stock, 2001).

$$Y^k = \frac{X^k}{\left| X^k_{\max} \right|} \tag{3.13}$$

where, Y^k is the normalized data vector and X^k is the kth sample vector.

- **Nonlinear Transformation:** Nonlinear transformation is another process of transforming data from its original value into a new range (Özdamar & Kalayci, 1998). In this study, hyperbolic sigmoid function is utilized as the nonlinear function

$$Y^k = \frac{2}{1 + e^{-2X^k}} - 1 \tag{3.14}$$

where Y^k is the transformed data matrix and X^k is the normalized original data matrix. (Before the nonlinear transformation is applied, the original data is normalized first).

- **Principle Component Analysis (PCA):** PCA is a linear transformation method (Bishop, 1995; Duda, Hart, & Stock, 2001; Wiskott, 2004). In this method, first the d-dimensional mean vector and *dxd* covariance matrix are computed for the full data set. Next, the eigenvectors and eigenvalues of the covariance matrix are found and stored according to decreasing eigenvalue. The representation of original data by PCA consists of projecting the data onto a new subspace whose dimensionality K could be equal to or less than the dimensionality of the

original data d. The PCA transforms X to Y by the following equation;

$$Y = XV \qquad (3.15)$$

where Y is the transformed data matrix, X is original data matrix and V is portioned matrix consisting of eigenvectors corresponding to the eigenvalues decreasing value of the covariance matrix.

• **Whitening Transformation:** The whitening transformation is also a linear transformation (Duda, Hart, & Stock, 2001; Tang, Suganthan, Yao, & Qin, 2005). It performs a coordinate transformation that converts an arbitrary multivariate normal distribution into a spherical one. Therefore, the new distribution of data has a covariance matrix proportional to the identity matrix I. The whitening transformation transforms X to Y by the following equation;

$$Y = VX \qquad (3.16)$$

where Y is the transformed data matrix, X is the original data matrix, and V is a transformation matrix calculated by

$$V = D^{-1/2} E^{T} \qquad (3.17)$$

Here D is the diagonal matrix of eigenvalues and E represents the portioned matrix consisting of the corresponding eigenvectors of the covariance matrix.

3.5 Visualization of Multidimensional Data using Self Organizing Maps

Self-organizing maps (SOMs) are biologically inspired neural network architectures trained by unsupervised learning algorithms based on competitive learning rule (Kohonen, 1982; Kohonen, 2001). The SOM was invented by Kohonen (1982). SOM usage is divided as two main categories in the literature. In the first one, the neurons in the SOM represent different clusters in the data space. The number of neurons in this network corresponds to the number of clusters that exist in the input data. So, neuron size is very small; it is generally less than twenty. The other usage of SOM is related to the low dimensional visualization of high dimensional data (Ultsch, 2003). Humans simply cannot visualize high dimensional data. Therefore, different techniques have been developed to help visualize this kind of high dimensional data. One of these methods is the Unified Distance Matrix (U-matrix). U-matrices are invented for the visualization purposes of these high dimensional structural features. The U-Matrix is the canonical tool for the display of the distance (and topological) structures of the input data (Ultsch, 1992). In these models of SOM, very large numbers of neurons are used, generally over 1000.

The SOM is an unsupervised type neural network architecture used to visualize and interpret high-dimensional data sets on the map. The map usually consists of a two-dimensional regular (rectangular or hexagonal) grid of nodes called neurons as shown Figure 3.4 and Figure 3.5. Each sample of high dimensional input data is associated with a unit which is the winner. Not only the winning neuron but also its neighbors on the lattice are allowed to learn and adapt their weights towards the input. This way, the representations will become ordered on the map. After training, the responses of the SOM network are ordered on the map. This is the essence of the SOM algorithm and its main distinction from other networks.
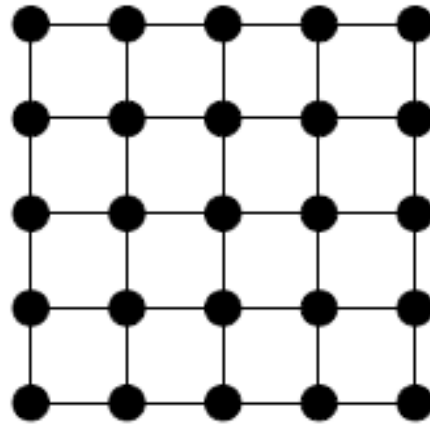
Figure 3.4 Rectangular grid structures.
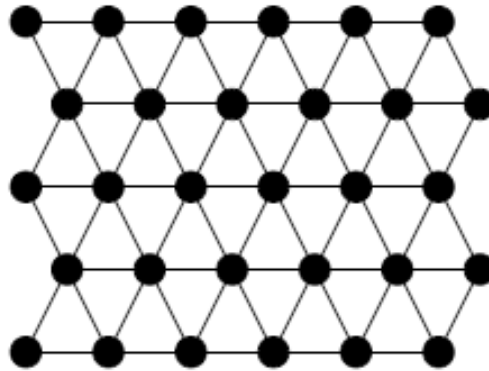
Figure 3.5 Hexagonal grid structures.

An N-dimensional input is presented to each neurons of a SOM network as shown in Figure 3.6. Then the winner unit (indicated by the index c), i.e. best match, is identified by the condition shown below for each sample,

$$\| x_i(t) - w_c(t) \| = \min_i \| x_i(t) - w_i(t) \| \qquad (3.18)$$

where x_i is input vector with N dimension, w_i is the i$^{\text{th}}$ weight, and c indicates the winning neuron.
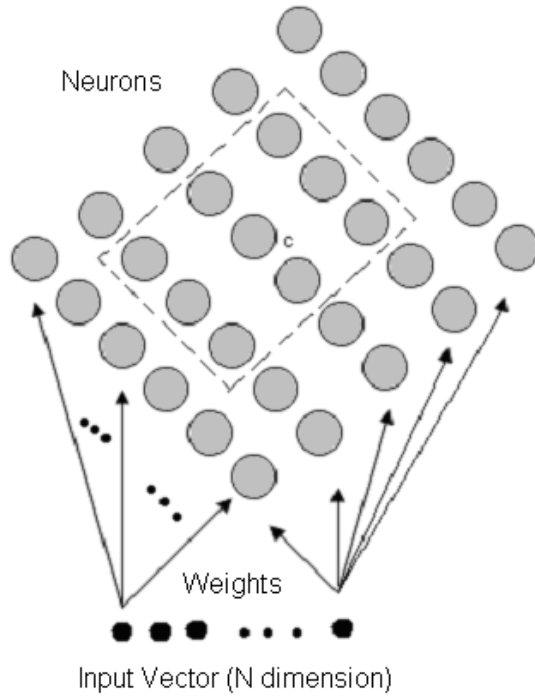
Figure 3.6 Self-Organizing Map Structure.

The update of the weights in the SOM network is limited by neighborhood function ($\Omega_c(i)$). The neighborhood function plays a main role in SOM algorithm regardless of the type of the learning algorithm. Three frequently used neighborhood functions are Gaussian, rectangular and cut Gaussian. The weight of the winning unit and its neighbors are updated by the formula

$$\Delta w_i = \eta (x - w_i)\Omega_c(i) \quad i \in NB_c \tag{3.19}$$

where η is the learning rate in the interval $0 < \eta < 1$, $\Omega_c(i)$ is the neighborhood function and NB_c indicates the neighbor neurons centered around node c, i.e. the winning neuron.

3.5.1 U-Matrix

After training the SOM network, the weight vectors that connect the high dimensional input vector space to 2-D output map grid are obtained. The distance

between the two mapped units on the projected plane is obtained through their respective weight vectors. The U-matrix method determines the distances between weight vectors of the adjacent map units. A U-matrix is originally defined on planar map spaces and a U-matrix representation of the Self-Organizing Map visualizes the distances between the neurons. The distance between the neighbor neurons is calculated and presented with different colorings.

There are various methods for U-matrix calculation from the trained weight vectors (Ultsch, 1992; Ultsch, 1993; Livarinen, Kohonen, Kangas, & Kaski, 1994; Oja et al., 2002). One of the methods used in the construction of the U-matrix uses the sum of the distances of the weight vectors to their neighboring weight vectors at each map coordinate (X;Y) (Ultsch, 1992). Another method is the median method. In this method, the distances between all adjacent neighbors are computed using the same distance metric. The median distance corresponds to the distance measure for that grid. Another commonly used approach uses a dummy grid in between every pair of map grids. In this method, the distance between two map grids are calculated and then assigned to the dummy grids as shown Figure 3.7. This is one simple way of calculation of the U-matrix with dummy grids (Oja et al. 2002). The value to be assigned to the original map grids are taken as the median distance of all its neighbors. A different method of U-matrix computation for various types of lattice grids is discussed in the literature (Livarinen et al., 1994).
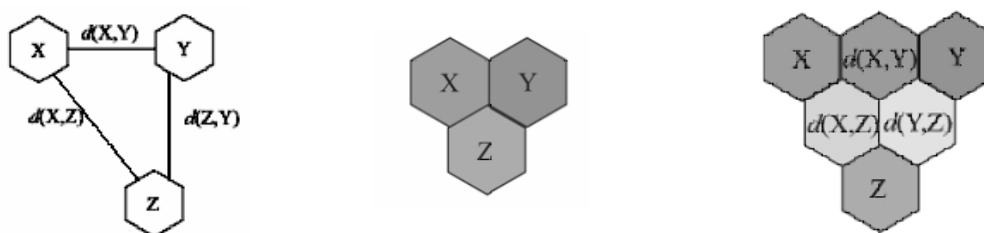
Figure 3.7 A simple way of calculating the U-matrix with dummy grids.

The computed U-matrix is visualized via a colored image or a gray-level image. The resultant gray-level image is a hexagonal grid map with different shades of gray-scale for the grids. The gray-scale map carries input pattern identification labels. The formation of clusters in the data and location of outlier observations become visible

from such a gray-scale image. Typically, lighter shade patches indicate the location of data vectors which are similar and have less mutual distance; darker shade patches, on the other hand, indicate the location of data vectors having larger distance with observations in adjoining lighter shade areas. The outliers are identified as observations located in the darkest patches of the projected map.

Figure 3.8 shows a U-matrix representation of a SOM network with gray-level image. The neurons of the network are marked as black dots. The representation shows that they correspond to separate clusters in the upper right corner of this representation. The clusters are separated by a dark gap.
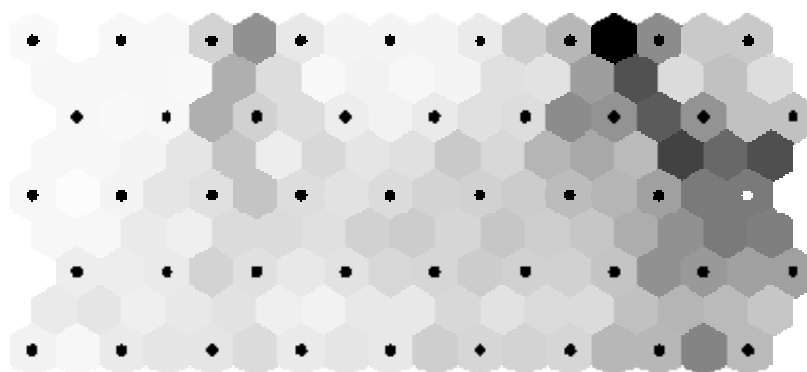
Figure 3.8 U-matrix representation of the SOM network with gray-level image.

The distances between the neighboring units are represented as heights in a 3-dimensional landscape. This is called as the hill-valley landscape visualization of the SOM. In this representation, there are valleys where the reference vectors in the lattice are close to each other and hills where there are larger mutual distances indicating dissimilarities in the input data. The height of the hills reveals the degree of dissimilarity among the data vectors. So the hills represent border of the clusters as shown in Figure 3.9. Outliers can be identified from this 'hill-valley' landscape visualization as they are typically located at higher locations on the hills. The degree of leverage of the outliers is associated with the height of the peaks of the corresponding hills.
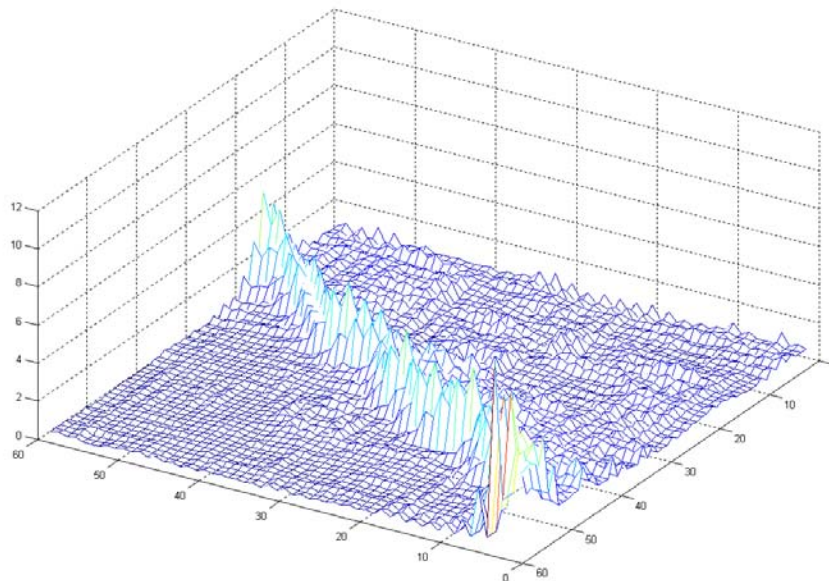
Figure 3.9 Three dimensional landscape visualization of high dimensional data.

3.6 Dimensionality Reduction

Determination of the relevant features and reduction of the dimension of the feature space is very important in a pattern classification task to improve the classification accuracy and reduce the computational cost. For this purpose there are three approaches that could be applied. In the first case, feature selection methods are used to find best subset from a large group of features to maximize classification performance. The selected features keep original physiological meaning, which may be important for understanding the physiological properties of the pattern. The other approaches are feature extraction and dimension reduction (Bhaskar, Hoyle, & Singh, 2006; Jain, Duin, & Mao, 2000). These methods create a reduced number of new features using combined features. These methods may not keep physiological meaning of the features. On the other hand, they may have better discriminative power (Jain, Duin, & Mao, 2000). PCA, SOM, and MLP are widely used effective methods in pattern recognition for feature dimensionality reduction and feature extraction (Bhaskar, Hoyle, & Singh, 2006; Jain, Duin, & Mao, 2000).

Feature selection is identified as: given a set of d features selects a subset of m

features that leads to the smallest classification error. Feature selection methods consist of detecting the relevant features and discarding the irrelevant features. Therefore, it improves generalization performance of the machine learning algorithm, and reduces data size for limiting storage requirements. Feature selection methods grouped as filter methods (open-loop methods) and wrapper methods (closed-loop methods) (Maroño, Betanzos, & Sanromán, 2007; John, Kohavi, & Pfleger, 1994; Kohavi & John, 1997) as shown Figure 3.10. Filter methods are based mostly on selection of features using the statistical measures and they do not depend on a classifier. Wrapper methods are, on the other hand, based on feature selection using a classifier performance as the selection criterion. Feature selection with wrapper method is used to find best subset from a large group of features that maximize classification performance of a specified classifier.

a)

b)
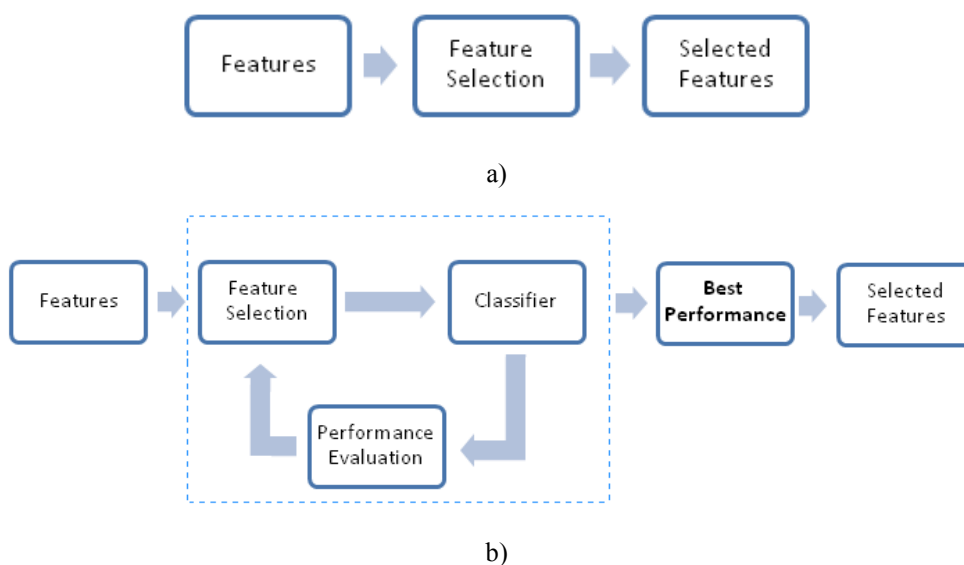
Figure 3.10 Block diagram of feature selection a) with filter methods and b) with wrapper methods.

In order to find optimal solution exhaustive search is used. But exhaustive search requires a lot of time to test the performance of the possible subset combinations of features (Jain, Duin, & Mao, 2000). So using deterministic or stochastic approach suboptimal feature set may be found in wrapper method. Sequential floating search and genetic algorithm are the most used methods for feature selection for finding suboptimal feature set.

3.6.1 Feature Selection with Sequential Floating Search

The sequential floating search methods (SFSM) are effective feature selection techniques (Pudil, Novovicova, & Kittler, 1994; Bhaskar, Hoyle, & Singh, 2006; Jain, Duin, & Mao, 2000; Duda, Hart, & Stock, 2001). The floating search method has two main categories: sequential forward search (SFS) and sequential backward search (SBS). The SFS algorithm starts with a null feature subset. For each step, the best feature that satisfies some criterion function is included to the current feature subset and this is repeated n times or it is repeated for all features and best subset which has best criterion value is chosen. The SBS algorithm starts with all features and for each step, the worst feature (concerning the criterion function) is eliminated from the subset and this is repeated r times or all features. For all features, the best subset which has best criterion value is chosen.

Extended case of SFSM uses both SFS and SBS as shown in Figure 3.11, which is called n-take r-away search algorithm or Plus-l-Minus-r method. This algorithm starts with a null feature set and in the case of forward search, for each step, the best feature that satisfies some criterion function is included to the current feature set and this is repeated n times. In the case of sequential backward search, the worst feature (concerning the criterion function) is eliminated from the set and this is repeated r times. SFS proceeds dynamically increasing the number of features and SBS proceeds decreasing the number of features until the desired feature size is reached or criterion function begin to decrease.
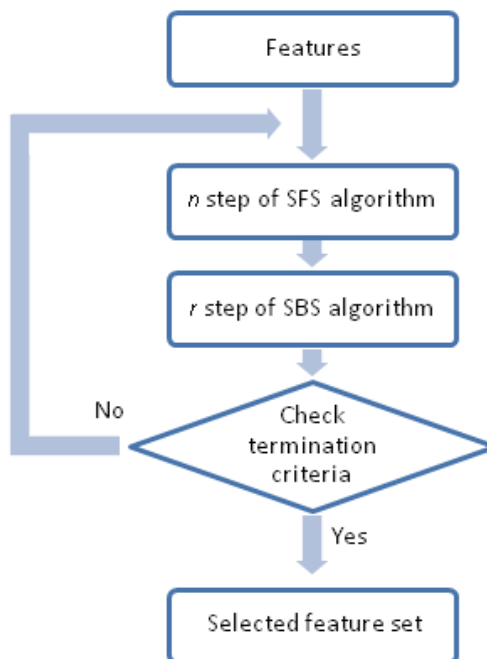
Figure 3.11 Block diagram of *n*-take *r*-away algorithm.

3.6.2 Feature Selection with Genetic Algorithm

Genetic Algorithms were discovered by John Holland (1975). It is a model for the evolution of a population in a special environment (Holland, 1975; Goldberg, 1989). Each member of the population is represented by a chromosome that consists of a series of genes. Each gene has two or more possible values and is transformed into a parameter of the problem space. A fitness function represents the environment. It evaluates each individual and determines a fitness value for each individual.

The algorithm is started with chromosomes which represents a set of solutions called population. The solutions from one population are taken and used to generate a new population. Generating new population, selection, crossover and mutation process are applied, then fitness values are evaluated. These processes are repeated until some criteria, i.e., reaching the best solution or certain number of population, elapsed time etc. Corresponding block diagram is shown in Figure 3.12.

The new population will be better than the old one since at least one best solution is copied without changes to a new population. It is called elitist strategy. The genetic operators such as representation, selection, crossover, mutation are described to construct GAs for optimization problems.
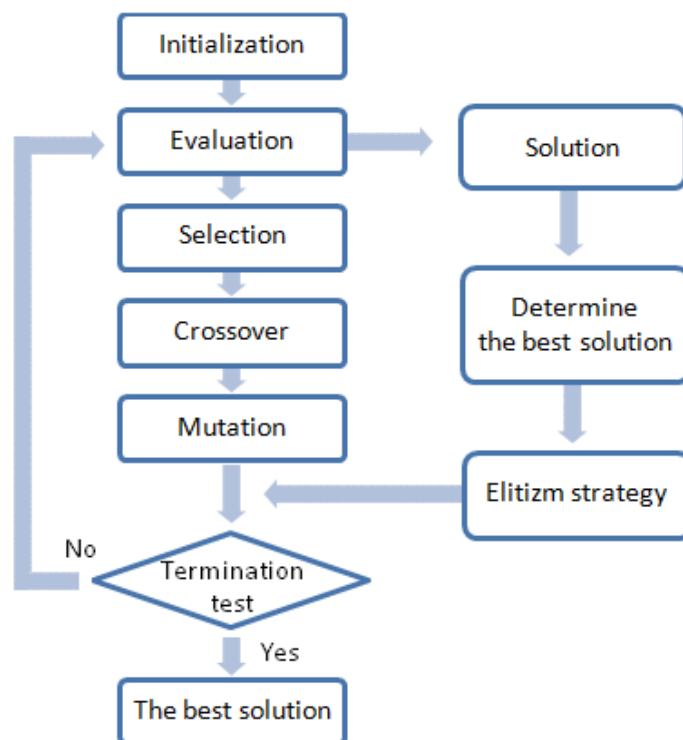
Figure 3.12 Block diagram of a typical genetic algorithm

The representation of chromosomes may be categorized into the two methods binary coding and real coding. For instance, the string shown Figure 3.13 is stored as a binary bit-string (binary representation) or as an array of integers (real-coded representation). The string by the binary coding consists of 0s and 1s. The binary string is decoded to the parameter value in integer, real number, or any parameter used with fitness function in GA. The binary representation is generally used in GA.

Binary Representation   Real-coded representation

| 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |

| 2 | 5 | 0 | 7 | 8 | 8 | 7 | 9 | 1 |

Figure 3.13 Representation of chromosomes

Each of candidate solutions is evaluated for optimization problems according to a fitness function. The fitness function is a criterion function which determines how well the candidate solution is. The GA searches a string (member of the population) with a better fitness value in the population.

Selection is an operator to select two parent strings for generating new strings. A string with a high fitness value has a higher probability to be selected. In GA parents are selected by random. The fitness value of each string is used for calculating the selection probability. The roulette wheel selection scheme and the rank-based selection scheme are often employed.

Crossover is an operator to generate offspring from parent strings. Various crossover operators have been proposed for GAs. One-point crossover for the binary coding is as an example of standard crossover operators. Two parents are selected randomly from the population. Then one crossover point is selected randomly and two new strings are generated by changing the substring along the crossover point. One point crossover operation is shown Figure 3.14.
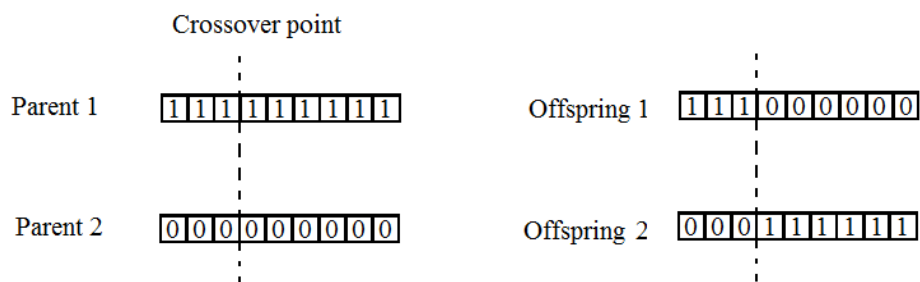
Figure 3.14 The standard one-point crossover for binary strings

Mutation is an operator to change elements in a string generated by a crossover operator with low probability. When applying mutation operator to strings in GAs, randomly selected one bit in a string is changed. An example of this mutation is shown in Figure 3.15. The stars are located as mutation points where 1 in the first and second position is changed to 0. Mutation supplies genetic variety and enables the genetic algorithm to investigate a broader space.

	*				*			
1	0	0	1	1	1	0	0	1

Mutation $\longrightarrow$

				*			*	
1	0	0	0	1	1	0	0	0
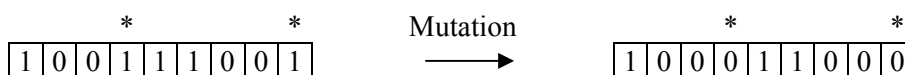
Figure 3.15 Two bit mutation for a binary string

Implementation of feature selection with GA is illustrated by the simple flow diagram shown in Figure 3.16. Each individual of population is represented by binary string. Fitness function for evaluation is the classifier performance.
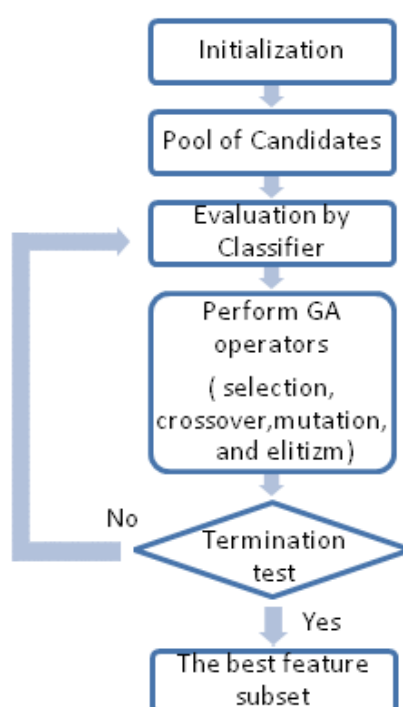
Figure 3.16 Wrapper type feature subset selection with GA.

3.6.3 Dimension Reduction using Neural Networks

A simple three layer linear network can be used as a dimension reduction tool. Figure 3.17 shows the structure of three layer neural networks. Each pattern of data set is applied to both input and output layer. Network trained by gradient descent on a sum squared error criterion. Activation functions of network are linear for all layers. So, dimension of the input data is d. The transformation F_1 is linear projection onto a K dimensional subspace. The transformation matrix is the weights of the first

layer of the network. The transformed data of *y* is the output of the linear hidden layer of networks.

$$\bar{y}_i = \sum_{j=1}^{d} w_{ij} \bar{x}_j \qquad (3.20)$$

The inverse transformation matrix is the weights of the last layer of the network. As a result of transformation, *d* dimension of input data is reduced to *K* dimension *(k<d)*. When linear activation function is used in the network, it provides the principle components (Duda, Hart, & Stock, 2001).
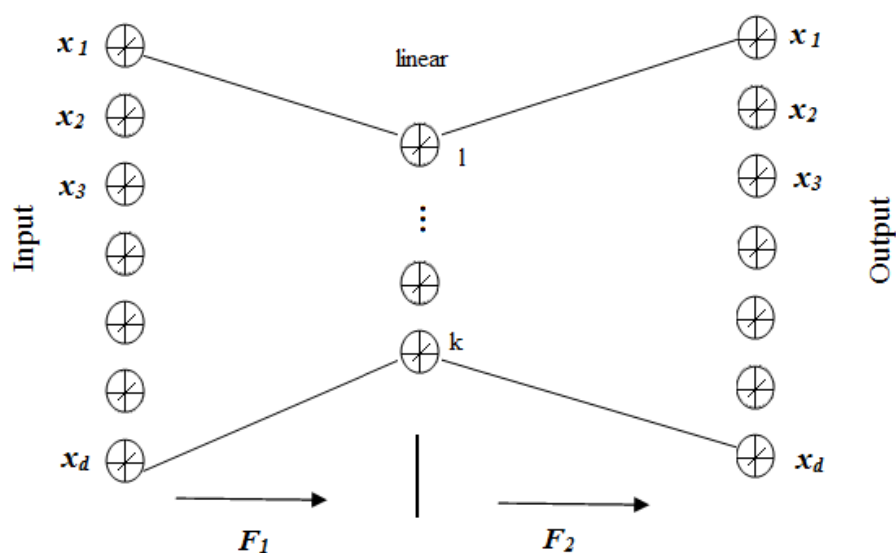
Figure 3.17 Three layer neural networks with linear hidden layer.

The SOM is also used as a dimension reduction method. The SOM is an unsupervised type of neural network architecture used to visualize and interpret high-dimensional data sets on the map (described in chapter 3.1.4). The map usually consists of a two-dimensional regular grid of neurons. Each sample vector of multidimensional input data is associated with a neuron on the map. An N-dimensional input is presented to each neurons of a map. Then the index of neuron which represents the input pattern is used as feature (Bhaskar, Hoyle, & Singh, 2006; Jain, Duin, & Mao, 2000; Kutlu & Kuntalp, 2009).

3.7 Classification

The final stage in pattern recognition is classification of a pattern. The recognition of a pattern generally consists of either supervised classification or unsupervised classification (Watanabe, 1985). In supervised classification, the classes are determined by the system designer and it is used as output of the system while training the system. On the other hand, in unsupervised classification, the class labels are not used. They are learned based on the similarity of patterns.

Increases in computing power have made possible the use of elaborate and diverse methods for data analysis and classification. In more recently, demands on automatic pattern recognition systems have raised enormously due to the availability of large databases and high performance computers. In great number of recognition applications, it is clear that there is no optimal approach for classification (Duda, Hart, & Stock, 2001). Thus, multiple methods are needed to employed and combining several methods and classifiers is now commonly used practice in pattern recognition problems (Jain, Duin, & Mao, 2000).

3.7.1 K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) algorithm is one of the most classical and effective nonparametric method in pattern recognition (Cover & Hart, 1968; Duda, Hart, & Stock, 2001). The KNN algorithm is a method for classifying objects based on closest training samples in the feature space. Because of identification of neighbors, the objects are represented by position vector in multidimensional feature space. More accurately, by a *K*-nearest neighbor method, a new pattern, X, is assigned to that category to which the plurality of its *K* closest neighbors belong. A training set is *m* labeled patterns, and a nearest-neighbor method decides that some new pattern, X, belongs to the same category as do its closest neighbors in training set.

The KNN algorithm is among the simplest of all machine learning algorithms, because it based on only the distance measure. Different distance measures can be

used in this algorithm. Simple Euclidean distance commonly used as the distance measure in nearest-neighbor methods. That is, the distance between two patterns, $(x_1, x_2, x_3, ... x_n)$ and $(y_1, y_2, y_3, ... y_n)$, is:

$$d = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2} \qquad (3.21)$$

K is usually chosen small positive integer. If $K=1$, then the object is simply assigned to the class of its nearest neighbor as shown in the example given in Figure 3.18. There are two classes which are illustrated as stars and circles in Figure 3.18. The test sample (triangle) is considered according to the distance and classified either to the first class of stars or to the second class of circles. If $K = 1$ it is classified to the first class due to the closest class members is star. If $K = 3$, it is classified to the second class because there are 2 circles and only 1 star inside the inner circle.
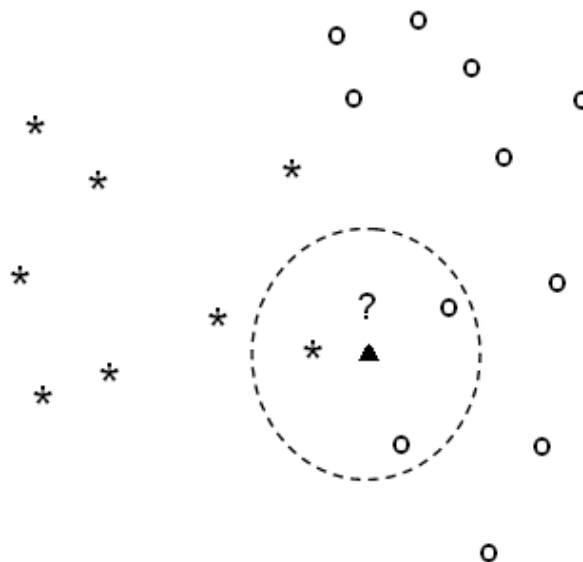
Figure 3.18 Example of KNN classification.

KNN is a supervised classifier. A set of object with class labels is required for the distance measurement of new object. This can be thought of as training set for the algorithm. As a matter of fact, it has no explicit training step. The algorithm is

independent from statistical distribution of training samples (Duda, Hart, & Stock, 2001).

Nearest-neighbor methods may require more memory. In order to achieve good generalization, a large number of training patterns must be stored. High memory cost is the major drawback of the method and its derivatives.

3.7.2 Artificial Neural Networks

ANNs, which are inspired by biologic neural networks, are composed of neuron-like units connected together through input and output paths that have adjustable weights (Bishop, 1995; Haykin, 1999). Each unit (neuron) produces an output signal, which is a function of the sum of its inputs. This function is formulated as:

$$y_i = f(\sum_{i=1}^{N} x_i w_i)$$
(3.22)

where, w_i represents the weights, x_i is the input, $f(.)$ is the activation function, and y_i is the output of the ith unit. A variety of functions can be utilized as the activation function but most often a sigmoid (or hyperbolic tangent) function is used.

Among different structures used in the ANNs, multi-layer perceptrons (MLPs) are the mostly used ones. An MLP consists of successive layers, each of which includes a different number of processing units. The units in the first layer receive inputs from the outside world and are fully connected to units in the hidden layer. The units in the hidden layer, in their turn, are fully connected to output layer units, The units in the output layer produce the output of the MLP (see Figure 3.19).
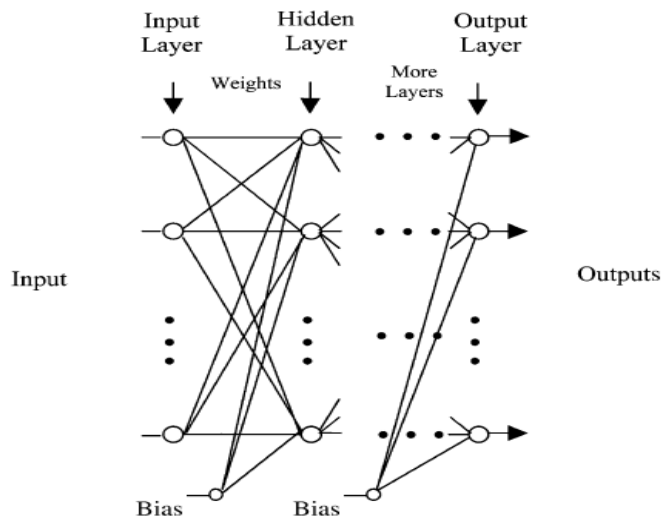
Figure 3.19 Architecture of the MLP network.

Learning Algorithms: An ANN should first be trained in order to accomplish the desired task. This means that the values of the connection weights are to be adjusted so that the network would produce the correct output for each given input pattern. The proper weights are determined under the control of a training algorithm. There are a large number of training algorithms and their variants (Haykin, 1999). It should be noted that the ultimate aim of training a neural network is not to force it to learn the training set perfectly. Instead good generalization ability is desired, this means producing correct output values for inputs which are not seen during the training process. The early stopping method (Amari, 1995; Demuth & Beale, 1998; Hagiwara & Kuno, 2000) is the approach used during training to increase the generalization performance of the network to avoid overtraining. In this method, a validation set, which is different from the training set, is chosen. During the training process, the validation error is used as the stopping criterion. As shown in Figure 3.20, when the validation error reaches its minimum, the training is finished. In this study, most of the training algorithms, except the last one which is based on the regularization method, are applied using the early stopping criteria.
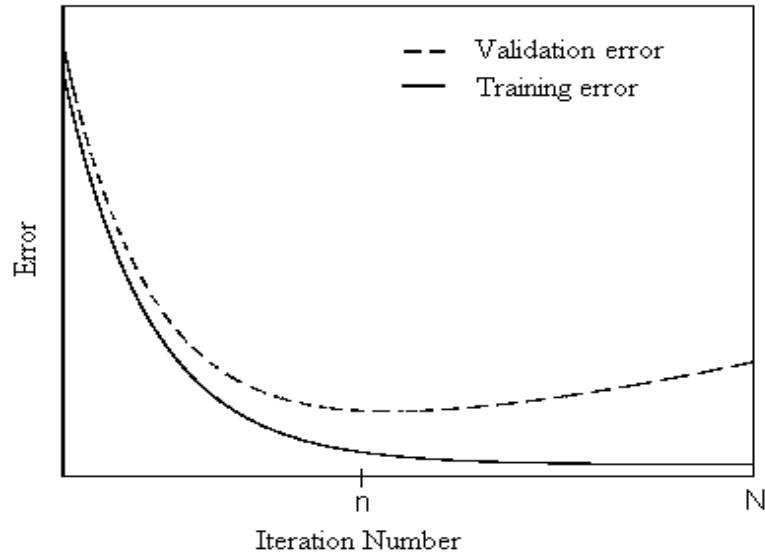
Figure 3.20 Training and validation errors versus iteration number.

Gradient Descent (GD): A gradient descent based optimization algorithm such as back propagation is the most common method used to adjust the connection weights in MLP iteratively in order to minimize an error function (Bishop, 1995; Duda, Hart, & Stock, 2001; Hecht-Nielsen, 1989; Yu, Efe, & Kaynak, 2002). Generally the error function used is the Mean Square Error (MSE):

$$E_{MSE}(y) = \frac{1}{2}\sum_{j=1}^{N}(t_j - y_j)^2 \tag{3.23}$$

where *t* is the target, *y* is the output, and E_{MSE} is the error function.

The errors calculated at the output units are then propagated backward to units in other layers. In order to minimize the error occurred in backpropagation phase, the value of each weight is updated by

$$\Delta w_{ji}^{l} = \eta \delta_j^{(l)} y_i^{(l-1)} \tag{3.24}$$

where η is the learning rate and δ is the derivative of error function with respect to the weight, i.e.

$$\delta(n) = \frac{\partial E}{\partial w} \tag{3.25}$$

Gradient Descent with Adaptive Learning Rate (GDALR): In plain gradient descent, as described above, the learning rate is held fixed during the training phase. However, changing the learning rate during the training process is a method that could increase the performance of the network (Yu & Liu, 2002). In this variant of gradient descent, when the new error exceeds the previous one, the learning rate is decreased and the new weight and bias values are discarded. If, on the other hand, the new error is less than the old one, the learning rate is increased by

$$e_r(n) = \frac{E(n) - E(n-1)}{E(n)} \tag{3.26}$$

where *E(n)* is the current error, *E(n-1)* is the previous error and e_r is the relative factor. During the training process, the learning rate is changed according to relative factor

$$\text{for } e_r < 0,\ \eta(n+1) = \eta(n)(1 + u.e^{-e_r(n)})$$
$$\text{for } e_r > 0,\ \eta(n+1) = \eta(n)(1 - u.e^{-e_r(n)}) \tag{3.27}$$

where *η(n+1)* is the updated learning rate, *η(n)* is the previous learning rate and u is the relative control parameter (0<u<1).

Levenberg-Marquart (LM) Algorithm : The LM method shows the fastest convergence during the training process based on gradient descent methods because it acts as a compromise between the stability of the first-order optimization methods (e.g., steepest-descent method) and the fast convergence properties of the second-order optimization methods (e.g., Gauss-Newton method) (Hagan & Menhaj, 1994; Chen, Han, Au, & Tham, 2003). When training with the LM method, the update of the weights are obtained as follows

$$\Delta w = [J^T J + \lambda I]^{-1} J^T e \qquad (3.28)$$

where J is the Jacobian matrix, λ is the learning parameter, and e is the sum of error squares.

Regularization Method: Another method that does not use early stopping but also increases the generalization performance of an ANN is the regularization method (Amari, 1995; Chan, Ngan, Rad, & Ho, 2002; Demuth & Beale, 1998; Hagiwara & Kuno, 2000). In this method, a penalty term is added to the error function as shown below

$$\tilde{E} = E + v\Omega \qquad (3.29)$$

$$\Omega = \frac{1}{n} \sum_{i=1}^{n} w_i^2 \qquad (3.30)$$

where E is the mean square error function, v is the control parameter of the penalty term, and Ω is the penalty term. Using this method has the similar effect of applying early stopping during the training process.

3.7.3 Modular Classifiers

In many of pattern recognition applications, it is clear that there is no optimal approach for classification (Duda, Hart, & Stock, 2001). Therefore, there have recently been widespread interests in the usage of multiple models for pattern classification. The aim is to solve a complex problem by dividing it into simpler problems whose solutions can be combined to yield a final solution. A multiple classifier system combines an ensemble of diverse classifiers as shown Figure 3.21. The combination of diverse experts has better results than single classifiers (Bhaskar, Hoyle, & Singh, 2006). Such classifier models which consist of more than one classifier are variously called mixture of expert models, ensemble classifiers, modular classifiers, committee machine or sometimes pooled classifiers as shown in Figure 3.21. Committee Machine

is a supervised learning method based on divide-and-conquer principle. It divides input space into subspaces and combines individual results of each expert. Experts share common input or each expert uses diverse input. Each expert is trained differently (if a number of neural networks are used). Individual outputs of experts are then combined to find the overall output.
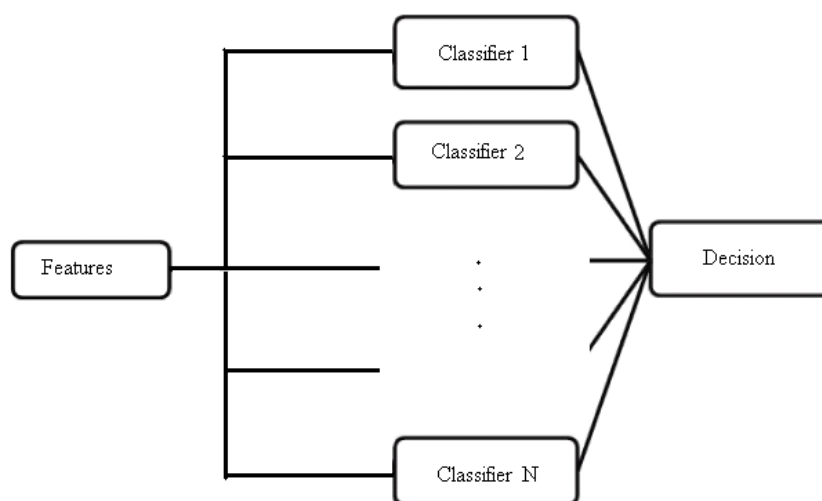
Figure 3.21 Simple structures of modular classifiers.

3.7.4 Performance Measures

Any diagnostic system decide either "1" or "0" which means "positive" or "negative". Each decision may be "true" or "false". So there are two kinds of responses for each decision. According to two-class case, there are four possible situations as a decision (Gibbons et al., 1997). If the instance is positive and it is classified as positive, it is assigned as true positive (TP); if it is classified as negative, then it is assigned as false negative (FN). If the instance is negative and it is classified as negative, it is assigned as true negative (TN); if it is classified as positive, it is assigned as false positive (FP).

Given a recognition system, a two-by-two decision matrix can be constructed according to decision of the test set. This matrix is also known as a contingency table or confusion matrix as shown in Table 3.1.

Table 3.1 Two-by-two decision matrix

		Output of the System	
		Normal	*Failure*
Experts	*Normal*	True Negative (*TN*)	False Positive (*FP*)
	Failure	False Negative (*FN*)	True Positive (*TP*)

The performance of a recognition system is measured by several parameters using the decision matrix (Eberhart & Dobbins, 1990). Sensitivity (SEN), selectivity (SEL), specificity (SPE), and overall accuracy (ACC) are the most used parameters.

Sensitivity is described as the ratio of the number of positive decisions correctly classified by the recognition system to the total number of positive decisions made by the expert. It shows the ratio of correctly classified abnormal patterns to abnormal pattern.

$$Sensitivity \ = \frac{TP}{TP+FN}x100\% \tag{3.31}$$

Specificity is described as the ratio of the number of negative decisions correctly made by the recognition system to the total number of negative decisions made by the expert. It shows the ratio of correctly classified normal patterns to normal pattern.

$$Specificity \ = \frac{TN}{TN+FP}x100\% \tag{3.32}$$

Selectivity is described as the ratio of the number of positive decisions correctly made by the recognition system to the total number of positive decisions made by the recognition system. It shows the ratio of correctly classified abnormal patterns to pattern which are classified as abnormal.

$$Selectivity = \frac{TP}{TP + FP} x100\% \qquad (3.33)$$

Overall Accuracy is the ratio of the total number of positive decisions and negative decisions correctly made by the recognition system to the all decisions.

$$OverallAccuracy = \frac{TP + TN}{TP + FN + TN + FP} x100\% \qquad (3.34)$$

When the system is a multi- class recognition system, the decision matrix can be also constructed according to decision of the test set. But calculation of the measures is a little different. For example, five different heartbeats (N, S, V, F, and Q) will be classified by the system. The decision matrix is constructed as shown in Table 3.2.

Table 3.2 Multi-class decision matrix

		Classifier Results				
		N	S	V	F	Q
Reference label	N	TP_N	FN_N	FN_N	FN_N	FN_N
	S	FP_N	TN	TN	TN	TN
	V	FP_N	TN	TN	TN	TN
	F	FP_N	TN	TN	TN	TN
	Q	FP_N	TN	TN	TN	TN

where

TP - True Positives:Number of heartbeats of an arrhythmia type correctly classified by the system

TN - True Negatives:Number of other arrhythmia heartbeats correctly classified by the system

FP - False Positives:Number of heartbeats incorrectly classified by the system

FN - False Negatives:Number of other heartbeats incorrectly classified by the system

SEN, SEL, SPE, and overall accuracy are given for each class as:

$$Sensitivity = \frac{TP_i}{TP_i + FN_i} x100\% \qquad\qquad (3.35)$$

$$Specificity = \frac{TN_i}{TN_i + FP_i} x100\% \qquad\qquad (3.36)$$

$$Selectivity = \frac{TP_i}{TP_i + FP_i} x100\% \qquad\qquad (3.37)$$

$$Over\ all\ accuracy = \frac{TP_1 + TP_2 + ... + TP_i}{All\ Beats} x100\% \qquad\qquad (3.38)$$

where the sub indices denote the heartbeat type.

Another performance evaluation tool is Receiver Operating Characteristic (ROC) analysis. It is a plot of sensitivity versus specificity values as shown in Figure 3.22. It is widely used in the medical applications to evaluate the performance of diagnostic tests. ROC curves contain a wealth of information for understanding and improving performance of classifiers. Area under the ROC curve is a measure of discrimination, or the performance measure of a diagnostic test. Overall accuracy or overall misclassification rate is not a useful measure when the disparity between classes is high (Alberg et al., 2004). Reported accuracies in this study are also measured by the area under the ROC curve as shown in Figure 3.22. An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. The traditional academic point system is used to evaluate the performance of a diagnostic test: if the area is between 0.90-1 it is excellent, between 0.80-.90 it is good, between 0.70-0.80 it is fair, 0.60-0.70 it is poor, and between 0.50-0.60 it is fail (Erkel & Pattynama, 1998; Fawcett, 2006; Alberg et al., 2004).
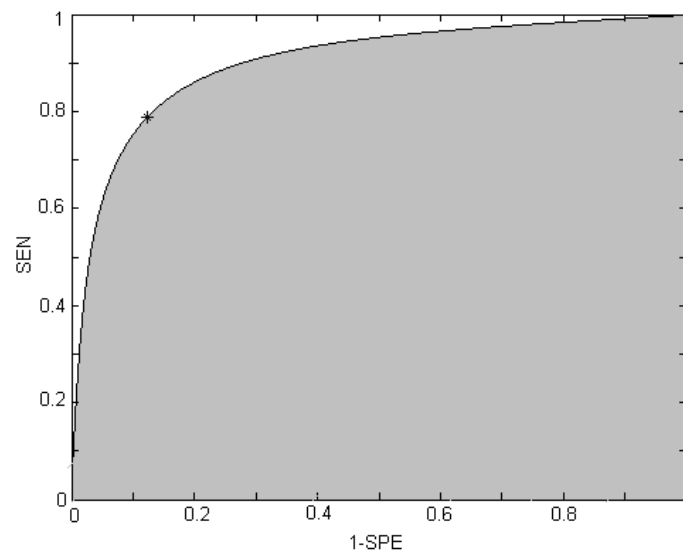
Figure 3.22 Area under ROC curve

There are many methods to determine area under ROC curve. Two methods are commonly used to compute the area of ROC curve: a non-parametric method based on constructing trapezoids under the curve to approximate the integral or the area under the curve and a parametric method, using a maximum likelihood estimator to fit a smooth curve to the data points.

CHAPTER FOUR
AUTOMATIC RECOGNITION OF EPILEPTIFORM EVENTS IN EEG

4.1 Introduction

Epilepsy, which is a very common neurological disorder, is defined as sudden, excessive and abnormal discharges in brain that may be caused by a variety of pathological processes of genetic or acquired origin. This disorder is often identified by sharp recurrent and transient disturbances of mental function or movements of different body parts (Göksan, 1998).

The evaluation of an EEG record for the detection of epileptic activity is usually performed by experienced electroencephalographers (EEGers) based on the visual scanning of the EEG record. However, this process is very time-consuming, error-prone, and too subjective (Ktonas, 1987). For this reason, there is an ever-increasing need for the development of automated systems to detect these abnormal wave patterns and there have been different attempts mainly based on artificial neural network (ANN) structures to automate the epileptic spike detection process. For example, (Webber, Litt, Wilson, & Lesser, 1994) utilized ANNs and mimetic methods. Kalaycı & Özdamar (1995) also used ANN-based systems for classification purposes and reported very satisfactory results. James, Jones, Bones, & Carroll, (1999) employed multi-stage approaches. Tarassenko, Khan, & Holt (1998) also used ANN-based recognition systems. (Özdamar & Kalayci, 1998 ; Özdamar, Yaylali, Jayaker, & Lopez, 1991) also used ANN-based systems. Dingle, Jones, Carroll, & Fright, (1993) used a multistage system to detect epileptiform activity. Nuh, Jazidie, & Muslim, (2002) utilized a different type of neural network, i.e. a wavelet neural network. Adjouadi et al. (2004) developed an algorithm using the Walsh transformation. Acır & Güzeliş (2004) utilized a two stage classification system based on support vector machine (SVM). Acır, Öztura, Kuntalp, Baklan, & Güzeliş, (2005) employed a two stage classification system based on the radial basis function network (RBFN). Exarchos, Tzallas, & Fotiadis, (2006) used a rule based

classification system. İnan & Kuntalp, (2007) used a two stage unsupervised classification system.

In some of these systems, the extracted waves were used as windowed raw input to the classification system; whereas, in others, specific waveform features are fed into the system as the input. The use of the parameterized approach reduces data load on the system and processing time and increases the performance of the classifier. In the parameterized approach, however, the success of the spike detection algorithm relies heavily on the proper selection of the features. These features may include frequency domain parameters like total power, or time domain parameters like amplitude, duration, and slope. Using raw data, on the other hand, avoids any data loss that parameterization techniques will inevitably introduce but provides the system with a high-dimensional input data. This could reduce the performance of the classifiers due to the curse of dimensionality effect. Therefore, both approaches have their own advantages and disadvantages. In this study, both the parameterized and raw forms of data are used as input and their effects are compared.

As the classification system, different multilayer perceptron (MLP) networks utilizing between 3 and 15 hidden neurons are constructed for the automatic detection of epileptic spikes in the EEG records for the diagnosis of epilepsy. For training the MLP networks, early stopping versions of backpropagation (Amari, 1995; Bishop, 1995; Demuth & Beale, 1998; Duda, Hart, & Stock, 2001; Hagiwara & Kuno, 2000; Hecht-Nielsen, 1989; Yu, Efe, & Kaynak, 2002), backpropagation with adaptive learning rate (Yu & Liu, 2002), Levenberg-Marquardt (LM) algorithms (Chen et al., 2003), and regularization methods (Amari, 1995; Hagiwara & Kuno, 2000) are used. The aim of using early stopping and regularization is to increase the generalization performance of the classifier.

The inputs used for the training of the networks are constructed as follows. As the first step, the individual spike-like waves are extracted from all records. These waves include both epileptic spike waves and non-epileptic spike waves which are similar to epileptic spikes. From here on, both of them will be referred to as spike-like

waves. As the next step, either of the following two methods is applied. In the first method, specific morphologic features, which are extracted by numerical techniques from the wave patterns, are given as input to the detection system. In the other method, the raw waveform is directly presented as input to the system after a proper scaling and windowing process. The aim here is to compare the effects of using either raw data or extracted features. However, in addition to the original forms of both raw data and extracted features, the networks are also fed with their transformed versions which are obtained by using different transformation methods. The performances of all the constructed classifiers are then evaluated and compared based on sensitivity, selectivity, and specificity measures since these parameters have been accepted and used as the standard for EEG spike detection algorithms (Pang, Upton, Shine, & Kamath, 2003).

In this chapter, the proposed automated recognition systems are described in detail. The general block diagrams of the constructed systems are shown in Figure 4.1. At first, single MLP based classification system is constructed. Then, a multi-stage classification system is constructed.
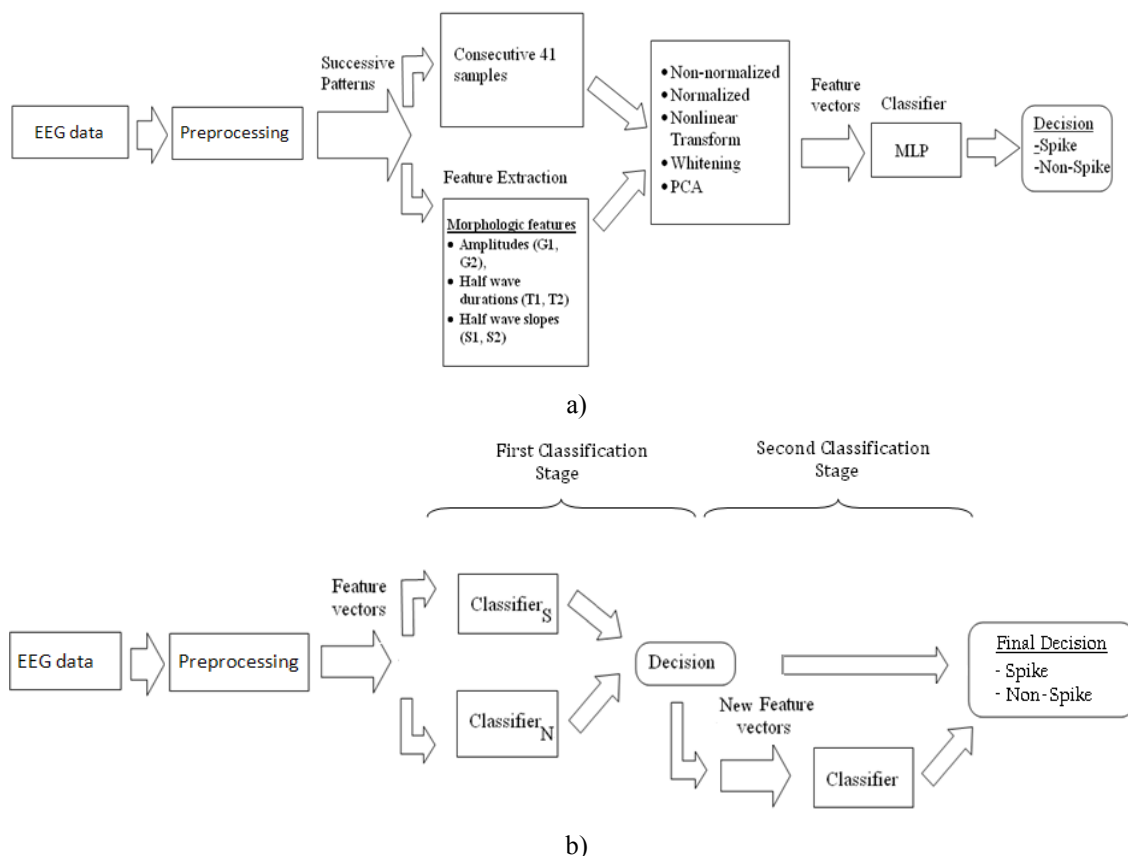
Figure 4.1 Block diagram of the constructed classification systems for spike detection in EEG a) Single MLP based classification system, b) multi-stage classification system.

4.2 EEG Data

4.2.1 Data acquisition and its properties

The EEG data used in this study are obtained from the Neurology Department of Dokuz Eylül University Hospital, İzmir, Turkey. The EEG data are acquired with Ag/AgCl disk electrodes placed using the 10–20 international electrode placement system. The EEG signals are recorded from 19 channels at a sampling frequency of 256 Hz and then band-pass filtered between 1 and 70 Hz. These EEG recordings are initially labeled for spikes by two experienced EEGers. Only the wave patterns which are labeled as epileptic spikes by both EEGers are accepted as spike waves.

4.3 Pre-processing

The first step of the pre-processing stage is the mean removal by which the average of each individual EEG record is calculated and then subtracted from the signal itself. The second step is the determination of the locations of all peaks (positive and negative) in the record since the spike waves generally reveal themselves in the form of transient spikes with a pointed peak. The peaks obtained are then processed based on the approach described in (Acır et al., 2005): First of all, the amplitude differences between each peak are calculated. This gives the slope between these data points. If the slope is positive, the signal is increasing; otherwise, the signal is decreasing. Whenever the slope changes from positive to negative, this means that it is a positive peak. In the same manner, if there is a change in the slope from negative to positive then it is a negative peak.

In order to eliminate the artifacts from the signal and ignore irrelevant small changes with high frequencies, the following simple algorithm is used. If the length of a segment between two adjacent peaks is shorter than the length of the previous and next segments and if the duration of this segment is shorter than 20 ms and its amplitude is smaller than $2\mu V$, then the peak is accepted as an artifact and eliminated (Acır et al., 2005). This filtering process also eliminates mains electricity interference and sharp, short-duration waves similar to spike activity resulting from the movement of the patient.

After the application of the processes mentioned above, a total of 119 spike-like waves are extracted from the available EEG data. Based on the views of the EEGers, 39 of these waves are labeled as epileptic spikes and 80 as non-epileptic waves. In order to improve the generalization performance of the MLP based classifiers, "training with noise" method is used whereby new training samples are generated by adding zero mean Gaussian noise with a variance of 10% to the available 119 spike-like waves (Amari, 1995; Bishop, 1995; Duda, Hart, & Stock, 2001; Minnix, 1991; Nicholson, 2002; Tsukuda, Kurokawa, & Mori, 1995, Kutlu, Kuntalp, & Kuntalp, 2006). As a result of this process, a new set with 390 spike and 400 non-spike waves

is created. These samples are used for training purposes only as if they were normal training data sampled from the same source distributions. The real spike-like waves, i.e. 119 waves without Gaussian noise added, are used for testing the classifiers.

4.4 Feature Extraction and Transformation

4.4.1 Raw EEG

For applying the extracted waves to the classifiers as raw input, a window with a length of 41 data points is used in which the peak of the wave is located at the 11th point. The reason for choosing this value is that it corresponds to approximately 150-160 ms of the EEG signal and is an average value for the epileptic spike duration. Several samples of spike-like waves are given in Figure 4.2.
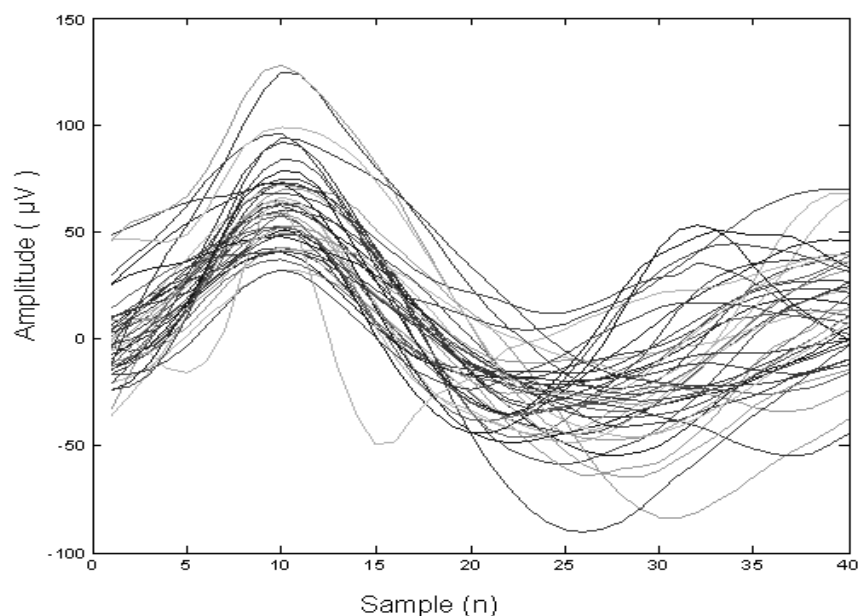
Figure 4.2 Samples of spike-like waves.

4.4.2. Morphological Features

Six morphologic features are acquired from each spike-like wave. These features include first half wave amplitude (G1), second half wave amplitude (G2), first half

wave duration (T1), second half wave duration (T1), first half wave slope (S1=G1/T1), and second half wave slope (S2=G2/T2) (see Figure 4.3).
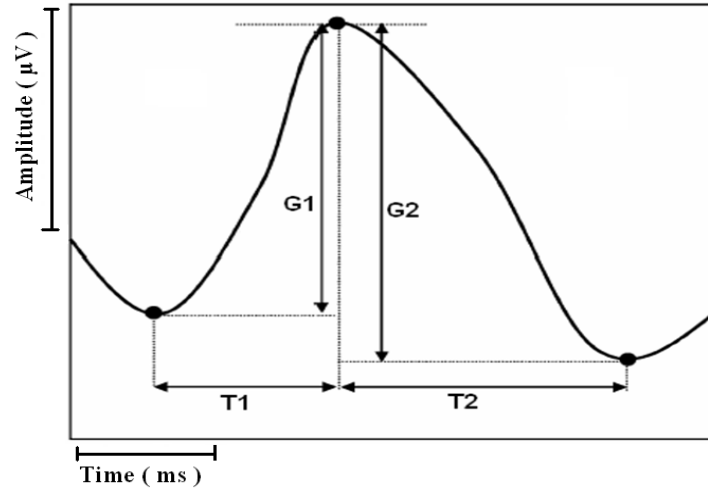
Figure 4.3 Morphologic features obtained from candidate waves.

4.4.3 Feature Transformation

Each spike candidate (i.e. a spike-like wave), whether it is represented by 41 consecutive points or by the six morphologic features, is further transformed by using four different transform methods: i) Normalization, ii) Nonlinear transformation, iii) Principal Component Analysis (PCA), and iv) Whitening transformation. These transformation methods are described in section 3.1.4.

4.5   Classification

4.5.1 Recognition with a MLP based classifier

The classifiers implemented in this study are developed by using Matlab. Classification procedure is performed off-line on data stored on the hard disk. Seven different MLP networks are constructed utilizing between 3, 5, 7, 9, 11, 13, and 15 hidden units, respectively (Kutlu, İşler, Kuntalp, & Kuntalp, 2006). Each network is trained with all four different training methods described in Section 3. For each method, 10 different data sets are used; these data sets consist of original and transformed versions of both raw data and extracted features. This corresponds to a

total of 280 (7 MLP structures x 4 training algorithms x 10 data sets) different MLP classifiers. All the network structures are trained five times starting from different initial conditions and tested separately. The outcome of each MLP classifier is based on the average of these 5 testing results.

The performance of each classification system for the automatic detection of epileptic spikes is obtained using three standard statistical measures: sensitivity, specificity, and selectivity.

However, instead of using these three measures separately, two different combinations of them are used: (1) the average of sensitivity and selectivity, and (2) the average of sensitivity and specificity. This way, it would be possible to directly compare the performances of our systems with other classifiers given in the literature using different performance measures (Kutlu, Kuntalp, & Kuntalp, 2009b).

Of all the 280 classifiers, the one that displays the best performance in terms of both average sensitivity/selectivity (ASenSel) (90.8%) and average sensitivity/specificity (ASenSpe) (94.9%) measures is found to be the one which has 15 hidden units and is trained with the Gradient Descent Algorithm with Early Stopping (with Adaptive Learning Rate) method using whitened transformed data. All the classifiers trained with other training algorithms using whitened data also revealed very high performances.

4.5.2 Recognition with a multi-stage classifier

Multi-stage classification procedure is performed both to reduce the computation time of the entire classification procedure and to increase the overall detection performance. The first stage classification is to eliminate trivial non-spikes and also to determine definite spikes and to determine spike like non-spikes. The peaks are classified into three groups (two dimensional description is shown in Figure 4.4): epileptiform waves (Region I), non- epileptiform waves (Region II), and possible epileptiform waves and possible non- epileptiform waves (Region III). For this

purpose, two classifiers are used in the first stage: One classifier determines definite epileptiform waves; the other one determine definite non-epileptiform waves. The Region III is represent possible epileptiform waves and possible non- epileptiform waves which are also intersection of the first stage classifiers. Neural networks are used in all stages of the classifier which are trained with adaptive learning rate algorithm. In the first stage, classifiers used 1 hidden unit. Raw ECG signal is used in the first stage of classifier as input vector. In the second MLP classifier is utilized 15 hidden units. Whitened parameters are used in the second stage. All classifiers are trained several times with different initial values. In the first stage, the best selectivity measure is used and in the second stage overall accuracy is used as performance criteria because of the aim of the classifier.
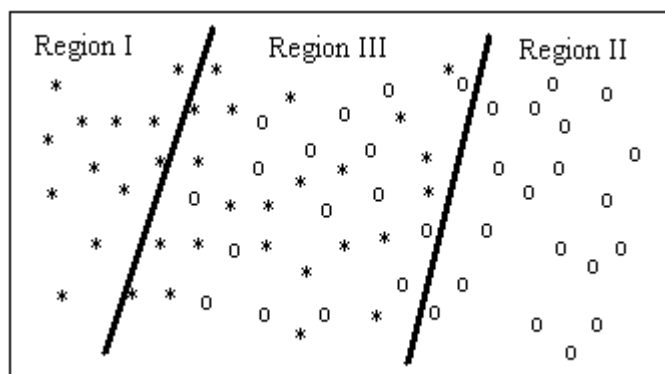
Figure 4.4 Two dimensional description of classification of first stage: The peaks are classified into three subgroups in the first stage: (*) represents spikes, (o) represents non-spikes, Region I represents definite spikes, Region II represents definite non-spikes, and Region III represents possible spikes and possible non-spikes.

4.6 Result and Discussion

Many MLP structures are constructed. Figure 4.5-4.8 and Figure 4.9-4.12 show in detail the effect of hidden layer size on the performance of the classifiers for raw data and extracted features, respectively. As can be seen from these figures, there is no direct correlation between the number of hidden units and the performance obtained.
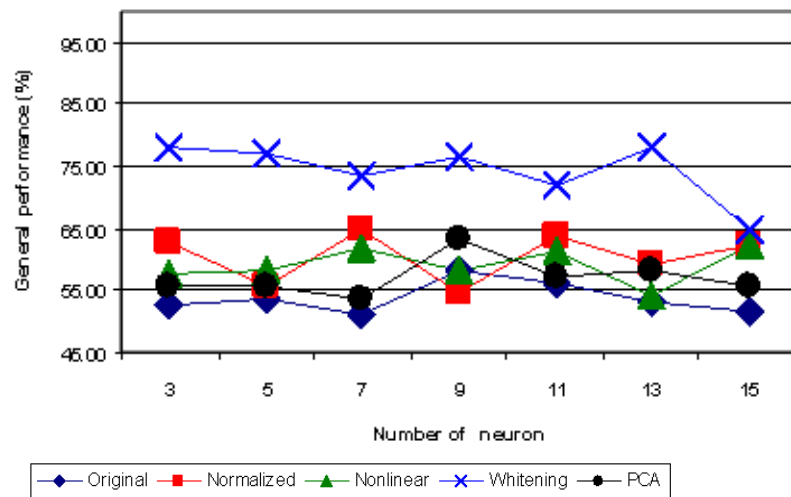
Figure 4.5 The average accuracy values versus hidden neuron size using raw data of 41 samples with Gradient descent.
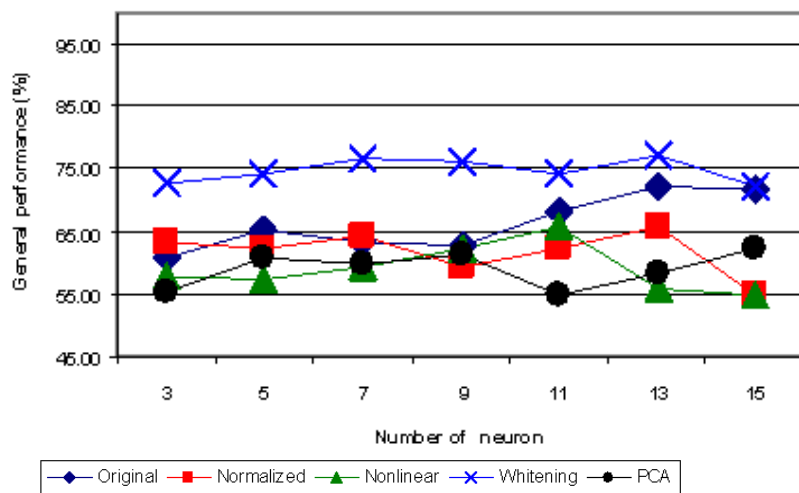
Figure 4.6 The average accuracy values versus hidden neuron size using raw data of 41 samples with Gradient descent with adaptive learning rate.
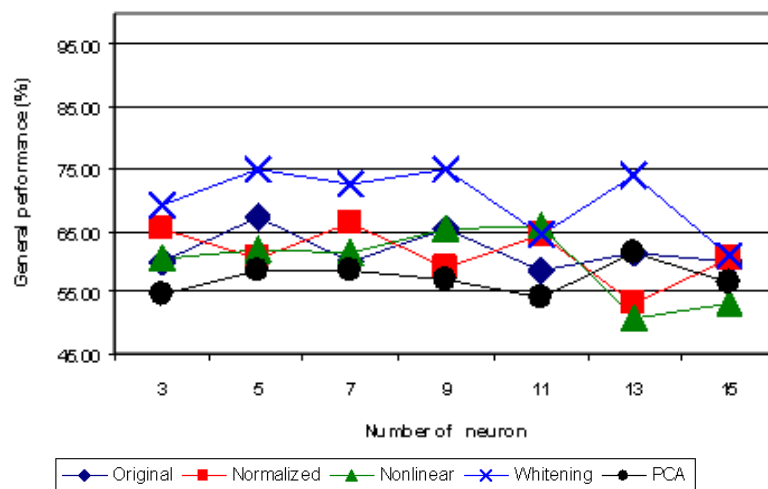
Figure 4.7 The average accuracy values versus hidden neuron size using raw data of 41 samples with LM.
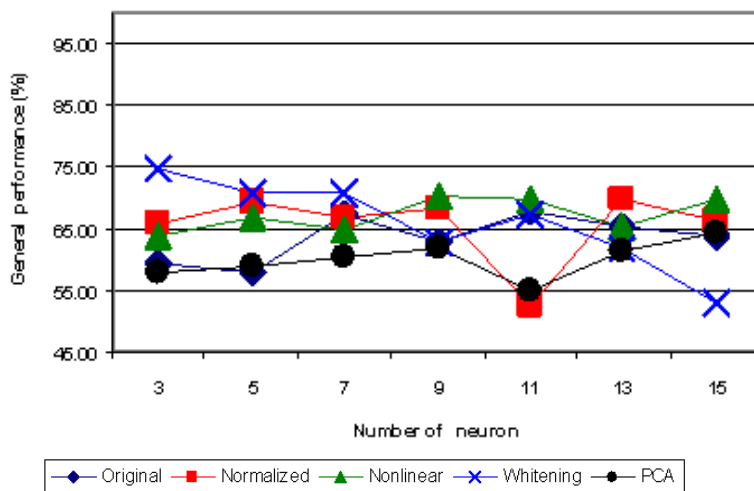
Figure 4.8 The average accuracy values versus hidden neuron size using raw data of 41 samples with regularization algorithm.
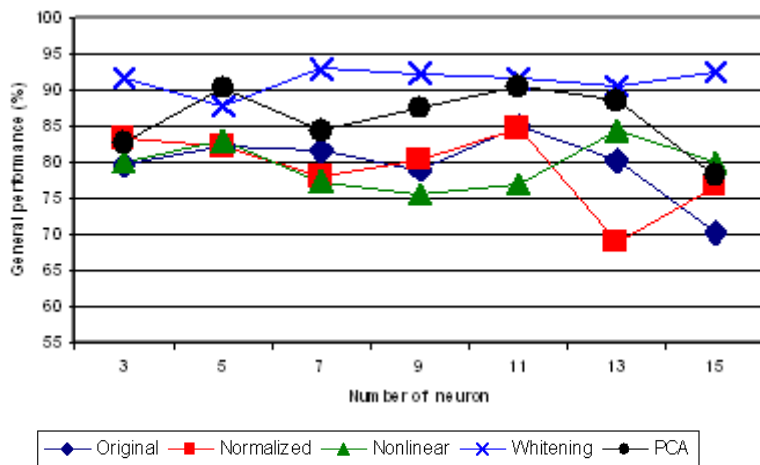
Figure 4.9 Using 6 features, the average accuracy of GD algorithm versus number of neuron at hidden layer.
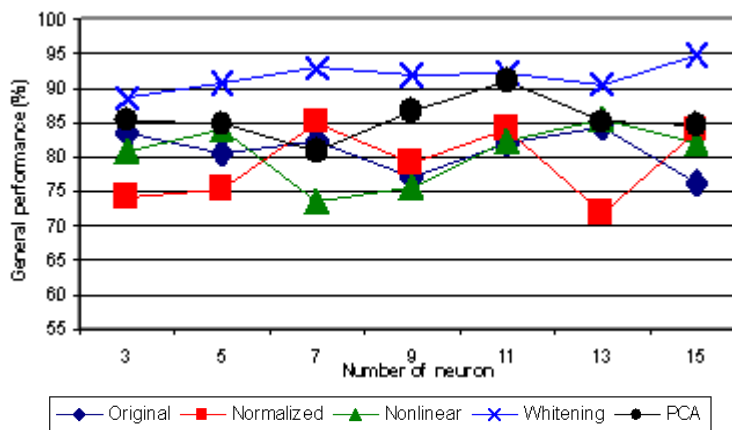
Figure 4.10 Using 6 features, the average accuracy of GD algorithm with adaptive learning rate, versus number of neuron at hidden layer.
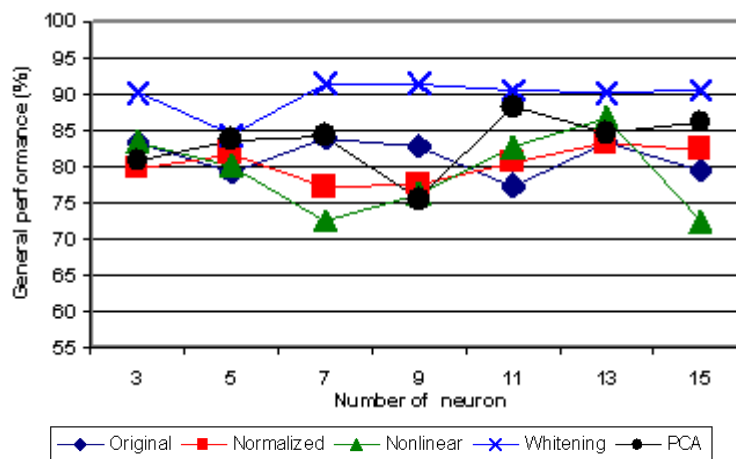
Figure 4.11 Using 6 features, the average accuracy of LM algorithm versus number of neuron at hidden layer.
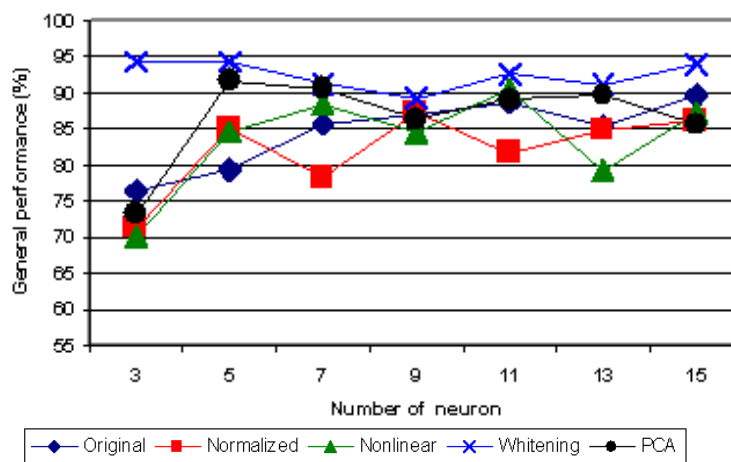
Figure 4.12 Using 6 features, the average accuracy of regularization methods versus number of neuron at hidden layer.

The best performance measures obtained for different structures constructed in this study are shown in Tables 4.1 and 4.2. Each value in these tables represents the best performance from among the ones obtained by using different hidden unit numbers. As can be seen from these results, the classifiers trained and tested with extracted features instead of raw data perform better. Additionally, the classifiers displaying the best performances in terms of both measures are the ones trained with whitened transformed input data.

Table 4.1 Measures for raw data sets

Training Algorithm	Original		Normalized		Non-linear transformed		PCA transformed		Whitened	
	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)
Regularization	67.65	72.66	69.69	72.11	70.14	69.63	64.46	70.68	74.80	70.92
GD with Earlystop	58.13	70.99	64.83	71.72	62.07	71.34	63.33	70.57	78.33	79.30
GDALR with Earlystop	72.42	72.60	65.78	74.26	65.99	72.83	62.59	71.25	77.42	76.16
LM with Earlystop	67.18	67.35	66.43	71.77	65.88	71.48	61.19	71.66	75.07	71.04

ASenSpe: average of sensitivity-specificity, ASenSel: average of sensitivity-selectivity

Table 4.2 Measures for extracted features

Training Algorithm	Original		Normalized		Non-linear transformed		PCA transformed		Whitened	
	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)	ASenSpe (%)	ASenSel (%)
Regularization	89.56	79.89	87.10	76.66	90.35	81.07	91.66	83.31	94.29	89.94
GD with Earlystop	85.10	74.41	84.70	73.97	84.47	73.40	90.57	83.62	93.01	89.20
GDALR with Earlystop	84.34	73.19	85.09	75.57	85.58	74.95	91.10	83.87	94.91	90.78
LM with Earlystop	83.85	72.82	83.23	72.28	86.70	76.41	88.17	81.29	91.43	88.17

ASenSpe: average of sensitivity-specificity, ASenSel: average of sensitivity-selectivity

Of all the 280 classifiers, the one that displays the best performance in terms of both average sensitivity/selectivity (ASenSel) (90.8%) and average sensitivity/specificity (ASenSpe) (94.9%) measures is found to be the one which has 15 hidden units and is trained with the Gradient Descent Algorithm with Early Stopping (with Adaptive Learning Rate) method using whitened transformed data.

The second classification system is a multi-stage system. MLP classifiers are used in the constructed system. The best performance in terms of both average sensitivity/selectivity (ASenSel) and average sensitivity/specificity (ASenSpe) measures are 93.7% and 95.6%, respectively.

In addition to the above mentioned performance criteria, the training times of the constructed systems are also observed. It is seen that training times vary according to the training algorithm used and the number of units in the hidden layer. The fastest algorithm in this respect is found to be the back-propagation algorithm with adaptive learning rate using early stopping criteria.

Epilepsy is a very common neurological disorder leading to disturbing seizures. In the inter-ictal period, i.e. in between seizures, epileptic transients, in the form of spikes and sharp waves, are typically observed in the EEG recordings. For this reason, the inter-ictal spike detection plays a crucial role in the diagnosis of epilepsy. In the present study, spike waves are aimed to be detected from EEG records by different classifiers constructed using MLP structures utilizing different number of hidden neurons, different training algorithms, and different preprocessing of input data.

ANN-based detection systems found in the literature generally use two approaches. In the first approach, the raw EEG data is directly used as the input to the ANN. In the second approach, some features are extracted from the EEG records and fed into the ANN as the input. Using the parameterized approach has the advantage of using fewer inputs. However, it requires the correct definition of input features that would be selected for the detection of spikes. As a result, in order to find the best set of features for optimum detection, this approach does not make full

use of the power of an ANN. Nevertheless some researchers have successfully made use of the extracted features for the detection of epileptic activity (Gabor & Seyal, 1992; Webber et al., 1994). On the other hand, using raw EEG data has the advantage of avoiding the possible false classification that could arise from data loss in the parameterization of the EEG data (Ko & Chung, 2000; Özdamar & Kalayci, 1998). However, the performance of the classifiers trained with this approach could also be reduced due to the curse of dimensionality effect since the dimension of input data is generally very high. In this study, both the parameterized and raw forms of data are used as the input and the results obtained are compared.

For training the MLP networks, early stopping versions of backpropagation, backpropagation with adaptive learning rate, Levenberg-Marquardt (LM) algorithm, and regularization methods are used. The aim of using early stopping and regularization method is to obtain better generalization performance from the classifiers. In addition, in order to increase the generalization performance of the constructed systems, training with noise method is used. For this purpose, zero mean Gaussian noise with a variance of 10% is added to real data. While these new samples are used only for training the networks as if they were normal training data sampled from the same source distributions, the real data is reserved for the testing process. Furthermore, to be able to observe the effect of applying preprocessing on input data, different linear and nonlinear transformations are used on the available data set. The classifiers constructed are also trained and tested with these transformed data. The overall performances of the constructed classification systems are computed based on the average sensitivity/specificity and average sensitivity/selectivity measures. The average sensitivity/specificity is also represents the measures of ROC area.

The first outcome obtained from this study is that all the classifiers perform better when trained and tested with extracted features instead of raw data. The second important result is that the classifiers displaying the best performances in terms of both measures are the ones trained with whitened transformed data. Being specific, the classifier that displays the best performance in terms of both average sensitivity/selectivity (90.8%) and average sensitivity/specificity (94.9%) measures is

found to be the one which has 15 hidden units and is trained with the Gradient Descent Algorithm with Early Stopping method using whitened transformed data. All the classifiers trained with the other training algorithms using whitened data also reveal very close performances. The second classification system is constructed as a multi-stage system using MLP network that has the best performance measure. The best performances in terms of both average sensitivity/selectivity (ASenSel) and average sensitivity/specificity (ASenSpe) measures become 93.7% and 95.6%, respectively.

The comparison of the systems constructed in this study with similar detection systems given in the literature is not straightforward due to the varieties in the network types, architectures, data sources (e.g., channel numbers, displaying montages, degrees of artifact presence, recording type, status of subject), and performance measures used. Nevertheless, some objective conclusions can be drawn. For example, Webber et al. (1994) have tested their system on parameterized EEG records and reported 74% sensitivity and 74% selectivity values by using mimetic and ANN methods. Tarassenko, Khan, & Holt (1998) have studied both time and frequency domain parameters of the EEG signal and reported the results of 90.5% and 93.2% for specificity and sensitivity, respectively. Özdamar et al. (1991) have reported good results for sensitivity (90%); but selectivity was found to be relatively low at about 69% using raw EEG data. Dingle et al. (1993) have given a very good result for selectivity (100%) although the sensitivity was relatively low at 58%. James et al. (1999) have also reported the results of 82% and 55% for selectivity and sensitivity, respectively. Kalayci & Özdamar, (1995) have reported the results of 93.3% and 87.3% for specificity and sensitivity, respectively, using the wavelet transform. Acır et al. (2005) have reported that the best performance was obtained with the RB-SVM method providing an average sensitivity of 89.1% and average selectivity of 85.9%. Nuh, Jazidie, & Muslim, (2002) have studied wavelet neural network which combines wavelet analysis and ANN in a single algorithm and obtained 82% sensitivity and 90.4% selectivity values. Adjouadi et al. (2004) have reported 79% and 85% for sensitivity and selectivity indices, respectively, using the Walsh transformation. Acır & Güzeliş (2004) have reported the results of 90.3% for

sensitivity and 88.1% for selectivity analysis using a two stage classification system based on only SVM. Acır et al. (2005) have obtained the values of 91.1% and 89.2% for sensitivity and selectivity measures, respectively, using an RBFN system. Exarchos, Tzallas, & Fotiadis (2006) have used morphological and power spectrum features and have reported 86% sensitivity, 92% specificity and 83% selectivity values by using a rule based classification system. İnan & Kuntalp, (2007) have reported 93% sensitivity, 74% selectivity and 28% specificity values using a two stage unsupervised classification system.

CHAPTER FIVE

AUTOMATIC RECOGNITION OF ARRHYTHMIAS IN ECG RECORD

5.1 Introduction

The goal of this study is to develop a robust system that will be capable of classifying a large number of arrhythmia types with higher accuracy than the other methods in the literature. The best set of features that could be used will be selected by a genetic algorithm from a large pool of candidate features including morphologic, spectral, time-frequency, and higher-order statistical ones. This project will differ radically from all the other systems in the literature, by taking into account all possible features simultaneously and by determining the best features for each arrhythmia type from among them.

Electrocardiography is an important tool in diagnosing the condition of the heart. It provides valuable information about the functional aspects of the heart and cardiovascular system. Early detection of heart diseases/abnormalities can prolong life and enhance the quality of living through appropriate treatment. Therefore, numerous research and work analyzing the electrocardiogram (ECG) signals have been reported.

In the literature, many researchers have addressed the problem of automatic detection and classification of cardiac rhythms. In most of the studies, MIT-BIH ECG database is used. Some techniques are based on the detection of a single arrhythmia type and its discrimination from normal sinus rhythm, or the discrimination between two different types of arrhythmia (Afonso & Tompkins, 1995; Ham & Han, 1996; Chen, Clarkson, & Fan, 1996; Thakor, Zhu, & Pan, 1990; Clayton, Murray, & Campbell, 1993; Clayton, Murray, & Campbell, 1994; Yang, Device, & Macfarlane, 1994). Other classes of proposed methods for arrhythmia detection and classification are based on the detection of different heart rhythms and their classification into two or three arrhythmia types and the normal sinus rhythm (Thakor, Natarajan, & Tomaselli, 1994; Khadra, Al-Fahoum, & Al-

Nashash, 1997; Minami, Nakajima, & Toyoshima, 1999; Al-Fahoum & Howitt, 1999; Zhang, Zhu, Thakor, & Wang, 1999; Wang, Zhu, Thakor, & Xu, 2001; Owis, Abou-ZiedYoussef, & Kadah, 2002). Another field of interest is the ECG beat-by-beat classification, where each beat is classified into several different rhythm types (Hu, Palreddy, & Tompkins, 1997; Lagerholm, Peterson, & Braccini, 2000; Dokur & Olmez, 2001; Osowski & Linh, 2001; Hosseini, Reynolds, & Powers, 2001; Tsipourasa, Fotiadisa, & Sideris, 2005; Übeyli, 2007; Chazal & Reilly, 2006; Besrour, Lachiri, & Ellouze, 2008; Melgani & Yakoub, 2008; Arif, Akram, & Afsar, 2009; Moazzen, Ahmadzadeh, Doost-Hoseini, & Omidi, 2009; Nasiri, Naghibzadeh, Yazdi, & Naghibzadeh, 2009; Yong, Wenxue,& Yonghong, 2009; Raghav & Mishra, 2008; Chia-Hung Lin, Chao-Lin Kuo, Jian-Liung Chen, & Wei-Der Chang, 2009; Osowski, Siroic & Siwek, 2009). These methods can classify more arrhythmic beat types.

In all these studies, the researchers used a variety of features to represent the ECG signal and a number of classification methods. The features has been based on higher order statistics (Alliche & Mokrani, 2003; Besrour, Lachiri, & Ellouze, 2008; Khadra, Al-Fahoum, & Binajjaj, 2005; Osowski, Hoai, & Markiewicz, 2004; Osowski & Linh, 2001; Torun, İşler, Kuntalp, & Kuntalp, 2006), wavelet transform (Acır, 2005; Arif, Akram, & Afsar, 2009; Song et al., 2005), Fourier transform (Acır, 2005; Heidari, Shahidi, Aminian, & Sadati, 1998; Minami, Nakajima, & Toyoshima, 1999; Minami, Nakajima, & Toyoshima, 1997), principle component analysis (Nadal & Bossan, 1993; Nasiri et al., 2009), Helmite function coefficients (Braccini et al. 1997; Osowski, Hoai, & Markiewicz, 2004), fractal properties (Chia-Hung Lin et al., 2009; Raghav & Mishra, 2008), morphological features (RR-interval, QRS complex, QRS duration in time, T wave duration in time, P wave flag, T-wave segment, etc. (Besrour, Lachiri, & Ellouze, 2008; Braccini et al. 1997; Hu, Palreddy, & Tompkins, 1997; Hosseini, Reynolds, & Powers, 2001; Minami, Nakajima, & Toyoshima, 1999; Melo, Caloba, & Nadal, 2000; Nasiri et al., 2009; Osowski & Linh, 2001; Nadal & Bossan, 1993).

Moreover, different systems are used as classifier in the researches. For instance, multi-layer perceptron are used as a classifier by (Acır, 2005; Chia-Hung Lin et al., 2009; Hosseini, Reynolds, & Powers, 2001; Minami, Nakajima, & Toyoshima, 1997; Minami, Nakajima, & Toyoshima, 1999; Nadal & Bossan, 1993;Osowski & Linh, 2001; Torun, İşler, Kuntalp, & Kuntalp, 2006; Song et al., 2005; Yong, Wenxue,& Yonghong, 2009); support vector machine are performed by (Acır, 2005; Besrour, Lachiri, & Ellouze, 2008; Melgani & Yakoub, 2008; Song et al., 2005; Nasiri et al., 2009; Osowski, Siroic & Siwek, 2009); mixture of experts approach (Hu, Palreddy, & Tompkins, 1997), fuzzy logic (Osowski & Linh, 2001; Song et al., 2005), RBF (Heidari, Shahidi, Aminian, & Sadati, 1998), *K*-nearest neighbor (Arif, Akram, & Afsar, 2009; Karimifard, Ahmadian, & Khoshnevisan, 2006; Isler & Kuntalp, 2007), and SOM (Braccini et al. 1997; Hosseini, Reynolds, & Powers, 2001) are also used as classifier.

The structure of the heartbeat classification system is shown in Figure 5.1. The first step of the constructed system consists of preprocessing. In this step, removing base line wander effect and power line interface, QRS detection, and segmentation of the raw data are performed. In the feature extraction step, higher order statistics of wavelet packet decomposition (WPD) coefficients, frequency domain features, morphological features and higher order statistic features are extracted. Then two classification stages follow. They both have feature selection and classification steps. At the first stage, all heartbeats are classified into five main groups. In other word, all heartbeats in database are grouped into five classes according to "Association for the Advancement of Medical Instrumentation" (AAMI) standards (Chazal, O'Dwyer, & Reilly, 2004). In the feature selection step, optimum feature set is determined for each beat group using feature selection algorithm. Genetic algorithm and sequential floating search methods are used as feature selection algorithm. For the second stage of classification, each main group are separated into subgroups of heartbeat types. For each subgroup, optimum feature set is determined for each beat type using feature selection algorithm. Combined KNN classifier and also combined MLP classifier are implemented and the classification performance of each is compared

with the other. For the second stage of classification, combined KNN classifiers are constructed to separate main groups into subgroups of heartbeat types.
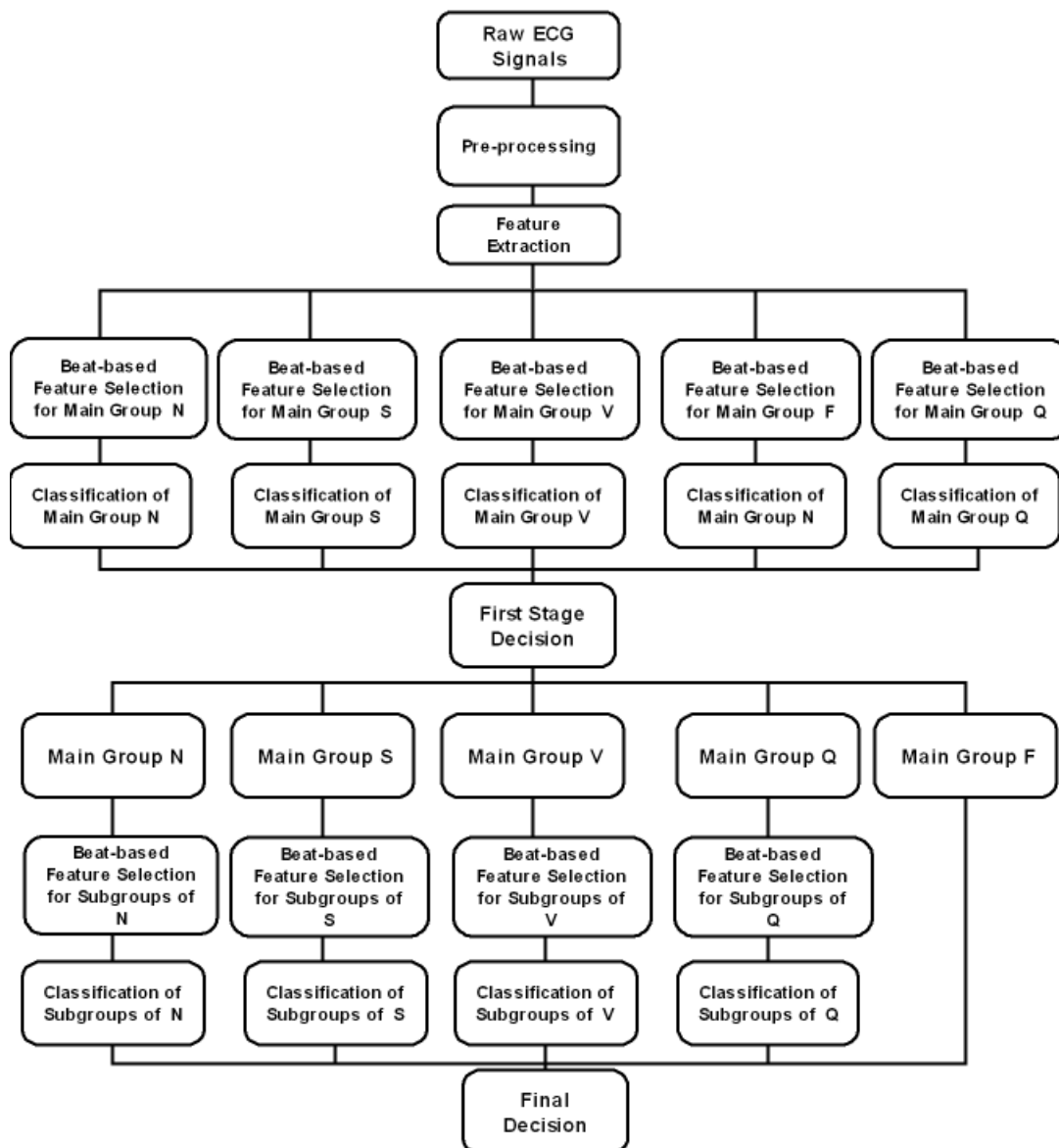
Figure 5.1 The general block diagram of the constructed system.

All ECG recordings will be provided from the publicly available MIT/BIH database which has become a standard for the ECG researchers in recent years. MATLAB will be the primary development platform for all the algorithms to be developed. The performance of the systems developed will be calculated and compared both with each other and with other systems in the literature based on the sensitivity, selectivity, specificity, and ROC area measures.

5.2. ECG Data

5.2.1. Data Acquisition and Its Properties

The source of the ECG records is obtained from the MIT-BIH Arrhythmia Database (Goldberger et al., 2000). The database is a set of over 4000 long-term Holter recordings that are obtained by the Beth Israel Hospital Arrhythmia Laboratory. The database contains 48 recording. The subjects are 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. But only 30 minutes long of the records is online available in pyhsionet freely. Each record contains two ECG lead signals denoted lead A and B. In 45 records of lead A are modified-lead II and the other three are lead V5. 40 records of Lead B are lead VI and the others are either lead II, V2, V4, or V5. The data are band-pass filtered at 0.1 - 100 Hz and sampled at 360 Hz. The lead A (lead II) is used in this study because of more clear QRS complexes.

Each record is annotated by two or more cardiologists independently. Amount of labeled beats of arrhythmias in the database is shown in the Table 5.1, and samples of each heartbeat type are shown in Figure 5.2.

Table 5.1 Mapping of MIT-BIH arrhythmia database heartbeat types to the AAMI heartbeat classes and amount of arrhythmias in the database

AAMI heartbeat classes	Labeled	Classes	Amount
Non-Ectopic Beats (**N**)	1	Normal Beat (NORMAL)	75054
	2	Left Bundle Branch Block Beat (LBBB)	8075
	3	Right Bundle Branch Block Beat (RBBB)	7259
	4	Nodal (Junctional) Escape Beat (NESC)	229
	5	Atrial Escape Beat (AESC)	16
Supraventricular Ectopic Beats (**S**)	6	Aberrated Atrial Premature Beat (AAPB)	150
	7	Premature or Ectopic Supraventricular Beat (PESB)	2
	8	Atrial Premature Contraction (APC)	2545
	9	Nodal (Junctional) Premature Beat (NPB)	83
Ventricular Ectopic Beats (**V**)	10	Ventricular Flutter Wave (VF)	472
	11	Ventricular Escape Beat (VESC)	106
	12	Premature Ventricular Contraction (PVC)	7129
Fusion Beats (**F**)	13	Fusion of Ventricular And Normal Beat (FUSION)	803
Unknown Beats (**Q**)	14	Paced Beat (PACE)	7028
	15	Unclassifiable Beat (Others)	33
	16	Fusion of Paced And Normal Beat (PFUS)	982
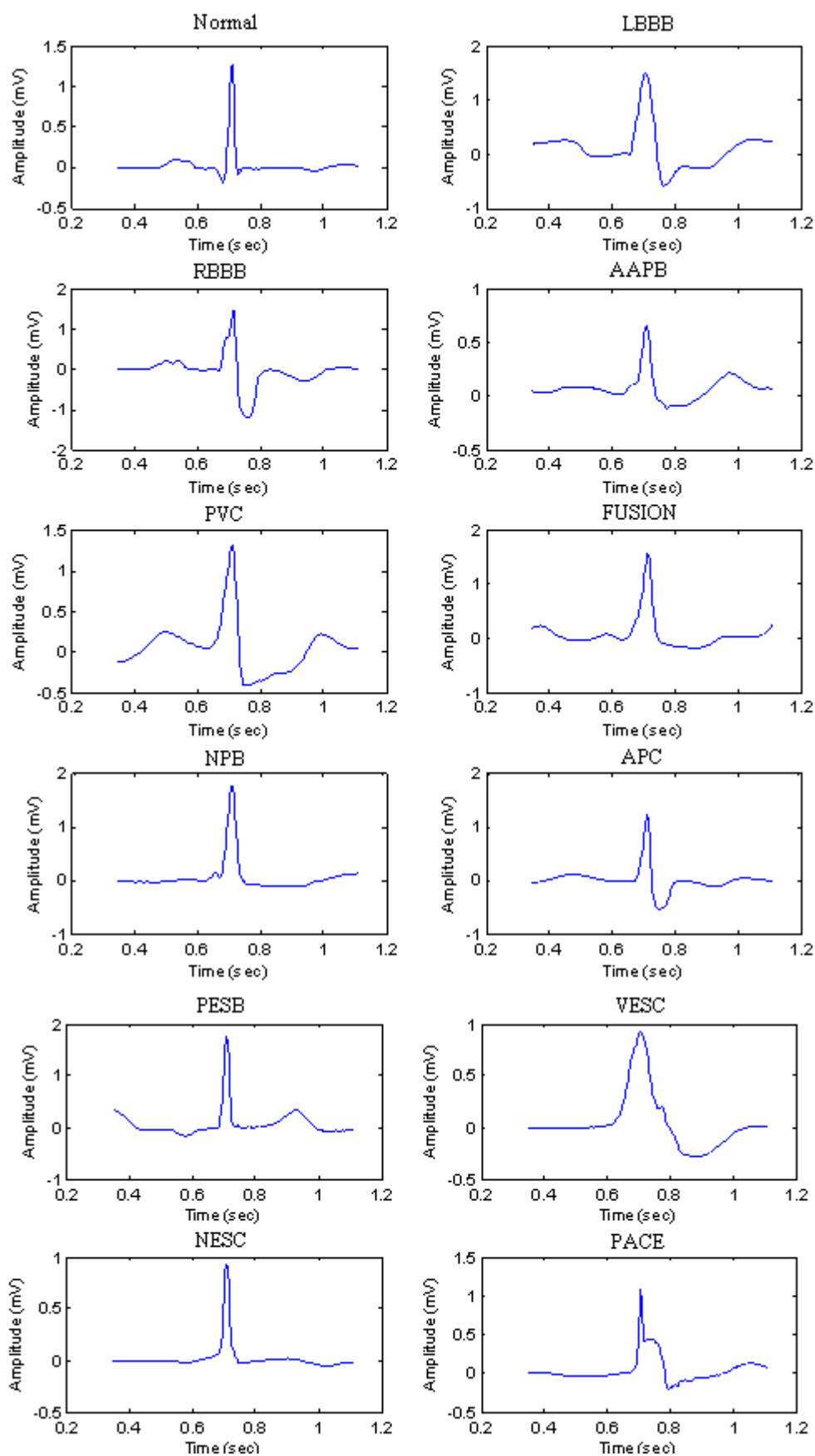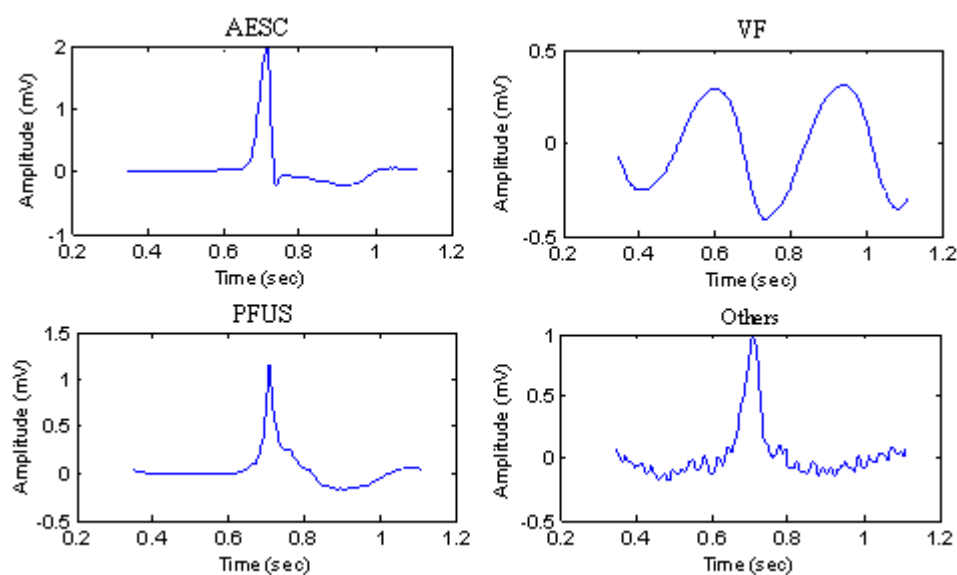
Figure 5.2 Samples of each heartbeat types.

5.3 Pre-Processing

5.3.1 Filtering

The baseline wandering and the power line interference are the most substantial noises and can strongly affect ECG signal analysis (Tompkins, 1993; Acharya, Suri, Spaan, & Krishnan, 2007). Baseline wander due to respiration contains low frequency components; power line interference contains high frequencies. All ECG signals are filtered with two median filters to remove the baseline wander. After the signal is processed with two median filters, which have 200ms and 600ms widths, respectively, the obtained signal is the baseline of the ECG signal. The result is subtracted from the original signal to remove the baseline wander effect of ECG signal.

Then power-line frequency is removed from the median filtered ECG by a notch filter. Records in database have 60 Hz power line interference. So with a proper notch filter power line frequency is removed. Result of filtering process is shown in Figure 5.3. The filtered ECG signals are used in all subsequent processing steps.
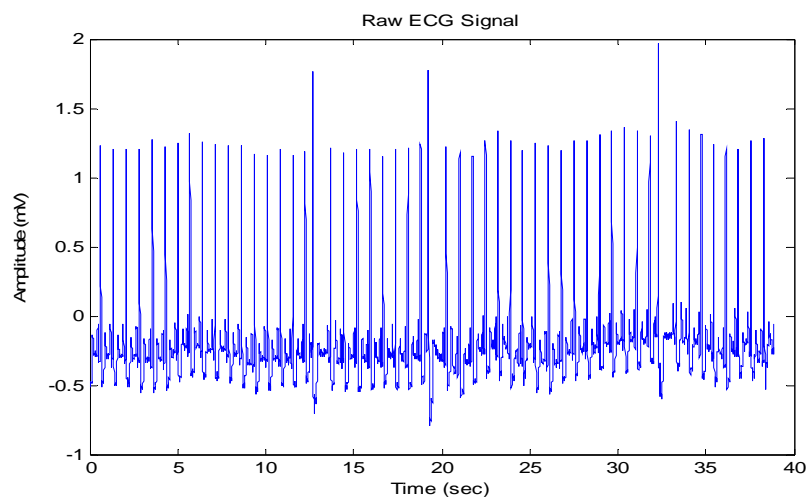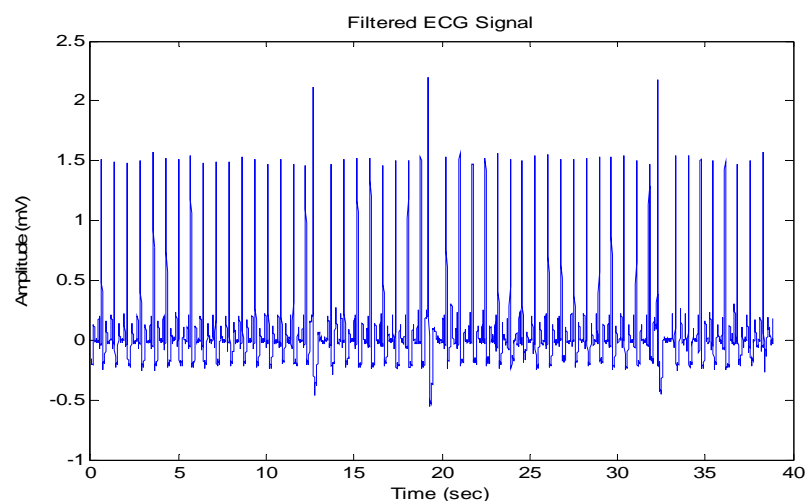
Figure 5.3 Result of filtering process a) Original signal with power line interference and base line wonder b) Filtered signal.

5.3.2 QRS Detection

In fact, many systems have already been constructed and executed for biomedical applications such as holter tape analysis, real-time patient monitoring etc… All these applications require an accurate detection of the QRS complex of the ECG. QRS detection is difficult, because of the physiological variability of the QRS complexes, and also because of the various types of noise that can be present in the ECG signal. In addition to the QRS complex, the ECG waveform contains P and T waves, power line interference, EMG from muscles, motion artifact from the electrode and skin

interface, and possibly other interference from electro surgery equipment in the operating room. It is important to extract the QRS complex from the other noise sources such as the P and T waves.

A QRS detection algorithm developed by Tompkins (Pan & Tompkins, 1985) is realized in this thesis. The energy of QRS complex is known to be concentrated approximately in the 5-15 Hz range (Hamilton & Tompkins, 1986; Pan & Tompkins, 1985). This QRS detection algorithm is based on analyses of the R point. The slope of the R wave is a common feature used to locate the QRS complex in many QRS detectors. However, a derivative amplifies the undesirable higher frequency noise components. Also, many abnormal QRS complexes with large amplitudes and long durations may be missed in a purely derivative approach because of their relatively low R wave slopes. Therefore, R wave slope lonely is insufficient for correct QRS detection. Hence, the algorithm consists of low-pass and high-pass filter, differentiator, squarer and moving average filter as seen in Figure 5.4.
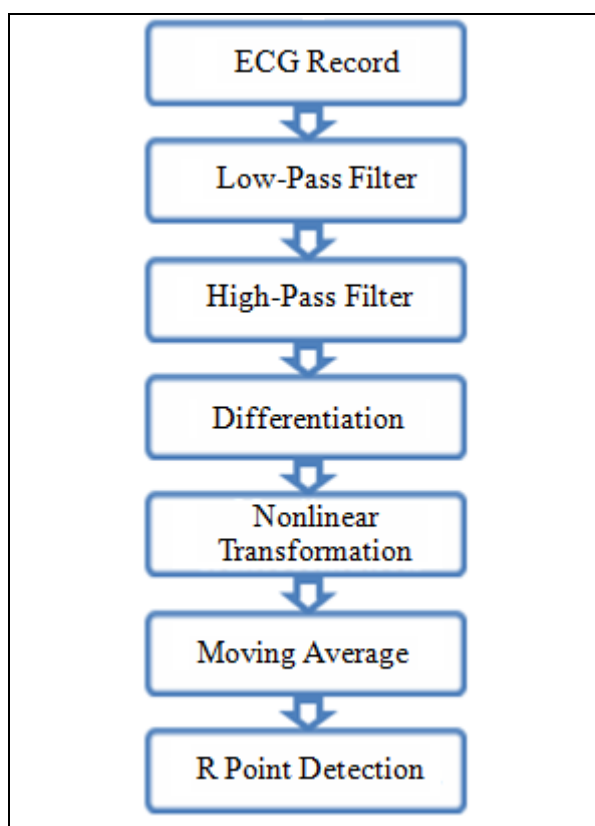
Figure 5.4 Block diagram of R point detection Algorithm.

In order to attenuate noise, the signal is passed through a band pass filter composed of cascaded high-pass and low-pass integer filters. Low-pass filter cut off frequency is about 11 Hz and the delay is five sample (or 25ms for a sampling rate of 200sps) and the gain is 36. The transfer function of second order low-pass filter is

$$H(z) = \frac{\left(1 - z^{-6}\right)^2}{\left(1 - z^{-1}\right)^2}$$ (5.1)

The difference equation of this filter is

$$y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) - 2x(nT - 6T) + x(nT - 12T)$$ (5.2)

Instead of the high-pass filter Tompkins used low-pass filter, which has cut off frequency at 5 Hz, gain of 32, and subtracted from the original signal. The delay is about 80ms. Transfer function of high-pass filter is

$$H(z) = z^{-16} - \frac{\frac{1 - z^{-32}}{1 - z^{-1}}}{32}$$ (5.3)

The difference equation of this filter is

$$y(nT) = x(nT - 16T) - \frac{y(nT - T) + x(nT) - x(nT - 32T)}{32}$$ (5.4)

The center of the pass-band frequency is at 10 Hz. At this process P, T waves are suppressed and the frequency characteristic of a QRS complex is optimally passed. Base line drift and power line interference are also eliminated from the signal.

The differentiation process is used to make clear the R point in the ECG signal. This process amplifies the higher frequency characteristic of QRS complex and

further attenuates the lower frequency of P and T waves. Five point derivatives has the transfer function

$$H(z) = 0.1(2 + z^{-1} - z^{-3} - 2z^{-4})$$

(5.5)

This derivative is implemented with the following difference equation

$$y(nT) = \frac{2x(nT) + x(nT - T) - x(nT - 3T) - 2x(nT - 4T)}{8}$$

(5.6)

The following process is the squaring process which is a nonlinear transformation. The equation is

$$y(nT) = [x(nT)]^2$$

(5.7)

This process increases the intensity of the output of the differentiation process. The last process is the time averaging process called moving window integral. The equation of moving window is

$$y(nT) = \frac{1}{N}\sum_{k=1}^{N-1} x((n-k)T)$$

(5.8)

This integrator sums the area under the squared waveform over a 150-ms interval while moving in time domain step by step.

Figure 5.5 shows the outputs of each processing steps. It starts from the original ECG and finishes with founded QRS point in the ECG signal. Figure 5.6 shows frequency response of the filters.
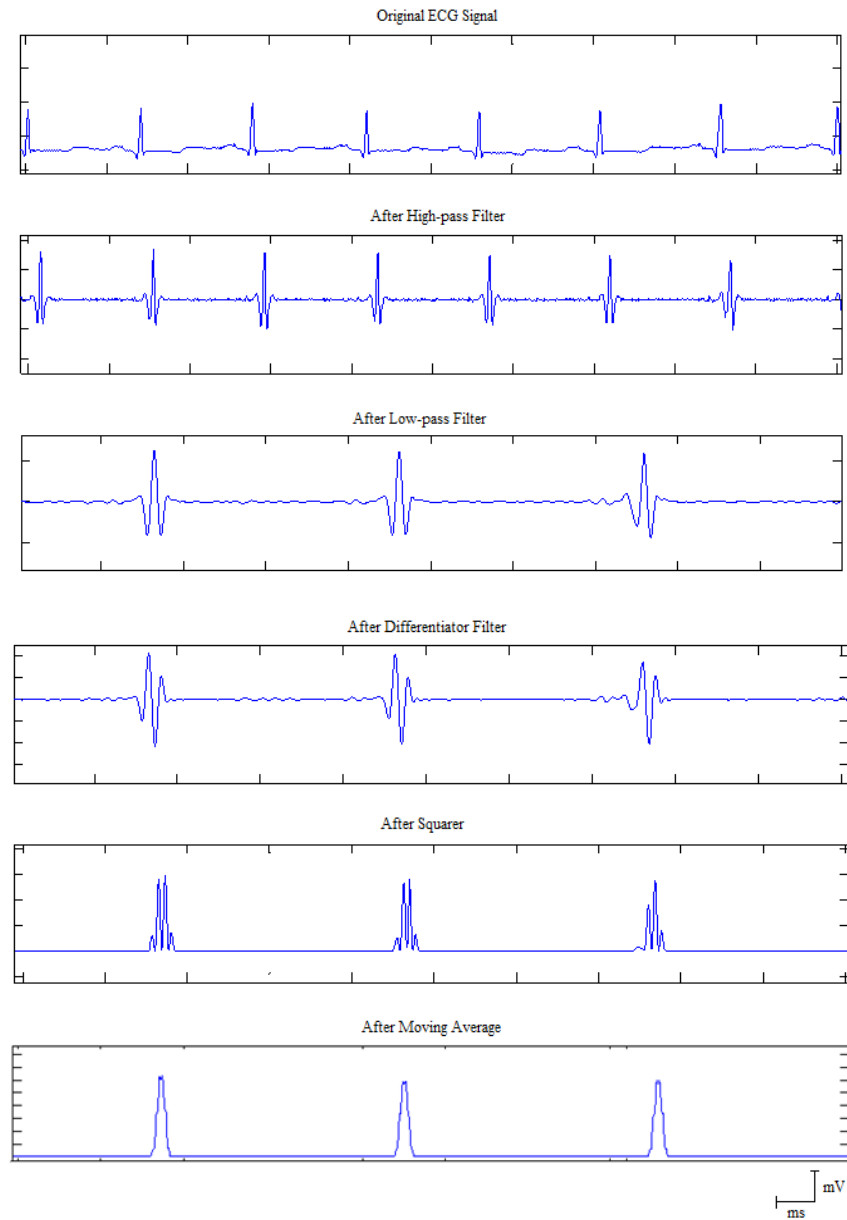
Figure 5.5 Output of QRS detection algorithm processing steps: Original ECG Signal, Output of the Notch Filter, Output of the High-pass Filter, Output of the Low-pass, Filter, Output of the Differentiator Filter, Output of the Squarer, and Output of the Moving Average Filter, respectively.
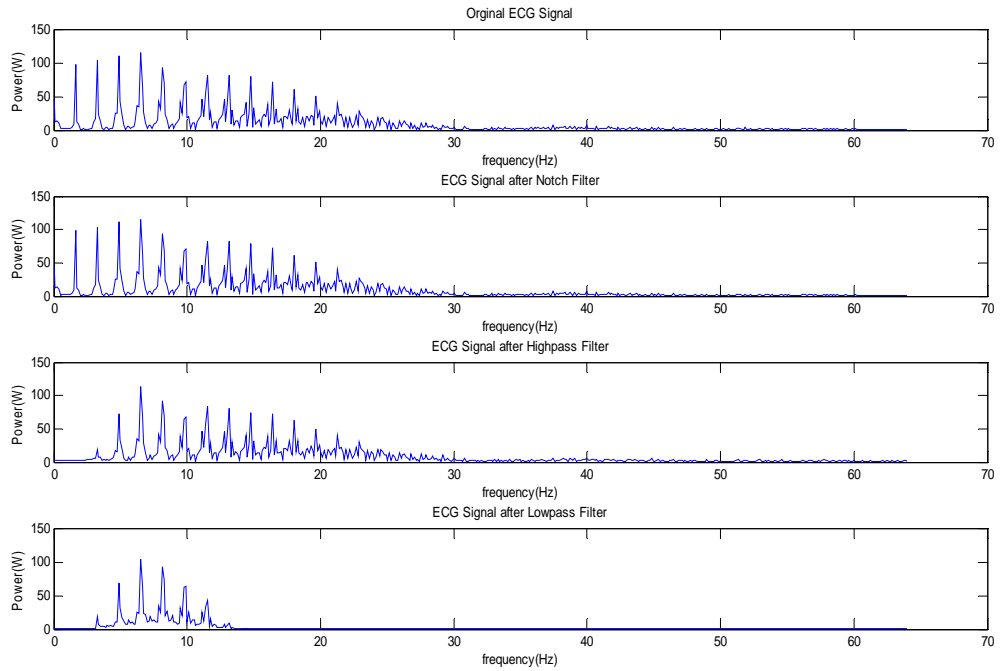
Figure 5.6 Frequency response of QRS detection algorithm processing steps: Original ECG Signal, Output of the Notch filter, Output of the High-pass filter, Output of the Low-pass filter, respectively.

Peak detector is the last process of the QRS detection algorithm. The locations of R peaks are determined by the algorithm. The output of the moving window process includes a large amplitude peak and some small amplitude peaks. Therefore, peak detector can sense these small peaks as R peak and can generate a wrong QRS detection. This can be avoided by using local maxima peak detector.

Two sets of thresholds are used to detect R points. Each set has two threshold levels. The set of thresholds that is applied to the waveform from the moving window integrator is

$$SPKI = 0.125(PEAKI) + 0.875(SPKI) \quad \text{if PEAKI is the signal peak} \quad (5.9)$$

$$NPKI = 0.125(PEAKI) + 0.875(NPKI) \quad \text{if PEAKI is the noise peak} \quad (5.10)$$

$$THRESHOLDI1 = NPKI + 0.25(SPKI-NPKI) \quad (5.11)$$

$$THRESHOLDI2 = 0.5 \; THRESHOLDI1 \quad (5.12)$$

where PEAKI is the overall peak, SPKI is the running estimate of the signal peak, NPKI is the running estimate of the noise peak, THRESHOLDI1 is the first threshold applied, and THRESHOLDI2 is the second threshold applied. Figure 5.7 shows the outputs of R point detection steps.
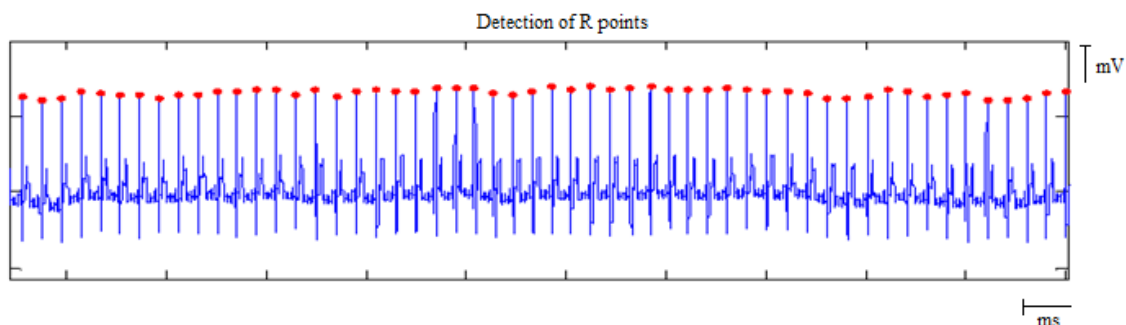
Figure 5.7 Outputs of R point detection steps.

5.3.3 Segmentation

The segmentation process is applied to the filtered signal after the R point is detected. The window with a length of 341 data points (the R peak of the wave is located at the 171th point) is extracted from the ECG record for each beat type. This corresponds to approximately 944 ms of the ECG signal. The window with a length of 944 ms is shown in Figure 5.8 for a normal ECG signal. The windowed beats are prepared as train and test sets as shown in Table 5.2.
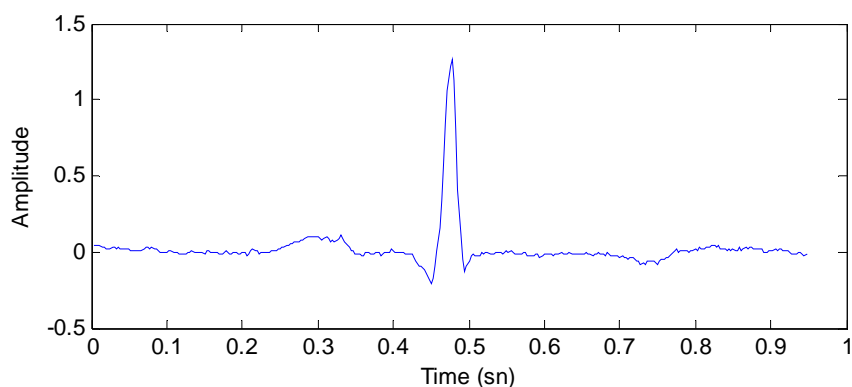
Figure 5.8 A window with a length of 341 data points for normal ECG signal.

Table 5.2. Mapping of MIT-BIH arrhythmia database heartbeat types to the AAMI heartbeat classes and amount of arrhythmias for training and testing

AAMI heartbeat classes (labels)	MIT-BIH heartbeat classes label	Train Set	Test Set
Non-Ectopic Beats (**N**)	1	350	250
	2	350	250
	3	350	250
	4	114	113
	5	8	8
Supraventricular Ectopic Beats (**S**)	6	74	74
	7	1	1
	8	350	250
	9	42	41
Ventricular Ectopic Beats (**V**)	10	236	236
	11	53	53
	12	350	250
Fusion Beats (**F**)	13	350	250
Unknown Beats (**Q**)	14	350	250
	15	17	16
	16	350	250
	Total :	3345	2542

5.4 Feature Extraction

Windowed data set from MIT-BIH arrhythmia database is further processed to extract other features.

5.4.1 Raw ECG

The raw ECG data which had different window sizes are used to investigate the effect of window size (Kutlu, Kuntalp, & Kuntalp, 2007). In order to take a long enough time segment for all probable cases, the feature sets are created with different window sizes. Each feature set are constituted by adding new part of 14ms (5 consecutive points) to this interval as shown in Figure 5.9. Total length of the window size is at most 1400ms (700ms right side and 700ms left side of R point). Figure 5.9 shows seven probable cases which are only a small part of right side of the R point.
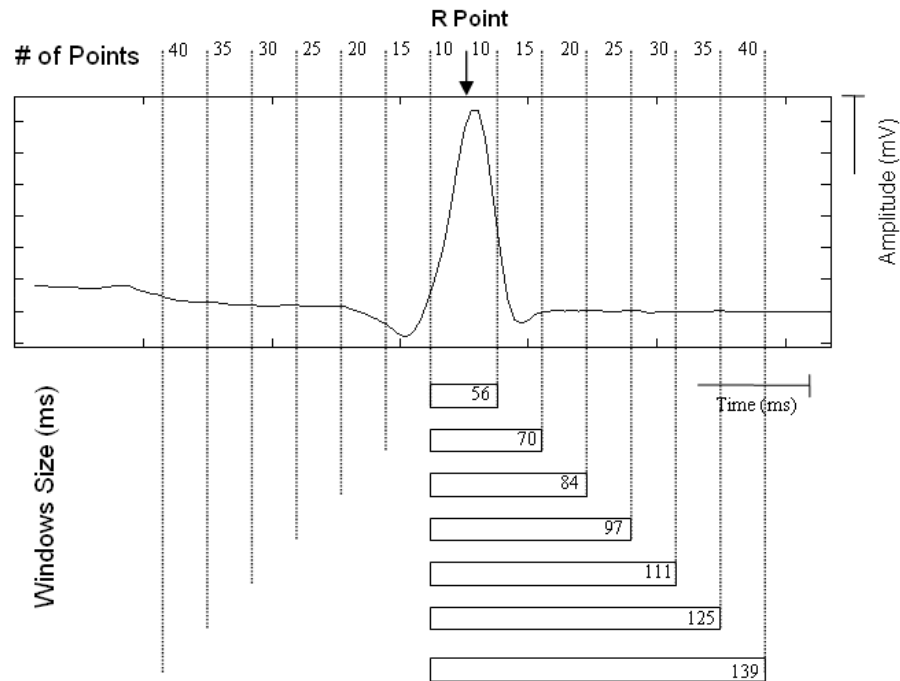
Figure 5.9 Amount of consecutive points and window size for each feature set.

K-nearest neighborhood method is used as classifier in search process. The performance is measured based on three standard statistical measures: sensitivity, specificity, and selectivity According to results of network structures. 161 raw ECG points (130 previous points and 30 next points) is the best performance for KNN based system.

5.4.2. *Higher Order Statistics*

The second, third and fourth order cumulants are calculated for each beat taking lag 0 (Kutlu, Kuntalp, & Kuntalp, 2008b). The zero-lag cumulants have special names: $c_2(0)$ is the variance and is denoted by σ^2; $c_3(0,0)$ and $c_4(0,0,0)$ are denoted by γ_{3x} and γ_{4x} known as the skewness and the kurtosis, respectively.

5.4.3. *Wavelet Packet Decomposition*

Different mother wavelets, such as Daubechies 2 (db2), Daubechies 4 (db4), and Daubechies 6 (db6) (as shown Figure 5.10), are examined as mother wavelet

function for estimating the wavelet packet coefficients in the study. Classification accuracy of extracted features from each mother wavelet is compared. ECG signals are decomposed up to level 4**.** Therefore the number of sub bands is 30 for the fourth level of wavelet packet decomposition.
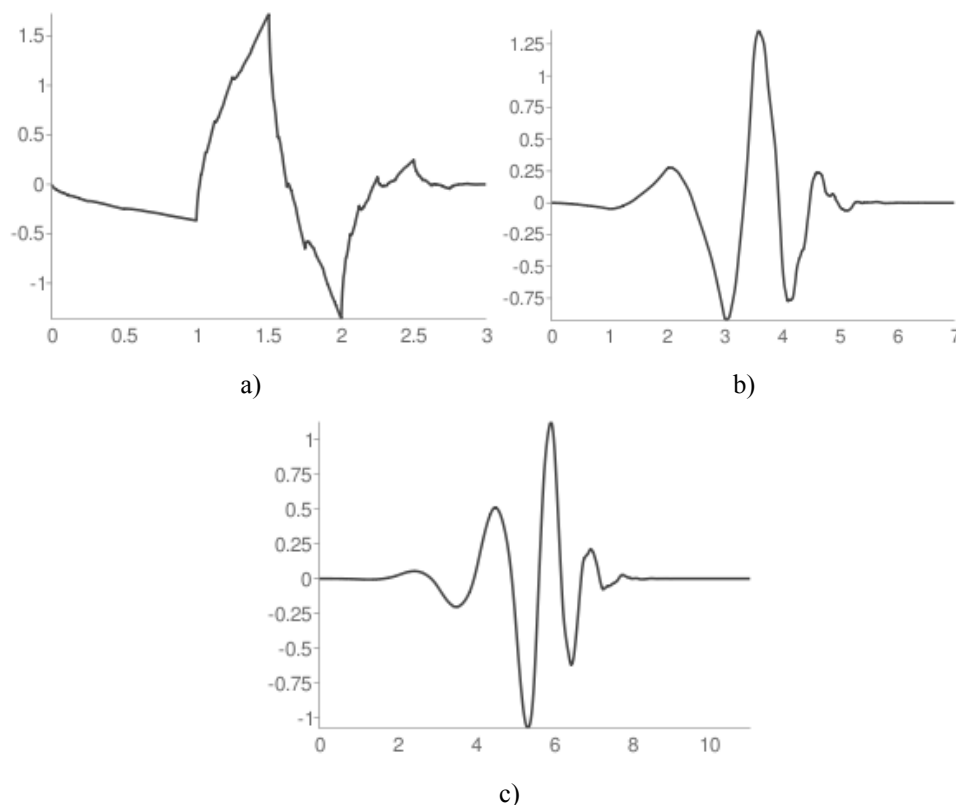
Figure 5.10 Mother wavelet function; a) daubechies 2, b) daubechies 4, c) daubechies 6.

The wavelet packet decomposition is a feature extraction tool. Not only does the WPD utilizes the low frequency components (approximations) but also the high frequency components (details). But the size of the feature vector might be too high to be applied as input to a classification system. The HOS methods are used to extract new and fewer number of features from the wavelet packet decomposition coefficients. Using the higher order statistical methods, the second, third and fourth order cumulants of each level of sub bands are calculated (Kutlu & Kuntalp, 2009b). There are thirty sub bands for the four levels. Three features are extracted for each sub band using HOS (2nd , 3rd , 4th order) and a total of 90 HOS features (Six features from first level, 12 features from second level, 24 features from third level, and 48 features from fourth level) are obtained for all sub bands of WPD. By using HOS of

sub band signals, it becomes possible to define hidden features embedded in the QRS complex.

HOS features are extracted for each mother wavelet. The classification performances are measured using KNN classifier. The results are shown in Table 5.3, 5.4, and 5.5. In the literature, many studies in the analysis of ECG preferred db4 wavelet as mother wavelet, because of the similarity of the QRS complex (Bakardijin &Yamamoto, 1995;Pic hot et al., 1999; Übeyli, 2008). In this application the best result for performance measures of the normal beat (for label 1) is obtained with db4 wavelet as well. But performance measures of other beats are not as good as the normal beat. The performance measures of db2 and db6 wavelet are almost the same as average value of performance measures. But overall accuracies are 72.8%, 73.3%, and 74.9% for db2, db4, db6, respectively. Although it is not clearly seen any difference, the best one is obtained with db6 wavelet. Consequently, these features are used in the constructed system.

Table 5.3 Performance measure of features for db2 wavelet

Label	SEN	SEL	SPE	Area Under Curve
1	60.00	64.02	96.92	0.78
2	83.00	77.18	97.81	0.90
3	82.50	76.91	97.86	0.90
4	53.00	57.00	98.69	0.76
5	62.00	67.82	97.46	0.80
6	75.00	69.65	96.97	0.86
7	41.24	47.00	99.12	0.70
8	72.00	63.38	96.07	0.84
9	0.00	0.00	99.95	0.50
10	88.45	86.09	99.72	0.94
11	63.96	70.04	98.71	0.81
12	85.00	81.85	98.56	0.92
13	6.25	11.29	99.73	0.53
14	75.33	83.59	98.53	0.87
15	9.50	13.67	99.77	0.55
16	70.00	73.04	97.71	0.84
Average	57.95	58.91	98.35	0.78

Table 5.4 Performance measure of features for db4 wavelet

Label	SEN	SEL	SPE	Area Under Curve
1	94.40	97.12	99.69	0.97
2	91.20	88.03	98.65	0.95
3	79.20	96.59	99.69	0.89
4	16.22	36.36	99.15	0.58
5	64.40	42.04	90.31	0.77
6	32.80	51.90	96.68	0.65
7	39.02	32.65	98.68	0.69
8	76.40	63.04	95.11	0.86
9	0.00	0.00	100.00	0.50
10	37.74	52.63	99.28	0.69
11	84.07	82.61	99.18	0.92
12	94.00	95.14	99.48	0.97
13	12.50	22.22	99.72	0.56
14	75.00	81.57	98.27	0.87
15	0.00	0.00	99.76	0.50
16	84.00	75.81	97.08	0.91
Average	**55.06**	**57.36**	**98.17**	**0.77**

Table 5.5 Performance measure of features for db6 wavelet

Label	SEN	SEL	SPE	Area Under Curve
1	86.40	94.32	99.43	0.93
2	90.80	88.67	98.73	0.95
3	84.00	97.67	99.78	0.92
4	32.32	45.86	98.99	0.66
5	63.60	46.39	92.23	0.78
6	38.80	64.12	97.82	0.68
7	37.15	37.84	99.04	0.68
8	79.60	53.29	92.28	0.86
9	0.00	0.00	100.00	0.50
10	38.85	76.00	99.76	0.69
11	88.50	87.21	99.34	0.94
12	94.80	96.73	99.65	0.97
13	0.00	0.00	98.81	0.49
14	74.61	83.64	98.35	0.86
15	0.00	0.00	99.76	0.50
16	87.20	73.72	96.34	0.92
Average	**56.04**	**59.09**	**98.14**	**0.78**

5.4.4 Morphological Features

This approach is based on R point in the ECG record. Therefore, informations of QRS detection are used in this step. The Q and S points are limited within the 150 ms period which is centered by the R point (Tompkins, 1993). The sampling frequency of the data set is 360 Hz. and 150ms equal to nearly 55 consecutive points. This is the region of interest (ROI). Hence the fiducial points are limited with nearly 28 point left and 28 point right. By using these statistical data with the first derivative of ECG, the Q and S points are determined.

The time-derivative of the signal $f(t)$ are calculated and its zero-crossings are found. For a discrete signal $f[n]=f(t)|_{t=nT}$ it can be obtained by searching the inflection points based on the following criteria:

- If $f[n]- f[n-1] \leq 0 \leq f[n+1]- f[n]$, then this is a negative peak; it is stored with index.

- If $[f[n]- f[n-1]].[f[n+1]- f[n]]>0$, then this is not a peak; it is discarded.

This procedure is indeed a numerical differentiation technique such that the equation $(f(n+1)-f(n))/T$ represents the first-order forward approximation to the derivative of $f(t)$ at $t=nT$.

Using R, Q and S points, amplitudes of QR (G1), amplitudes of RS (G2), QR width (T1), RS width (T2), and slope of right and left side of R point are extracted as features as shown in Figure 5.11.
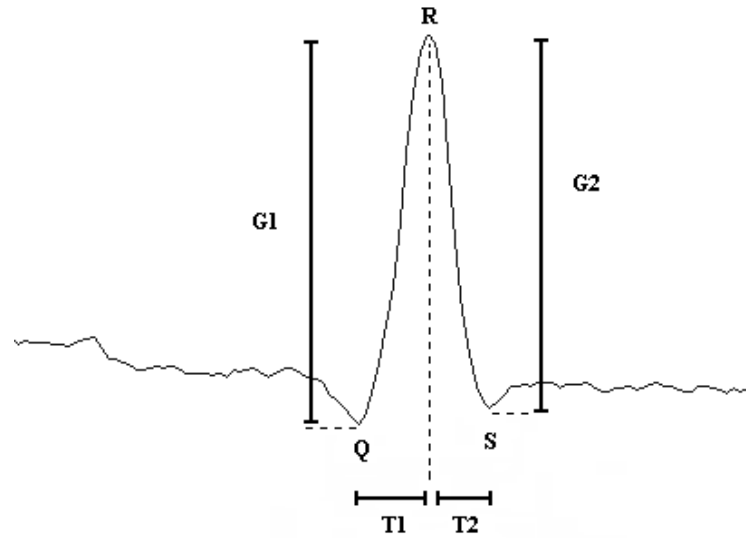
Figure 5.11 Morphological features of a sample normal signal.

5.4.5. Discrete Fourier Transform

Fourier transform is often called the frequency domain representation of the original signal. Frequency of ECG signals is in the range of 0-100Hz. But in clinical process band of 0-50Hz is analyzed. So in this study frequency band of 0-50Hz is considered. A sample energy spectrum of a normal ECG signal is shown in Figure 5.12. Using 256 sample Fourier transform, 46 energy values of the coefficients which are less than 50 Hz are used as features.
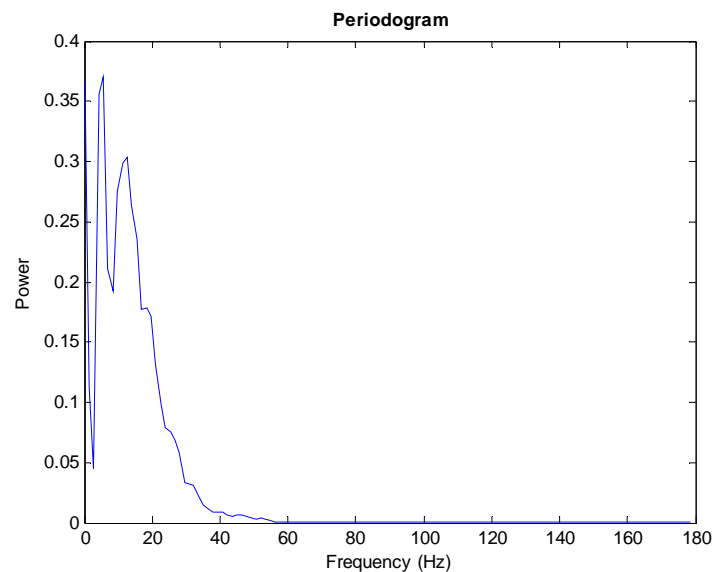
Figure 5.12 A sample energy spectrum of normal signal.

5.5 Visualization of Feature Sets Using Self Organizing Maps

SOM is used for different approach such as visualization data, clustering, feature extracting (Kutlu, Kuntalp, & Kuntalp, 2008a, Kutlu, Kuntalp, & Kuntalp, 2008c, Kutlu & Kuntalp, 2009a, Kutlu, Kuntalp, & Kuntalp 2009a, Kutlu, Kuntalp, & Kuntalp 2009c). In this step, the arrhythmias in the electrocardiograph (ECG) signals are analyzed by using Self Organizing Maps (SOM). The feature sets obtained with different methods are used for training the SOM networks diversely. The maps are examined using U-matrix representation method. That way, high dimensional data are examined in two dimensions. The clusters that appear in U-matrix representation is examined for different feature sets. U-matrix representation of a SOM network is given using three visualization: colored image, gray level image with clusters, and three-dimensional the hill-valley landscape visualization. They are described in section 3.5. The cluster color code is shown in Table below.

SOM structures are constructed with 3500 (70x50) neurons and hexagonal topology. Then U-matrix is calculated from trained SOM structures. The computed U-matrix is visualized through a colored image.

Table 5.6 Cluster color code in U-matrix representation

	Class
	Normal Beat (NORMAL)
	Left Bundle Branch Block Beat (LBBB)
	Right Bundle Branch Block Beat (RBBB)
	Aberrated Atrial Premature Beat (AAPB)
	Premature Ventricular Contraction (PVC)
	Fusion of Ventricular And Normal Beat (FUSION)
	Nodal (Junctional) Premature Beat (NPB)
	Atrial Premature Contraction (APC)
	Premature or Ectopic Supraventricular Beat (PESB)
	Ventricular Escape Beat (VESC)
	Nodal (Junctional) Escape Beat (NESC)
	Paced Beat (PACE)
	Unclassifiable Beat (Others)
	Ventricular Flutter Wave (VF)
	Atrial Escape Beat (AESC)
	Fusion of Paced And Normal Beat (PFUS)

Figures 5.13 - 5.17 show the U-matrix analysis of HOS features, Fourier Transform, Morphological Features, HOS features of only third level of WTP analysis, respectively.
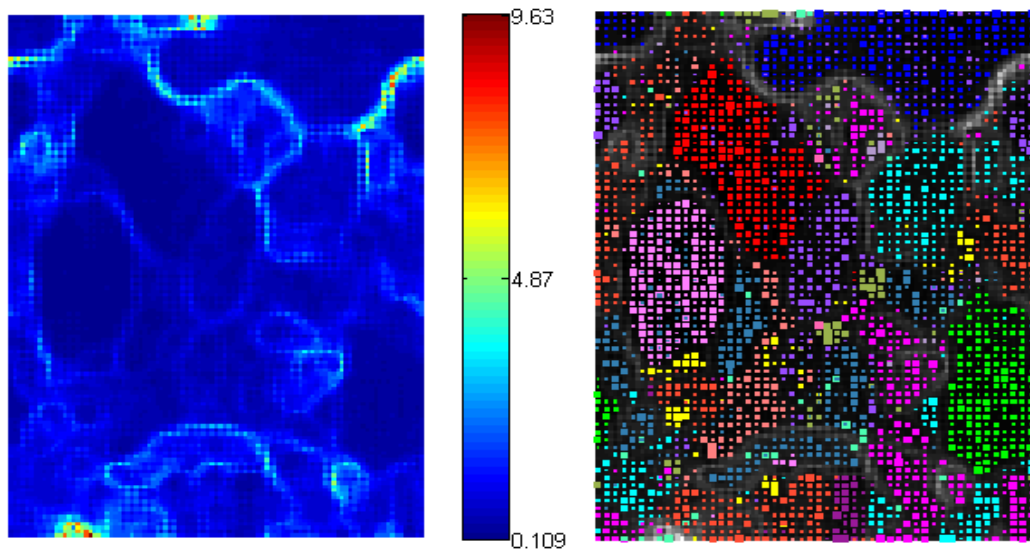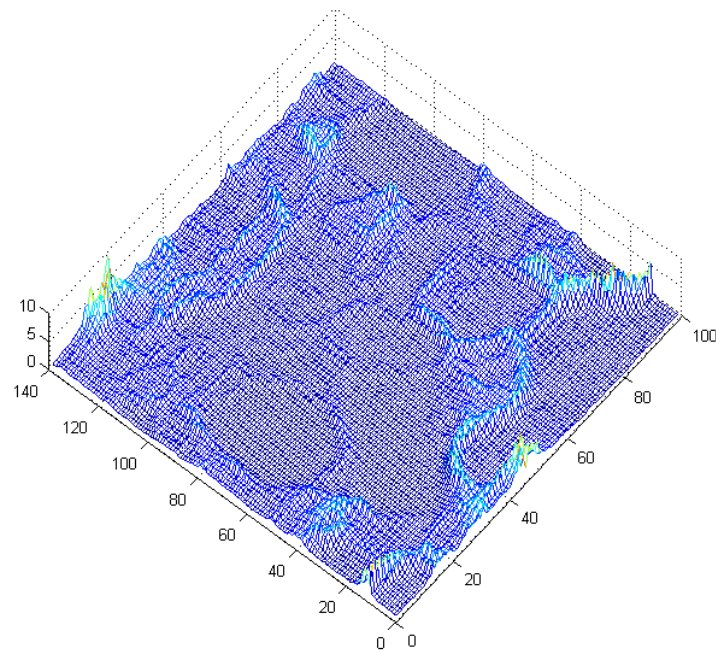
Figure 5.13 U-matrix representation of a SOM network for raw ECG data, a) colored image visualization, b) gray level visualization with colored clusters, c) 3D hill-valley landscape visualization.

a)

b)

c)
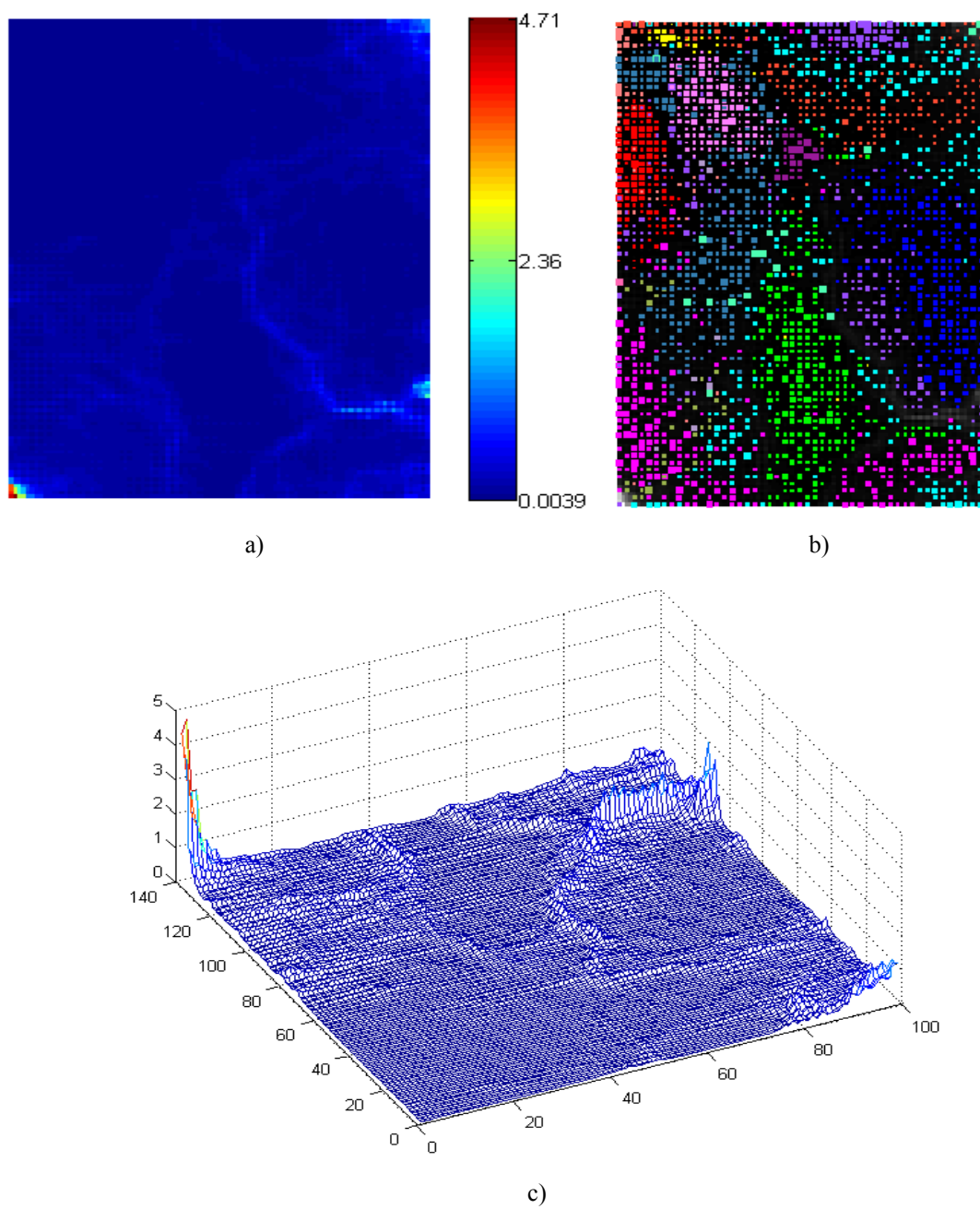
Figure 5.14 U-matrix representation of a SOM network for HOS features, a) colored image visualization, b) gray level visualization with colored clusters, c) 3D hill-valley landscape visualization.

a)

b)

c)
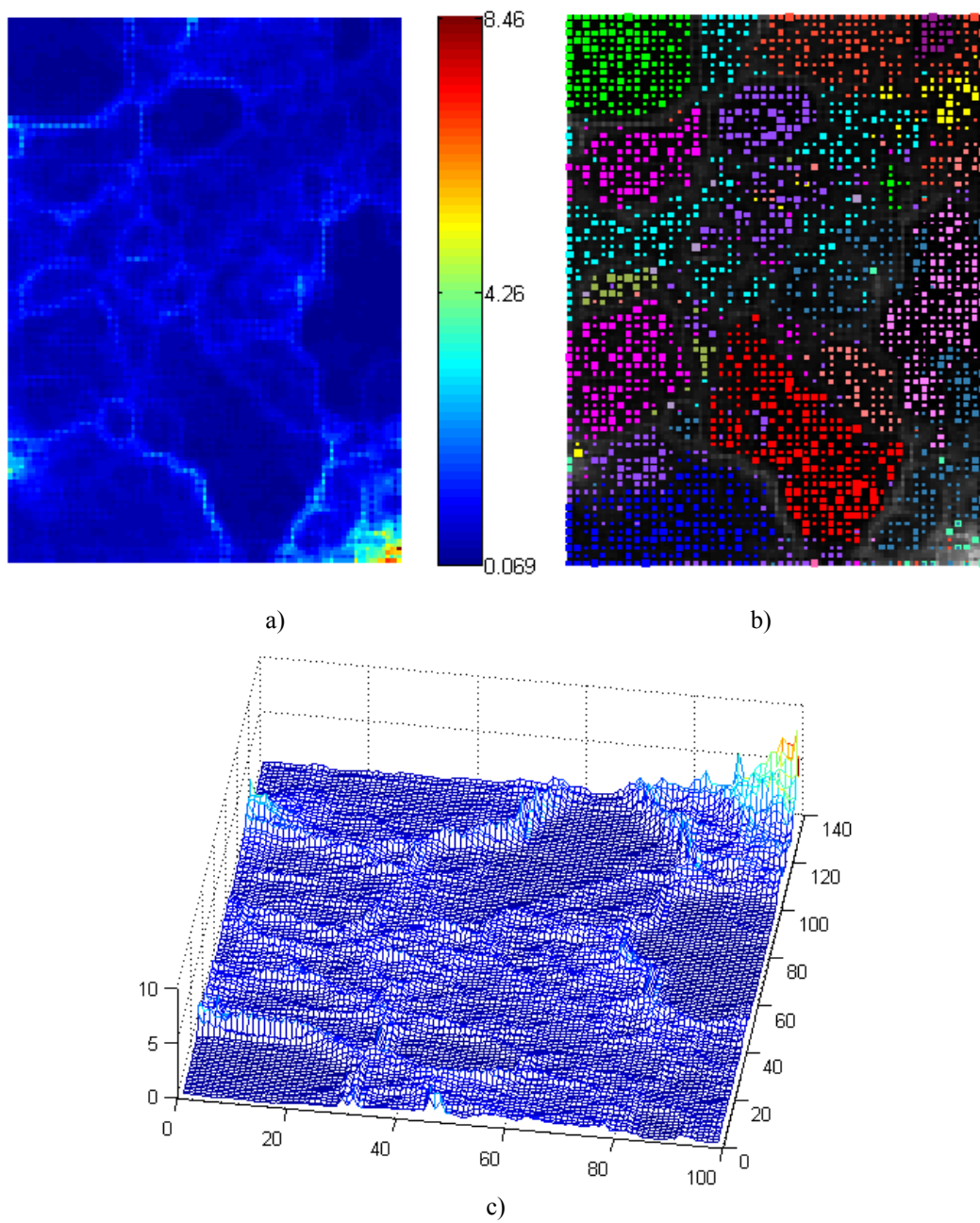
Figure 5.15 U-matrix representation of a SOM network for Fourier Transform coefficient, a) colored image visualization, b) gray level visualization with colored clusters, c) 3D hill-valley landscape visualization.

a)

b)

c)
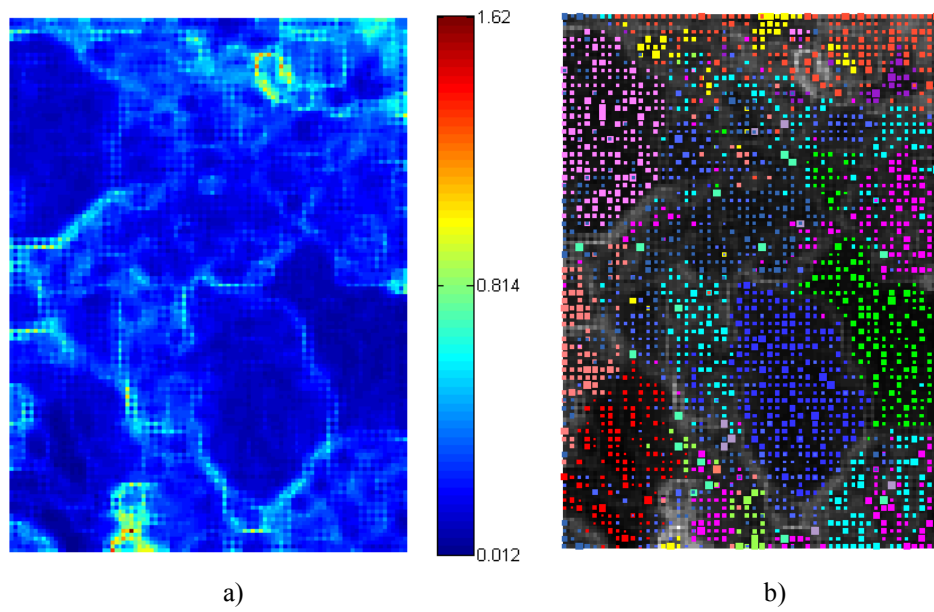
Figure 5.16 U-matrix representation of a SOM network for morphological Features, a) colored image visualization, b) gray level visualization with colored clusters, c) 3D hill-valley landscape visualization.

a)   b)

c)
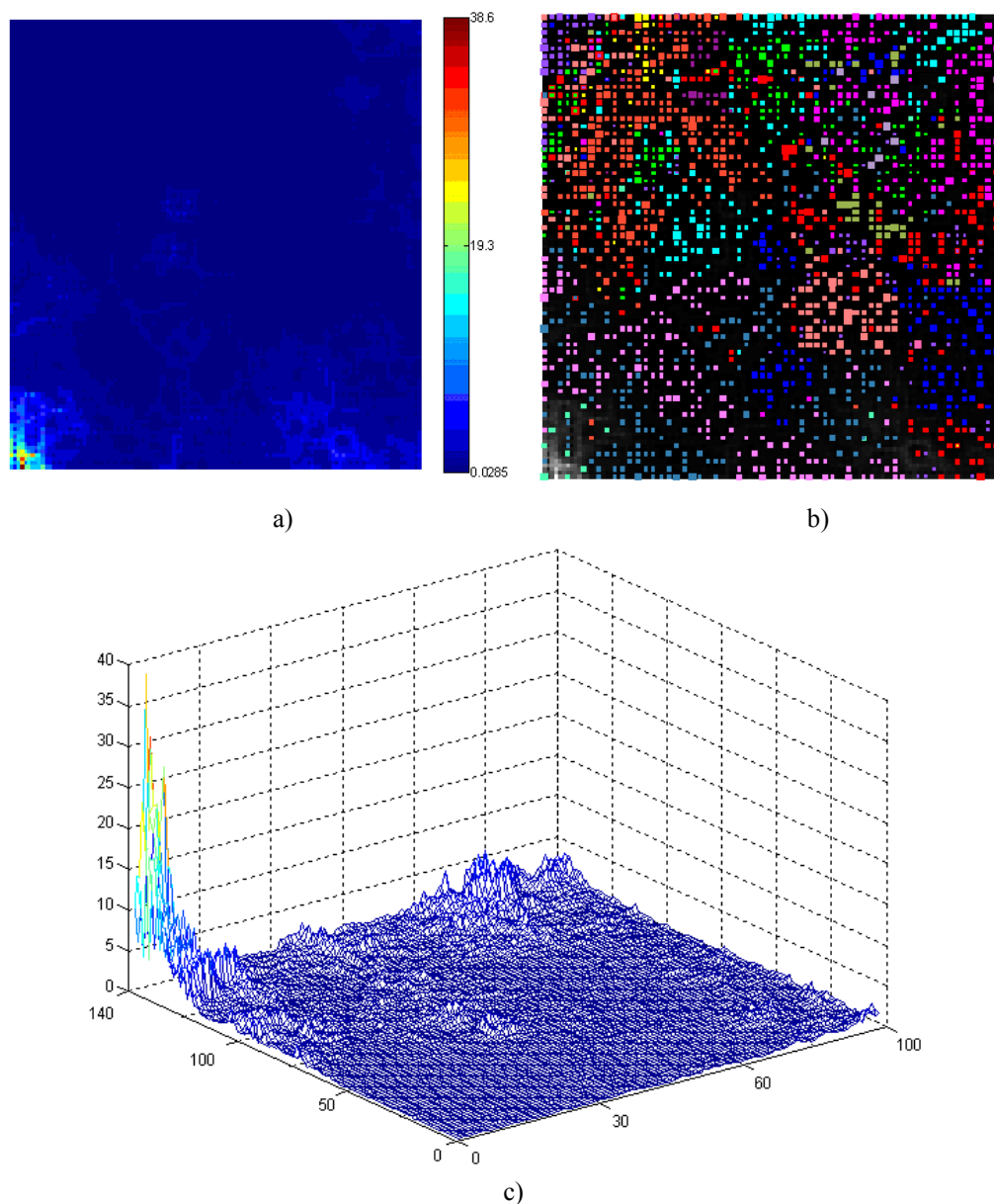
Figure 5.17 U-matrix representation of a SOM network for HOS features of only third level of WTP analysis, a) colored image visualization, b) gray level visualization with colored clusters, c) 3D hill-valley landscape visualization.

In general looking at these results two patterns are identified in colored image visualizations. First, there are big homogeneous areas which are colored blue. Here it can be said that the datasets have a high degree of similarity in each dark area. For example, colored image visualization in Figure 5.16.a the big blue region (top left) represents cluster of an arrhythmia type (seen in Figure 5.16.b). On the other hand, lighter shade indicates location of data vectors having high distance. It also indicates the border of the clusters where the dark areas indicate the clusters. For example,

colored image visualization in Figure 5.16.a the big dark blue region bordered lighter color (bottom right) represents cluster of an arrhythmia type (seen in Figure 5.16.b) and the lighter color indicates border of the cluster. 3-dimensional visualization of the SOM has valleys which indicate similarities in the input data and hills which indicate distance in the input data. That means hills indicate the border of the clusters. For example, 3-dimension visualizations of the same regions (bottom right and top left) in Figure 5.16.a are also given in Figure 5.16.c.

According to results of different feature sets, no unique robust feature set is able to classify all heartbeat types. For example, no cluster is clearly separated in Figure 5.14. LBBB, RBBB and paced beat types are clearly separated in Figure 5.16. Therefore the ECG heartbeat classification becomes a typical problem of classification which requires the exploration of diverse set of features.

5.6 Dimensionality Reduction

After the feature extraction processes, total of 150 features were extracted which are 8 form HOS, 90 from HOSofWPT, 6 from morphology, and 46 from Fourier transform. Features are labeled as numbers for the selection algorithm. HOS features label are numbered as f1-f8. Label of morphological features are numbered as f9-f14. Label of HOS features of WPD are numbered as f15-f104. And Fourier coefficient labels are numbered as f105-f150. To decrease dimension of feature set, wrapper method is used.

5.6.1 Feature Dimension Reduction using Selection Algorithm

Automatic feature selection is an optimization technique that, given a set of m features, attempts to select a subset of size n that leads to the maximization of some criterion function. Feature selection algorithms are important in recognition and classification systems because if a feature space with a large dimension is used, the performance of the classifier will decrease with respect to execution time and to recognition rate. Two feature selection algorithms are used in this process: sequential

floating search methods (SFSM) and Genetic Algorithm (GA).

In sequential search algorithm, the feature selection criterion is the overall accuracy of the KNN classifiers. Two step SFS ($n=2$) and one step SBS ($r=1$) are utilized.

In the GA, population size is taken 300; one point Crossover is utilized; elitizm strategy is used and the number of best individuals is two that are guaranteed to survive to the next generation; stochastic uniform selection is utilized; mutation is performed by randomly selecting a bit in a string and changing its value.

Feature selection process block diagram is shown in Figure 5.18. In the selection algorithm *K*-nearest neighborhood algorithm is used as classifier. Overall accuracy of classifier and ROC area measures are used as performance criteria of selection algorithm.
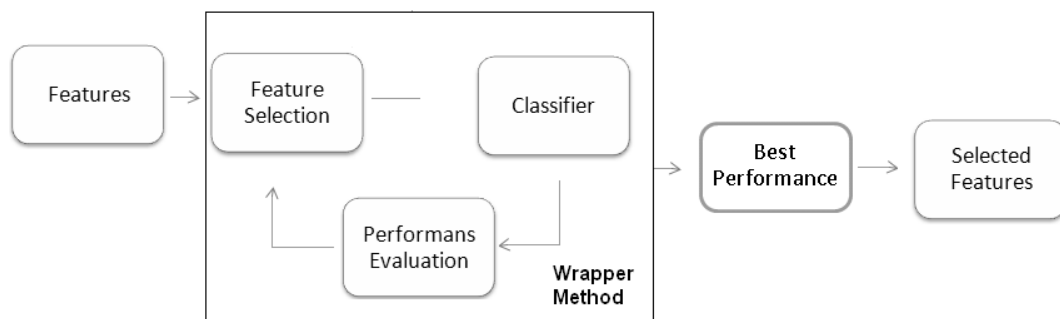
Figure 5.18 Block diagram of feature selection.

The constructed system has two classification stages. In the first stage, it classifies all heartbeats into five main groups. Therefore, feature selection algorithm is performed for each of these five groups. In addition, genetic algorithm and sequential floating search methods are both used in feature selection stage. At first, overall accuracy of classifier is used as performance criteria of selection algorithm. The results of the selection process are shown in Table 5.7 for GA and in Table 5.8 for SFSM. The feature sizes of the each main group (N, S, V, F, and Q) are 51, 64,

64, 61, and 69, respectively, with GA, and 61, 62, 67, 74, and 100, respectively, with SFSM.

Table 5.7. Amount of selected features for main groups with GA using overall accuracy as performance measure

Label	Selected Features
N	2 HOSF, 1 MF, 24 HOSofWPT, 24 FTF
S	2 HOSF, 1 MF, 45 HOSofWPT, 16 FTF
V	2 HOSF, 1 MF, 47 HOSofWPT, 14 FTF
F	2 HOSF, 3 MF, 37 HOSofWPT, 19 FTF
Q	4 HOSF, 3 MF, 40 HOSofWPT, 22 FTF

Table 5.8. Amount of selected features for main groups with SFSM using overall accuracy as performance measure

Label	Selected Features
N	1 HOSF, 37 HOSofWPT, 23 FTF
S	2 HOSF, 31 HOSofWPT, 29 FTF
V	1 HOSF, 2 MF, 28 HOSofWPT, 36 FTF
F	3 HOSF, 1 MF, 36 HOSofWPT, 34 FTF
Q	4 HOSF, 56 HOSofWPT, 40 FTF

*HOSF: Higher Order Statistics Features,
*MF: Morphological Features,
*HOSofWPT:HOS Features of Wavelet Packet Transform,
*FTF: Fourier Transform Features

Performance measures of classifiers for five main groups are shown in Table 5.9 for the cases of using all features, using selected features with SFSM and using selected features with GA. Overall accuracies are %83.01, %87.49 and %93.04 for all features, selected features with SFSM and selected features with GA, respectively. Results show that the best performance is obtained with features of GA selection. Therefore, first stage of the system is constructed using features obtained with GA these results.

Table 5.9 Performance measures of KNN based classifier using all features and selected features with SFSM and GA using overall accuracy as performance measure

Label	All Features				Selected Features with SFSM				Selected Features with GA			
	SEN	SEL	SPE	ROC Area	SEN	SEL	SPE	ROC Area	SEN	SEL	SPE	ROC Area
N	87.72	93.40	96.77	92.24	94.37	99.40	99.70	97.04	95.18	99.04	99.52	97.35
S	86.89	62.48	91.22	89.05	83.06	94.41	99.17	91.12	87.70	94.13	99.08	93.39
V	85.16	75.62	92.61	88.88	86.64	96.29	99.10	92.87	88.87	96.57	99.15	94.01
F	27.20	70.83	98.78	62.99	48.80	99.19	99.96	74.38	92.00	95.44	99.52	95.76
Q	97.09	97.85	99.46	98.28	98.64	64.84	86.38	92.51	98.06	80.70	94.03	96.04
Average	76.81	80.04	95.77	86.29	82.30	90.82	96.86	89.58	92.36	93.18	98.26	95.31

In the second stage, the constructed system classifies each main group into heartbeats. The feature selection algorithm is also applied in the second stage to select optimal features for subgroups of each main group. For this purpose, the following features (see Table 5.10) are chosen as the best discriminative features. Since the main group F has no subgroup, the selected features for the subgroups of remaining four main groups are used at the second stage. Results of overall accuracy of subgroup classifiers are %100, %96.06, %99.62 and %98.8 for subgroups of N, subgroups of S, subgroups of V and subgroups of Q, respectively.

Table 5.10 Number of selected features for subgroups of each main group using overall accuracy as performance measure

Main group	Label	Selected Features of Subgroups
Main group N	1	1 MF, 1 HOSofWPT, 2 FTF
	2	1 HOSofWPT, 2 FTF
	3	2 FTF
	4	2 MF, 1 HOSofWPT, 41 FTF
	5	1 MF, 2 HOSofWPT, 1 FTF
Main group S	6	1 HOSF, 1 MF, 1 HOSofWPT, 5 FTF
	7	1 MF, 1 HOSofWPT, 4 FTF
	8	1 HOSF, 1 MF, 2 HOSofWPT, 5 FTF
	9	1 HOSF, 20 FTF
Main group V	10	9 HOSofWPT, 8 FTF
	11	3 HOSofWPT, 2 FTF
	12	2 MF, 7 HOSofWPT, 34 FTF
Main group F	13	2 HOSF, 3 MF, 37 HOSofWPT, 19 FTF
Main group Q	14	2 HOSF, 10 HOSofWPT, 38 FTF
	15	4 HOSofWPT, 39 FTF
	16	5 HOSofWPT, 14 FTF

*HOSF: Higher Order Statistics Features,
*MF: Morphological Features,
*HOSofWPT:HOS Features of Wavelet Packet Transform,
*FTF: Fourier Transform Features

When ROC area measure of classifier is used as performance criteria of selection algorithm: In the first stage, the results of the selection process are shown in Table 5.11 for GA and in Table 5.12 for SFSM. The feature sizes of the each main group (N, S, V, F, and Q) are 63, 68, 64, 68, and 76, respectively, with GA, and 70, 63, 71, 70, and 95, respectively, with SFSM.

Table 5.11. Amount of selected features for main groups with GA using

ROC area as performance measure

Label	Selected Features
N	0 HOSF, 1 MF, 37 HOSofWPT, 25 FTF
S	2 HOSF, 1 MF, 47 HOSofWPT, 18 FTF
V	2 HOSF, 1 MF, 43 HOSofWPT, 18 FTF
F	2 HOSF, 3 MF, 41 HOSofWPT, 22 FTF
Q	5 HOSF, 4 MF, 49 HOSofWPT, 18 FTF

Table 5.12. Amount of selected features for main groups with SFSM

using ROC area as performance measure

Label	Selected Features
N	1 HOSF, 1 MF, 44 HOSofWPT, 24 FTF
S	0 HOSF, 0 MF, 40 HOSofWPT, 23 FTF
V	1 HOSF, 2 MF, 30 HOSofWPT, 38 FTF
F	1 HOSF, 2 MF, 31 HOSofWPT, 36 FTF
Q	3 HOSF, 0 MF, 52 HOSofWPT, 40 FTF

*HOSF: Higher Order Statistics Features,
*MF: Morphological Features,
*HOSofWPT:HOS Features of Wavelet Packet Transform,
*FTF: Fourier Transform Features

Performance measures of classifiers for five main groups are shown in Table 5.13 for the cases of using selected features with SFSM and using selected features with GA. Overall accuracies are %88.00 and %92.62 for selected features with SFSM and selected features with GA, respectively.

Table 5.13 Performance measures of KNN based classifier using selected
features with SFSM and GA using ROC area as performance measure

Label	Selected Features with SFSM				Selected Features with GA			
	SEN	SEL	SPE	ROC Area	SEN	SEL	SPE	ROC Area
N	84.73	99.33	99.70	0.92	95.29	98.46	99.22	0.97
S	80.05	96.38	99.49	0.90	85.79	95.73	99.36	0.93
V	88.13	97.14	99.30	0.94	88.13	97.14	99.30	0.94
F	83.20	98.11	99.83	0.92	91.20	95.40	99.52	0.95
Q	99.03	64.36	86.03	0.93	98.26	78.85	93.29	0.96
Average	87.03	91.06	96.87	0.92	91.73	93.11	98.14	0.95

Results show that there is no clear difference between using overall accuracy and using ROC area as performance criteria of selection algorithm. This may be limitation of feature pool. Results show that the best performance is obtained with

features of GA selection using overall accuracy as performance measure. Therefore, first stage of the system is constructed using the best features sets.

In the second stage, the feature selection algorithm is also applied in the second stage to select optimal features for subgroups of each main group. The best discriminative features are shown in Table 5.14. Results of overall accuracy of subgroup classifiers are %100, %95.86, %99.42 and %99.01 for subgroups of N, subgroups of S, subgroups of V and subgroups of Q, respectively.

Table 5.14 Number of selected features for subgroups of each main group using ROC area as performance measure

Main group	Label	Selected Features of Subgroups
Main group N	1	1 MF, 1 HOSofWPT, 2 FTF
	2	1 HOSofWPT, 2 FTF
	3	2 FTF
	4	2 MF, 1 HOSofWPT, 41 FTF
	5	1 MF, 2 HOSofWPT, 1 FTF
Main group S	6	1 HOSF, 1 MF, 1 HOSofWPT, 7 FTF
	7	1 MF, 1 HOSofWPT, 4 FTF
	8	1 HOSF, 1 MF, 2 HOSofWPT, 6 FTF
	9	1 HOSF, 20 FTF
Main group V	10	10 HOSofWPT, 9 FTF
	11	3 HOSofWPT, 5 FTF
	12	2 MF, 7 HOSofWPT, 34 FTF
Main group F	13	2 HOSF, 3 MF, 41 HOSofWPT, 22 FTF
Main group Q	14	2 HOSF, 10 HOSofWPT, 38 FTF
	15	4 HOSofWPT, 39 FTF
	16	5 HOSofWPT, 14 FTF

*HOSF: Higher Order Statistics Features,
*MF: Morphological Features,
*HOSofWPT:HOS Features of Wavelet Packet Transform,
*FTF: Fourier Transform Features

Results of second stage using ROC area as performance criteria of selection algorithm are almost same as using overall accuracy as performance criteria of selection algorithm. The second stage of the system is constructed overall accuracy as performance criteria of selection algorithm.

5.6.2 Feature Dimension Reduction with Neural Network

A simple three layer linear network can be used as feature dimension reduction tool. Figure 5.19 shows the structure of three layer neural network. All of the extracted features are used at both input and output layer. Network trained by gradient descent on a sum squared error criterion. Activation functions of network are linear for all layers. Dimension of input data is 150. The transformation is linear projection onto a different dimensional subspace which is represented by hidden neuron size of network.
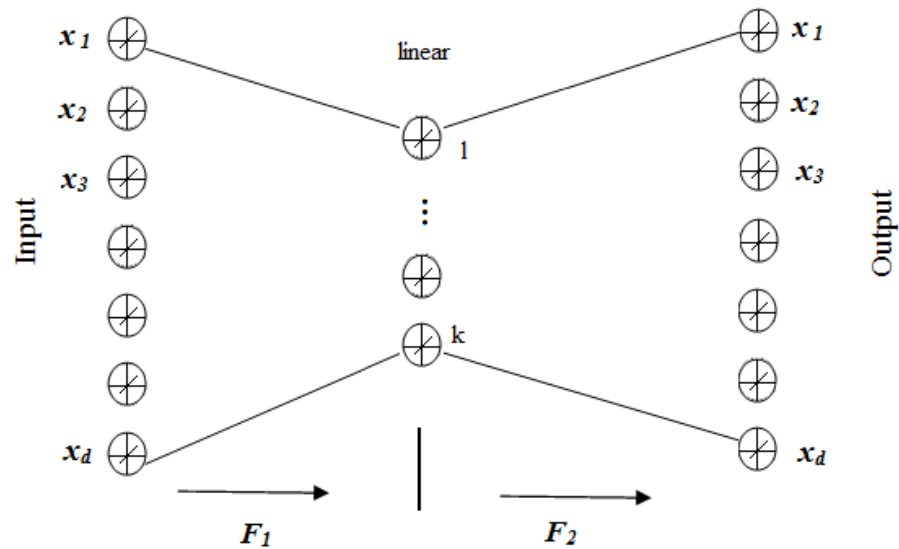
Figure 5.19 Three layer neural networks with linear hidden layer.

In this work, many neural networks with different hidden neuron size are constructed to investigate lower dimensional features set. The hidden neuron size, which represents new and reduced feature vector size, changes between 30 and 150. Each neural network with new hidden neuron is performed several times with different initial values of the network. Network with the best value of mean squared error (MSE) is used for new feature set extraction. Extracted feature sets with different sizes are used to investigate the classification accuracy of five main groups. Multilayer perceptron (MLP) networks are used as classifier. MLP structure is constructed as shown Figure 5.20 and Figure 5.21. The first model (in Figure 5.20) used single MLP structures with five outputs. The second model (in Figure 5.21)

used five MLP structures. Each network separates only one main group from others. Then the outputs of the networks are combined to make a decision.
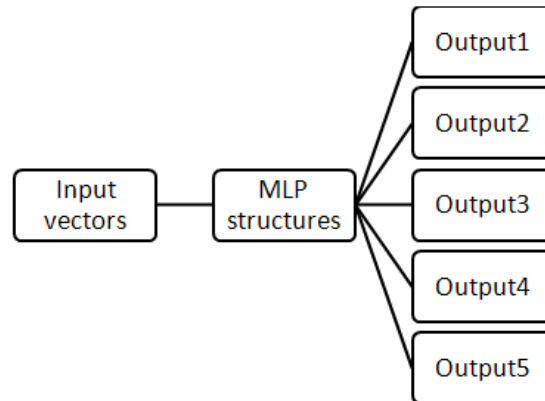
Figure 5.20 Classification systems for five classes with single MLP structure.
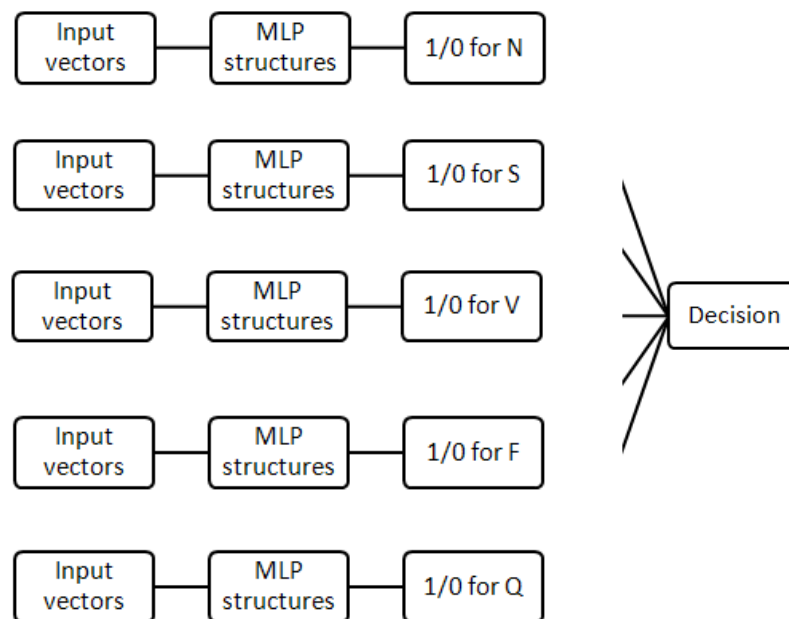
Figure 5.21 Classification system for five classes with five different MLP.

At first, the single MLP network is constructed with different hidden neuron size using all features as input vector. Each network is trained several times with different

initial values. The best values of overall accuracy of classifiers are shown in Table 5.15.

Table 5.15 Results of overall accuracy of single MLP classifiers using all features

		Overall Accuracy (%)
Hidden Neuron size of MLP classifier	5	72.90
	8	73.72
	11	78.48
	14	81.78
	17	**82.49**
	20	80.80
	23	80.61
	26	76.75
	29	81.00
	32	78.99
	35	77.93
	38	79.71
	41	77.42
	44	79.94
	47	79.31
	50	76.44

The single MLP network is constructed with different hidden neuron size and with different input vectors to classify five main groups with better performances. Each network is trained several times with different initial values. The best values of overall accuracy of classifiers are shown in Table 5.16.

The MLP network is constructed for each main group. Each network is built with 19 different hidden neurons and trained for 13 new feature sets. Therefore, total of 247 MLP structure are constructed for each main group. Each MLP structure is trained several times with different initial values of the network. The best values of overall accuracy of classifiers are shown in Table 5.17 for N, in Table 5.18 for S, in Table 5.19 for V, in Table 5.20 for F, and in Table 5.21 for Q.

Table 5.16 Result of overall accuracy of single MLP classifiers for five main groups using different input vectors

Hidden Neuron size of MLP classifier	New Feature Size												
	30	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**
2	71.32	64.42	55.48	71.91	73.60	59.99	56.14	60.30	73.01	63.97	59.36	72.19	64.26
5	76.63	74.00	78.52	83.28	83.91	81.83	85.17	83.08	84.15	84.62	81.43	84.03	84.66
8	75.45	75.26	81.00	85.56	81.59	83.44	**86.90**	83.32	82.97	81.04	84.26	84.42	83.12
11	76.28	75.61	80.88	81.47	81.55	81.75	84.58	83.01	84.03	84.38	83.87	84.74	82.69
14	77.07	72.78	81.08	83.48	80.25	80.45	83.40	83.36	83.28	83.91	82.97	82.10	81.67
17	75.10	74.35	80.57	82.53	81.04	81.67	83.48	82.18	83.04	83.24	82.57	82.73	83.67
20	73.72	77.93	80.68	80.72	80.72	83.08	83.48	81.98	81.63	84.82	82.97	83.63	83.12
23	76.44	73.52	79.90	82.57	81.71	82.14	84.85	82.69	82.53	84.74	83.48	83.24	83.75
26	74.59	76.12	81.75	77.58	80.02	81.94	84.15	80.92	82.65	83.71	83.36	84.26	82.22
29	75.10	79.31	81.63	82.69	82.22	82.57	84.11	82.26	82.65	84.15	83.40	82.65	83.08
32	74.00	75.18	83.01	81.16	79.78	81.55	83.87	81.86	82.49	83.40	83.16	83.91	80.49
35	74.67	80.65	81.20	83.12	80.76	81.71	85.72	84.03	81.86	83.60	83.83	83.71	81.55
38	76.55	77.22	79.39	82.49	83.52	83.44	83.79	82.38	83.08	84.62	83.48	85.09	83.40
41	75.18	78.32	80.49	81.75	81.55	81.98	84.07	82.49	82.93	84.89	85.21	83.71	83.08
44	75.53	72.58	82.85	81.39	78.99	80.29	86.00	82.30	84.70	84.97	83.79	83.32	82.18
47	80.68	76.40	81.27	82.89	81.12	82.77	83.87	83.60	84.11	84.34	82.89	84.34	82.69
50	73.80	79.78	82.10	82.22	82.53	82.53	85.29	83.83	81.43	86.15	82.49	84.23	83.71
FS	75.81	81.20	81.35	81.83	81.16	80.25	83.16	80.33	83.16	83.48	84.97	83.99	83.28
2xFS	79.43	75.45	81.67	82.53	82.34	81.20	83.28	82.93	82.93	85.33	84.62	82.57	82.77

*FS=Feature Size

Table 5.17 Result of overall accuracy of classifiers for N

Hidden Neuron size of MLP classifier	New Feature Size												
	150	140	130	120	110	100	90	80	70	60	50	40	30
2	93.63	94.37	94.40	93.08	94.61	94.49	92.04	94.81	92.76	93.30	91.90	87.69	91.19
5	94.69	95.28	94.14	94.34	94.45	94.73	95.32	95.16	93.55	95.44	92.13	89.93	92.72
8	94.65	93.94	93.39	94.49	94.61	95.40	94.53	95.16	92.68	95.08	92.68	90.40	91.62
11	93.71	94.18	94.81	95.28	94.53	95.16	94.89	94.69	93.67	95.44	91.97	89.14	92.96
14	94.53	94.53	95.00	95.08	95.44	94.89	95.20	95.71	93.15	95.44	91.97	89.30	93.47
17	94.26	94.37	95.04	94.49	95.55	95.20	95.48	95.67	93.63	94.96	92.01	93.19	92.45
20	95.00	94.26	94.30	95.40	94.65	94.73	95.08	94.89	94.37	95.52	91.74	93.27	93.90
23	94.73	94.61	95.16	94.49	94.61	95.52	94.85	95.52	93.90	95.59	91.50	91.11	94.81
26	95.08	94.34	94.73	94.45	94.77	94.85	95.16	95.71	92.80	95.63	91.15	92.13	92.41
29	95.00	94.34	95.48	94.81	94.96	95.36	94.96	95.71	93.55	95.52	92.21	88.51	93.82
32	94.34	94.57	95.20	95.55	94.69	94.69	95.12	95.48	94.57	95.32	91.90	90.64	91.35
35	95.16	94.06	94.89	94.93	95.40	94.73	94.89	94.96	93.55	95.36	91.23	89.38	90.91
38	94.85	93.04	94.57	94.41	95.04	94.96	95.67	95.79	93.98	95.75	92.09	89.30	91.46
41	94.18	94.30	94.45	95.11	95.36	95.67	94.96	94.83	94.34	94.77	92.84	90.52	91.70
44	94.26	94.14	94.73	94.57	95.32	95.52	94.96	95.83	94.14	95.44	93.27	90.68	92.72
47	94.96	94.61	94.73	94.34	94.85	94.93	94.10	95.28	93.31	95.32	91.58	88.83	92.68
50	94.18	94.18	94.61	95.11	94.61	95.55	94.69	95.37	93.63	95.20	91.07	87.84	93.55
FS	95.48	95.03	94.77	95.24	95.12	94.81	94.37	95.22	93.31	94.87	91.86	88.75	94.41
2xFS	94.96	93.90	95.08	95.24	95.08	94.73	95.16	95.44	94.22	95.55	91.66	89.73	92.01

*FS=Feature Size

Table 5.18 Result of overall accuracy of classifiers for S

		New Feature Size											
Hidden Neuron size of MLP classifier		150	140	130	120	110	100	90	80	70	60	50	40	30
	2	88.39	90.01	87.25	86.55	87.92	90.09	89.93	85.33	88.55	89.42	89.46	88.95	86.39
	5	87.53	91.62	89.65	91.54	86.00	89.89	89.77	86.94	85.41	90.60	88.83	89.61	88.55
	8	86.39	92.33	90.05	88.43	86.94	88.32	89.85	88.99	88.32	85.17	89.81	92.45	87.53
	11	84.66	88.87	88.39	89.93	87.33	88.59	90.76	86.51	90.13	84.07	90.60	89.50	87.77
	14	86.15	87.80	90.20	91.11	86.11	88.20	89.65	89.46	88.04	84.07	90.36	**94.73**	85.92
	17	85.60	90.83	88.79	88.24	87.29	89.30	89.89	88.63	85.25	90.95	90.20	93.15	84.07
	20	91.11	92.01	86.55	90.24	85.88	91.23	88.67	86.35	85.64	90.44	90.28	88.20	88.55
	23	83.91	91.19	89.22	90.32	87.69	89.26	88.08	79.15	90.28	87.96	90.48	88.67	85.01
	26	89.30	90.64	86.86	90.28	87.10	89.46	91.19	89.06	87.18	86.03	89.46	91.62	85.88
	29	90.79	92.56	87.29	90.17	87.80	89.89	87.53	88.87	88.71	89.38	88.67	87.77	86.39
	32	87.21	92.01	88.83	90.91	91.11	89.54	90.05	88.24	86.23	90.01	88.67	92.80	86.90
	35	87.69	90.99	88.47	89.65	86.03	89.93	88.24	80.06	88.04	83.95	90.64	91.86	86.70
	38	90.40	88.59	89.42	87.73	86.07	90.28	91.38	81.04	87.80	87.53	87.53	91.23	89.14
	41	87.33	90.87	91.50	87.41	89.14	90.01	90.17	82.57	87.57	90.48	89.38	93.19	86.23
	44	89.10	91.07	87.92	89.34	88.08	91.31	88.51	81.43	86.98	88.12	88.99	92.68	88.79
	47	82.42	87.65	82.57	89.54	86.82	89.54	89.30	85.09	85.88	85.64	89.54	87.45	86.86
	50	88.32	89.02	89.54	90.24	89.54	89.26	88.12	85.56	86.78	89.58	92.13	90.60	85.25
	FS	86.66	90.48	87.21	91.70	84.62	90.20	90.52	80.68	85.33	84.66	91.62	91.03	89.18
	2xFS	85.60	86.39	85.60	87.92	87.33	89.58	89.69	85.60	85.60	88.99	90.44	86.23	85.76

*FS=Feature Size

Table 5.19 Result of overall accuracy of classifiers for V

		New Feature Size												
Hidden Neuron size of MLP classifier		150	140	130	120	110	100	90	80	70	60	50	40	30
	2	91.90	94.15	94.02	93.43	94.11	93.59	91.50	93.59	94.05	94.07	92.41	87.77	89.14
	5	92.13	91.50	91.86	94.37	91.66	93.19	93.94	90.87	94.02	92.17	90.95	87.88	90.20
	8	91.50	91.90	93.71	92.88	92.09	91.82	92.13	94.30	90.68	93.19	92.09	89.02	88.87
	11	90.91	93.12	92.64	93.08	91.50	92.25	91.50	91.94	90.32	92.33	91.78	86.74	88.51
	14	90.64	92.09	91.27	91.94	91.90	91.54	91.27	91.50	91.38	91.54	91.54	87.73	89.38
	17	90.09	91.42	90.91	92.96	92.60	92.09	91.23	90.60	91.03	92.45	91.31	87.73	94.02
	20	90.99	91.23	90.87	93.04	92.21	90.64	92.21	91.78	92.56	92.33	91.38	87.37	90.05
	23	90.05	91.66	92.05	92.33	92.96	91.27	90.36	91.38	91.38	92.92	90.76	88.63	91.03
	26	90.13	92.29	93.47	92.53	92.29	91.07	91.19	90.56	91.62	92.21	90.60	88.91	90.48
	29	90.20	91.58	92.29	92.41	91.78	91.27	90.72	91.46	91.07	92.17	91.03	87.96	88.51
	32	91.03	91.35	92.13	91.74	92.64	91.82	91.46	91.62	90.20	92.17	90.44	86.82	86.03
	35	90.05	91.54	92.41	92.84	92.60	91.31	91.19	90.79	91.27	92.25	89.38	87.77	87.92
	38	91.11	91.07	90.64	92.33	91.90	92.21	91.82	91.15	91.27	91.11	91.94	88.43	90.83
	41	89.85	91.62	91.94	91.78	92.33	89.97	90.95	90.72	90.79	93.00	92.25	89.06	86.31
	44	89.85	91.11	92.72	92.29	92.88	92.21	91.38	90.36	90.76	92.33	90.44	87.18	87.41
	47	90.17	91.50	92.09	92.80	92.25	90.56	91.03	90.48	91.90	91.82	91.42	87.37	88.75
	50	90.44	91.19	91.94	93.00	90.99	91.54	90.99	90.79	90.13	92.21	91.38	88.59	89.50
	FS	89.81	91.27	90.28	92.56	91.74	90.56	92.37	91.27	90.72	92.21	91.54	86.66	89.58
	2xFS	88.51	78.80	91.78	91.66	92.01	90.72	91.07	90.91	91.42	92.17	91.86	87.06	88.47

*FS=Feature Size

Table 5.20 Result of overall accuracy of classifiers for F

		New Feature Size											
Hidden Neuron size of MLP classifier		30	40	50	60	70	80	90	100	110	120	130	140	150
	2	90.95	90.76	90.95	92.60	90.76	92.29	91.66	91.70	91.38	92.05	92.33	91.23	91.78
	5	91.54	91.23	90.87	90.44	90.83	91.82	91.42	91.54	91.54	91.66	91.74	91.35	91.82
	8	91.90	90.99	90.40	89.85	90.87	91.78	91.31	91.38	91.31	91.82	91.54	91.31	91.27
	11	90.48	91.15	90.52	90.91	90.72	91.62	91.19	91.42	91.54	91.35	91.62	91.78	91.90
	14	90.48	91.86	90.48	90.76	90.64	91.78	91.86	91.11	91.46	91.62	92.01	91.42	91.46
	17	91.19	91.11	90.68	90.91	90.83	91.70	92.05	91.38	91.38	91.58	91.90	91.19	91.74
	20	90.79	91.50	90.44	91.50	90.95	91.90	91.38	91.42	91.86	91.74	91.19	91.38	91.70
	23	89.77	90.64	90.40	90.60	91.07	91.58	91.46	91.35	91.23	92.49	91.54	91.42	91.50
	26	90.76	90.48	90.95	91.07	91.15	91.62	91.54	91.42	91.07	91.27	91.70	91.31	91.70
	29	90.68	90.56	90.40	92.25	90.17	91.58	91.50	93.75	91.23	91.90	91.58	91.50	91.66
	32	90.83	91.58	90.01	91.27	90.79	91.54	91.31	91.38	91.35	91.50	92.01	91.27	91.66
	35	90.44	91.11	90.76	90.52	91.15	91.86	91.54	91.27	91.66	91.58	91.31	91.38	91.35
	38	90.95	91.74	90.56	90.68	90.60	91.70	91.90	91.15	91.35	91.54	91.35	91.38	95.87
	41	91.58	90.60	90.52	90.95	91.03	91.74	91.19	91.27	91.31	91.07	91.54	90.87	91.54
	44	90.32	90.44	90.32	91.78	90.91	91.74	91.23	91.38	91.23	91.31	92.13	91.27	91.54
	47	90.24	90.64	89.89	90.95	91.11	91.86	91.46	91.38	90.99	91.31	91.62	91.42	91.38
	50	90.83	90.28	90.24	91.66	90.87	91.62	91.31	91.31	91.38	93.31	91.86	91.23	91.66
	FS	89.73	92.09	90.36	92.45	90.99	91.70	91.19	91.15	91.19	91.66	91.42	90.79	91.70
	2xFS	91.19	90.17	90.40	90.17	91.15	91.66	91.23	91.38	91.11	94.06	91.42	90.91	90.17

*FS=Feature Size

Table 5.21 Result of overall accuracy of classifiers for Q

Hidden Neuron size of MLP classifier	New Feature Size												
	150	140	130	120	110	100	90	80	70	60	50	40	30
2	99.02	98.98	99.25	99.31	99.21	99.02	99.33	99.10	98.66	98.58	98.94	96.30	98.35
5	99.37	99.13	99.25	99.33	99.29	99.06	99.33	99.02	98.70	98.66	99.02	96.58	98.11
8	99.25	99.10	99.33	99.33	99.31	98.98	99.25	99.02	98.70	98.82	98.62	96.73	98.03
11	99.37	99.17	99.29	99.17	99.29	99.10	99.25	99.06	98.66	98.62	99.10	96.42	98.03
14	99.25	99.02	99.21	99.25	99.34	99.13	99.29	98.70	98.74	98.43	98.90	96.66	97.92
17	99.37	99.21	99.29	99.29	99.37	99.02	99.29	98.54	98.74	98.51	98.78	96.66	97.88
20	99.34	98.98	99.17	99.25	99.33	99.17	99.29	99.17	98.82	98.94	99.17	97.25	97.72
23	99.37	99.29	99.33	99.21	99.33	99.10	99.33	99.21	98.62	98.70	98.90	96.34	98.03
26	99.21	99.06	99.33	99.25	99.29	99.06	99.45	98.90	98.58	98.86	99.06	96.81	98.27
29	99.37	99.02	99.25	99.06	99.37	99.06	99.33	99.10	98.74	98.70	98.90	96.77	97.92
32	99.35	98.86	99.21	99.25	99.29	99.02	99.37	98.94	98.58	98.78	98.15	96.14	98.19
35	99.17	99.02	99.29	99.02	99.37	99.10	99.29	99.02	98.70	98.70	99.06	96.70	97.84
38	99.29	98.94	99.29	99.17	99.21	99.02	99.29	98.62	98.66	98.58	98.43	97.56	98.11
41	99.29	98.66	99.29	99.21	99.33	99.02	99.25	99.13	98.70	98.66	98.94	96.58	98.27
44	99.10	98.86	99.33	99.33	99.33	99.06	99.31	98.54	98.82	98.78	98.62	96.85	98.54
47	99.17	99.33	99.29	99.29	99.37	99.17	99.33	98.94	98.66	98.78	98.94	97.01	97.92
50	99.37	98.98	99.21	99.21	99.37	99.02	99.29	99.02	98.66	98.74	99.21	96.70	97.72
FS	99.34	98.86	99.37	99.33	99.33	99.02	99.29	98.90	98.78	98.66	99.06	96.34	97.88
2xFS	99.33	98.27	99.37	99.37	99.29	98.98	99.21	98.98	98.70	98.70	99.17	95.44	98.03

*FS=Feature Size

The best performance of the single MLP structures is 82.49% of overall accuracy with 17 hidden neurons using all features.

The best performance of the single MLP networks is 86.90% of overall accuracy with 90 new feature vector size and MLP structures with 8 hidden neurons.

In the case of different MLPs for each main group, the best performance is 95.83% obtained with feature vector size of 80 and MLP structure with 8 hidden neurons for N; the best performance is 94.73% obtained with feature vector size of 40 and MLP structure with 14 hidden neurons for S; the best performance is 94.37% obtained with feature vector size of 120 and MLP structure with 5 hidden neurons for V; the best performance is 95.87% obtained with feature vector size of 150 and MLP structure with 35 hidden neurons for F ; the   best performance is 99.45% obtained feature vector size of 90 and MLP structure with 26 hidden neurons for Q. Networks with best performance are used to construct an ensemble system and the outputs of the networks are combined to make a decision.

The MLP network is constructed for each main group. Each network is built with 19 different hidden neurons and trained for 13 new feature sets. Therefore, total of 247 MLP structure are constructed for each main group. Each MLP structure is trained several times with different initial values of the network. The best values of ROC area measure of classifiers are shown in Table 5.22 for N, in Table 5.23 for S, in Table 5.24 for V, in Table 5.25 for F, and in Table 5.26 for Q.

Table 5.22 Result of ROC area measure of classifiers for N

Hidden Neuron size of MLP classifier	New Feature Size												
	30	40	50	60	70	80	90	100	110	120	130	140	150
2	0.9228	0.8846	0.9106	0.9363	0.9172	0.9421	0.9356	0.9345	0.9348	0.9431	0.9408	0.9336	0.9321
5	0.9352	0.8992	0.9118	0.9483	0.9111	0.9385	0.9388	0.9393	0.9276	0.9283	0.9362	0.9446	0.9340
8	0.9132	0.8999	0.9163	0.9408	0.9119	0.9404	0.9301	0.9372	0.9252	0.9317	0.9248	0.9185	0.9304
11	0.9186	0.9073	0.9131	0.9447	0.9136	0.9357	0.9325	0.9338	0.9304	0.9369	0.9310	0.9257	0.9304
14	0.9217	0.8922	0.9082	0.9474	0.9097	0.9465	0.9377	0.9281	0.9422	0.9318	0.9309	0.9290	0.9246
17	0.9374	0.9249	0.9035	0.9458	0.9136	0.9410	0.9403	0.9393	0.9387	0.9229	0.9334	0.9267	0.9327
20	0.9305	0.9233	0.9025	**0.9507**	0.9209	0.9438	0.9357	0.9290	0.9263	0.9389	0.9256	0.9222	0.9304
23	0.9446	0.9293	0.9065	0.9492	0.9143	0.9429	0.9314	0.9392	0.9271	0.9243	0.9354	0.9282	0.9247
26	0.9198	0.9140	0.9147	0.9486	0.9021	0.9427	0.9415	0.9289	0.9299	0.9215	0.9363	0.9212	0.9321
29	0.9357	0.9015	0.9089	0.9405	0.9127	0.9427	0.9328	0.9350	0.9334	0.9272	0.9414	0.9239	0.9304
32	0.8927	0.9060	0.9100	0.9499	0.9246	0.9440	0.9393	0.9255	0.9264	0.9398	0.9352	0.9312	0.9231
35	0.9057	0.8940	0.9114	0.9444	0.9149	0.9413	0.9355	0.9269	0.9386	0.9298	0.9444	0.9177	0.9332
38	0.8955	0.8958	0.9077	0.9479	0.9254	0.9468	0.9451	0.9325	0.9312	0.9215	0.9351	0.9225	0.9278
41	0.9204	0.9056	0.9101	0.9437	0.9250	0.9446	0.9344	0.9402	0.9391	0.9441	0.9229	0.9212	0.9193
44	0.9092	0.9024	0.9211	0.9496	0.9213	0.9485	0.9334	0.9392	0.9347	0.9268	0.9372	0.9211	0.9189
47	0.9160	0.8959	0.9060	0.9460	0.9159	0.9435	0.9321	0.9284	0.9390	0.9207	0.9313	0.9252	0.9378
50	0.9296	0.8836	0.9105	0.9481	0.9188	0.9445	0.9459	0.9384	0.9292	0.9470	0.9276	0.9170	0.9265
FS	0.9399	0.9030	0.9105	0.9493	0.9126	0.9409	0.9371	0.9270	0.9371	0.9360	0.9381	0.9497	0.9381
2xFS	0.9145	0.9038	0.9109	0.9459	0.9233	0.9441	0.9390	0.9277	0.9384	0.9360	0.9368	0.9332	0.9326

*FS=Feature Size

Table 5.23 Result of ROC area measure of classifiers for S

| | New Feature Size | | | | | | | | | | | | |
Hidden Neuron size of MLP classifier	150	140	130	120	110	100	90	80	70	60	50	40	30
2	0.8333	0.8098	0.7540	0.7429	0.7755	0.7966	0.7809	0.7098	0.7926	0.7395	0.7563	0.6888	0.5791
5	0.8113	0.8613	0.8373	0.8551	0.7504	0.8227	0.8434	0.7578	0.8420	0.7326	0.7659	0.7087	0.7594
8	0.8546	0.8756	0.8294	0.8199	0.8124	0.8302	0.8419	0.7891	0.8646	0.6933	0.7952	0.8513	0.7703
11	0.8149	0.8509	0.8168	0.8355	0.7851	0.8550	0.8437	0.7541	0.8582	0.6194	0.8155	0.7921	0.6686
14	0.8157	0.8464	0.8121	0.8435	0.8127	0.8617	0.8203	0.8384	0.8313	0.6850	0.8031	0.8784	0.7492
17	0.8454	0.8544	0.8084	0.7621	0.7644	0.8497	0.8617	0.8336	0.8241	0.7847	0.7796	0.8373	0.6899
20	0.8572	0.8706	0.7851	0.8157	0.7857	0.8738	0.8122	0.7442	0.8282	0.7919	0.7989	0.7542	0.6775
23	0.8019	0.8787	0.8564	0.8730	0.7997	0.8734	0.8224	0.6134	0.8222	0.7422	0.7899	0.6906	0.6986
26	0.8457	0.8673	0.8227	0.8478	0.7849	0.8688	0.8054	0.8032	0.8467	0.7071	0.7714	0.8442	0.6550
29	0.8610	0.8783	0.8301	0.8323	0.8072	0.8479	0.8181	0.8179	0.8533	0.7186	0.7622	0.6787	0.6725
32	0.8071	**0.8886**	0.8268	0.8495	0.8401	0.8682	0.8203	0.7711	0.8341	0.7874	0.7990	0.8170	0.6714
35	0.8248	0.8699	0.8247	0.8729	0.7627	0.8491	0.7915	0.6960	0.8551	0.7139	0.7885	0.8479	0.7327
38	0.8352	0.8288	0.8106	0.8159	0.7937	0.8784	0.8485	0.6847	0.8251	0.7442	0.7649	0.8419	0.7082
41	0.8314	0.8706	0.8629	0.8753	0.8218	0.8662	0.8664	0.7780	0.8547	0.7887	0.7768	0.8580	0.6963
44	0.8545	0.8640	0.8310	0.8445	0.7963	0.8333	0.7532	0.6870	0.8454	0.7718	0.7936	0.8232	0.6824
47	0.8274	0.8324	0.8047	0.8658	0.7556	0.8711	0.7886	0.7508	0.8354	0.6957	0.7769	0.7808	0.7019
50	0.8363	0.8704	0.8457	0.8799	0.8332	0.8395	0.8600	0.7531	0.8319	0.7380	0.7999	0.7667	0.6059
FS	0.8289	0.8683	0.6821	0.8598	0.7443	0.8655	0.8605	0.6906	0.8291	0.6683	0.7999	0.7942	0.7027
2xFS	0.5500	0.8535	0.7257	0.8185	0.7703	0.8596	0.7409	0.6055	0.8187	0.7993	0.7748	0.7977	0.6535

*FS=Feature Size

Table 5.24 Result of ROC area measure of classifiers for V

Hidden Neuron size of MLP classifier	New Feature Size												
	30	40	50	60	70	80	90	100	110	120	130	140	150
2	0.8138	0.8085	0.8955	0.9071	0.9153	0.8957	0.8953	0.9037	0.9094	0.9142	0.9291	0.9297	0.8931
5	0.8416	0.8088	0.8900	0.9288	0.9234	0.8916	0.9256	0.9127	0.9004	**0.9317**	0.9147	0.9225	0.9175
8	0.8432	0.8354	0.9088	0.9127	0.9042	0.9130	0.9230	0.9027	0.9132	0.9243	0.9214	0.9154	0.9149
11	0.8376	0.8108	0.9118	0.9018	0.8976	0.9014	0.9122	0.9122	0.9101	0.9181	0.9242	0.9251	0.9057
14	0.8673	0.8226	0.9087	0.8952	0.9026	0.9013	0.9148	0.8921	0.9090	0.9138	0.9093	0.9186	0.9064
17	0.8895	0.8182	0.9074	0.9141	0.9139	0.8877	0.9085	0.9007	0.9192	0.9228	0.9091	0.9117	0.8972
20	0.8480	0.8253	0.8979	0.9100	0.9098	0.9078	0.9163	0.8982	0.9151	0.9233	0.9123	0.9192	0.9105
23	0.8543	0.8305	0.8999	0.9083	0.9033	0.9060	0.9022	0.9042	0.9235	0.9207	0.9173	0.9200	0.9140
26	0.8657	0.8408	0.9180	0.8978	0.9109	0.9055	0.9129	0.8986	0.9151	0.9288	0.9233	0.9226	0.9048
29	0.8232	0.8345	0.9042	0.9076	0.8952	0.9058	0.9031	0.8998	0.9134	0.9052	0.9206	0.9134	0.9100
32	0.8528	0.8174	0.8968	0.9069	0.9147	0.8987	0.9174	0.9092	0.9205	0.9190	0.9202	0.9098	0.9064
35	0.8641	0.8335	0.8892	0.9129	0.9070	0.8989	0.9147	0.8919	0.9230	0.9207	0.9198	0.9165	0.9023
38	0.8645	0.8195	0.9113	0.8916	0.9107	0.8940	0.9088	0.9033	0.9128	0.9118	0.9047	0.9163	0.9076
41	0.8285	0.8343	0.9190	0.9076	0.8968	0.8953	0.9053	0.8970	0.9071	0.9187	0.9185	0.9177	0.9071
44	0.8573	0.8142	0.9143	0.9041	0.9013	0.8883	0.9116	0.9079	0.9141	0.9179	0.9209	0.9178	0.8936
47	0.8567	0.8074	0.8961	0.9084	0.9126	0.8953	0.9094	0.8987	0.9156	0.9177	0.9193	0.9142	0.9037
50	0.8498	0.8290	0.9118	0.9147	0.9006	0.9009	0.9035	0.9023	0.8976	0.9264	0.9183	0.9132	0.9103
FS	0.8640	0.8276	0.9118	0.8963	0.9018	0.8910	0.9068	0.8891	0.9118	0.9237	0.9031	0.9161	0.9068
2xFS	0.8475	0.8209	0.9076	0.9022	0.9090	0.8969	0.8921	0.8950	0.9114	0.9280	0.9180	0.9121	0.8830

*FS=Feature Size

Table 5.25 Result of ROC area measure of classifiers for F

Hidden Neuron size of MLP classifier	New Feature Size												
	30	40	50	60	70	80	90	100	110	120	130	140	150
2	0.6701	0.6423	0.6612	0.7077	0.5995	0.6383	0.6205	0.6225	0.5941	0.6263	0.6510	0.5930	0.6165
5	0.7197	0.5985	0.5857	0.5874	0.5821	0.5983	0.5872	0.5921	0.6123	0.6272	0.6094	0.5928	0.6000
8	0.6468	0.6097	0.5828	0.5985	0.5866	0.6052	0.5930	0.6309	0.6052	0.6000	0.5989	0.5919	0.6023
11	0.6013	0.5968	0.5819	0.6121	0.5850	0.5961	0.5948	0.5941	0.6021	0.6164	0.6090	0.6372	0.6183
14	0.6425	0.6787	0.5891	0.6168	0.5828	0.5927	0.6127	0.5890	0.6034	0.6364	0.6225	0.5936	0.6052
17	0.6104	0.5971	0.5839	0.5939	0.5861	0.5923	0.6281	0.5954	0.5899	0.6130	0.6112	0.5879	0.6067
20	0.6140	0.6019	0.5888	0.6085	0.5875	0.6040	0.6065	0.5961	0.6020	0.5979	0.6035	0.5861	0.6003
23	0.6237	0.5993	0.6031	0.6056	0.6030	0.5919	0.5909	0.5885	0.5950	0.6429	0.6168	0.5874	0.5929
26	0.6038	0.5945	0.5862	0.6368	0.5945	0.5972	0.5932	0.5919	0.5897	0.6059	0.5994	0.5901	0.6078
29	0.5991	0.6152	0.5851	0.6381	0.5881	0.5999	0.5943	0.7016	0.5992	0.6112	0.6019	0.5929	0.5992
32	0.6088	0.6144	0.5845	0.6077	0.5872	0.5941	0.5915	0.5908	0.6028	0.6054	0.6159	0.5899	0.5876
35	0.6746	0.6075	0.5852	0.6143	0.6057	0.6127	0.6341	0.6080	0.6079	0.6126	0.6025	0.5995	0.5952
38	0.6697	0.6301	0.5859	0.5875	0.5911	0.5889	0.6183	0.5935	0.6010	0.5985	0.6065	0.5855	0.8755
41	0.6023	0.5951	0.6486	0.6217	0.5870	0.5996	0.5930	0.5881	0.5915	0.5975	0.5976	0.5895	0.5959
44	0.6079	0.5866	0.6113	0.6479	0.5899	0.5996	0.6352	0.5972	0.5941	0.5977	0.6374	0.5983	0.5927
47	0.6091	0.6022	0.5849	0.5917	0.5890	0.5949	0.5945	0.5869	0.5903	0.5907	0.5954	0.5930	0.5905
50	0.6365	0.6059	0.5853	0.6366	0.5886	0.5881	0.5919	0.6005	0.5959	0.6885	0.6056	0.5965	0.5974
FS	0.6009	0.6373	0.5853	0.6748	0.5877	0.6003	0.5881	0.5856	0.5876	0.6068	0.5961	0.5855	0.5923
2xFS	0.6144	0.6019	0.6011	0.5842	0.6160	0.5938	0.5843	0.5861	0.5908	0.7283	0.5889	0.5897	0.5843

*FS=Feature Size

Table 5.26 Result of ROC area measure of classifiers for Q

Hidden Neuron size of MLP classifier	New Feature Size												
	30	40	50	60	70	80	90	100	110	120	130	140	150
2	0.9713	0.9624	0.9815	0.9803	0.9830	0.9835	0.9857	0.9844	0.9857	0.9869	0.9859	0.9835	0.9835
5	0.9723	0.9612	0.9816	0.9783	0.9829	0.9806	0.9857	0.9847	0.9869	0.9864	0.9857	0.9837	0.9859
8	0.9718	0.9602	0.9791	0.9813	0.9820	0.9830	0.9850	0.9842	0.9788	0.9864	0.9864	0.9849	0.9852
11	0.9718	0.9580	0.9828	0.9805	0.9822	0.9832	0.9852	0.9849	0.9862	0.9854	0.9857	0.9854	0.9862
14	0.9705	0.9595	0.9815	0.9793	0.9827	0.9796	0.9869	0.9852	0.9869	0.9859	0.9850	0.9837	0.9867
17	0.9693	0.9639	0.9830	0.9783	0.9834	0.9808	0.9862	0.9844	0.9867	0.9857	0.9862	0.9850	0.9867
20	0.9695	0.9612	0.9840	0.9832	0.9825	0.9833	0.9862	0.9840	0.9787	0.9866	0.9857	0.9830	0.9869
23	0.9710	0.9620	0.9830	0.9801	0.9832	0.9842	0.9864	0.9849	0.9864	0.9864	0.9864	0.9854	0.9869
26	0.9740	0.9651	0.9840	0.9820	0.9824	0.9827	0.9878	0.9835	0.9854	0.9845	0.9864	0.9847	0.9857
29	0.9710	0.9610	0.9815	0.9803	0.9825	0.9835	0.9821	0.9847	0.9859	0.9859	0.9859	0.9837	0.9867
32	0.9720	0.9585	0.9816	0.9815	0.9817	0.9825	0.9867	0.9837	0.9862	0.9859	0.9857	0.9820	0.9787
35	0.9698	0.9612	0.9818	0.9810	0.9825	0.9830	0.9862	0.9844	0.9867	0.9862	0.9862	0.9837	0.9852
38	0.9730	0.9667	0.9764	0.9796	0.9822	0.9805	0.9869	0.9844	0.9854	0.9862	0.9862	0.9840	0.9869
41	0.9718	0.9612	0.9847	0.9808	0.9825	0.9837	0.9849	0.9844	0.9864	0.9854	0.9862	0.9825	0.9866
44	0.9743	0.9644	0.9818	0.9793	0.9825	0.9800	0.9862	0.9840	0.9864	0.9864	0.9859	0.9820	0.9869
47	0.9738	0.9610	0.9832	0.9815	0.9827	0.9822	0.9857	0.9847	0.9862	0.9864	0.9854	0.9857	0.9862
50	0.9684	0.9605	0.9837	0.9820	0.9822	0.9823	0.9866	0.9844	0.9867	0.9850	0.9852	0.9835	0.9859
FS	0.9720	0.9568	0.9837	0.9825	0.9830	0.9823	0.9867	0.9837	0.9864	0.9857	0.9867	0.9835	0.9862
2xFS	0.9718	0.9619	0.9847	0.9830	0.9827	0.9835	0.9832	0.9842	0.9862	0.9787	0.9867	0.9805	0.9859

*FS=Feature Size

In the case of different MLPs for each main group, the best performance of ROC area is 0.9507 obtained with feature vector size of 60 and MLP structure with 20 hidden neurons for N; the best performance of ROC area is 0.8886 obtained with feature vector size of 140 and MLP structure with 32 hidden neurons for S; the best performance of ROC area is 0.9317 obtained with feature vector size of 130 and MLP structure with 5 hidden neurons for V; the best performance of ROC area is 0.8755% obtained with feature vector size of 150 and MLP structure with 38 hidden neurons for F ; the best performance of ROC area is 0.9878 obtained feature vector size of 90 and MLP structure with 26 hidden neurons for Q. Networks with best performance of ROC area are used to construct an ensemble system and the outputs of the networks are combined to make a decision.

Overall accuracy of single MLP classifier with all features, overall accuracy of single MLP classifier with new feature vector, overall accuracy of MLP based combined classifier using overall accuracy as performance criteria for each network and overall accuracy of MLP based combined classifier using ROC Area as performance criteria for each network are 82.49%, 86.90%, 91.54%, and 91.07%, respectively. The other performance measures are shown Table 5.27.

Table 5.27 Performance measures of MLP based classifiers

Label	Single MLP Classifier using all features				Single MLP Classifier using new features vector			
	SEN	SEL	SPE	ROC Area	SEN	SEL	SPE	ROC Area
N	82.43	95.73	98.09	90.26	90.59	98.26	99.16	94.87
S	83.33	59.11	90.30	86.82	93.17	73.81	94.44	93.80
V	91.47	74.02	91.36	91.41	90.54	76.97	92.71	91.62
F	31.60	76.70	98.95	65.28	35.20	72.73	98.56	66.88
Q	97.29	99.01	99.75	98.52	97.48	96.36	99.06	98.27
Average	77.22	80.92	95.69	86.46	81.39	83.63	96.79	89.09

Label	MLP Based combined Classifiers using Overall Acc. as performance criteria				MLP Based combined Classifiers using ROC Area as performance criteria			
	SEN	SEL	SPE	ROC Area	SEN	SEL	SPE	ROC Area
N	93.92	97.61	98.80	96.36	90.82	97.65	98.86	0.948
S	85.52	89.17	98.25	91.89	90.98	80.24	96.23	0.936
V	92.02	88.26	96.70	94.36	90.54	89.71	97.20	0.939
F	78.40	69.26	96.20	87.30	79.60	75.67	97.21	0.884
Q	97.67	99.21	99.80	98.74	97.67	98.82	99.70	0.987
Average	89.51	88.70	97.95	93.73	89.92	88.42	97.84	0.939

Results show that the ensemble classification system is performed better than single network classifier. MLP based feature extraction and reduction method and the performances of MLP classifiers are as good as KNN based classifiers with features selected with GA features. Measures of MLP based combined classifier using ROC Area as performance criteria are a bit better than the MLP based combined classifiers using overall accuracy as performance criteria of each network structure.

In the construction of the whole system, combined KNN classifier with the selected features by GA are used in the first stage of the classification system.

5.7   Classification

The classification model is an important part of the pattern recognition systems. It should be easily comprehensible and have high performance. The purpose of this

research is to explore methods for improving the performance of automated ECG beat classifiers. Therefore, the constructed system has two stages: in the first stage, it separates all beats to few classes which are called main groups (as shown in Table 5.2.) then in the second stage, each main group is further separated into subclasses (Kutlu, & Kuntalp, 2010). Proposed classification model is shown in Figure 5.22. In the first classifier stage, five KNN classifiers are used. One KNN classifier separates one main group from others. Then combination of individual decisions produces the final decision of the first stage. In the second classification stage, separated main groups are classified into the subgroups as similar to the first stage. Each main group is separated into subgroups by combined KNN classifiers. At the end of the constructed system final decision is evaluated. The constructed system with combined KNN classifiers is shown in Figure 5.23.

Each classifier in constructed system uses different input sets because of the beat-based feature selection algorithm. For the first stage classifier, the best input vectors are obtained from the GA based feature selection algorithm. At the second stage classifier, the input vectors are obtained from the SFSM based feature selection algorithm.
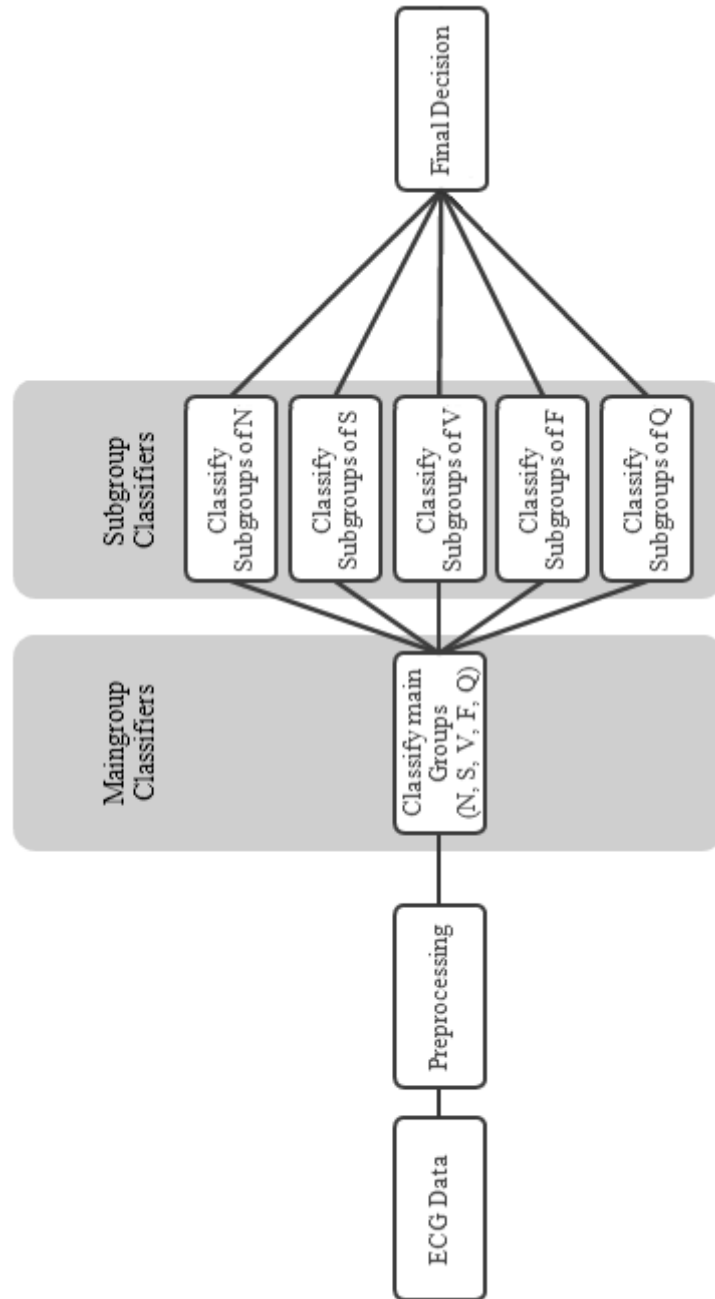
Figure 5.22 Two stage classification system.
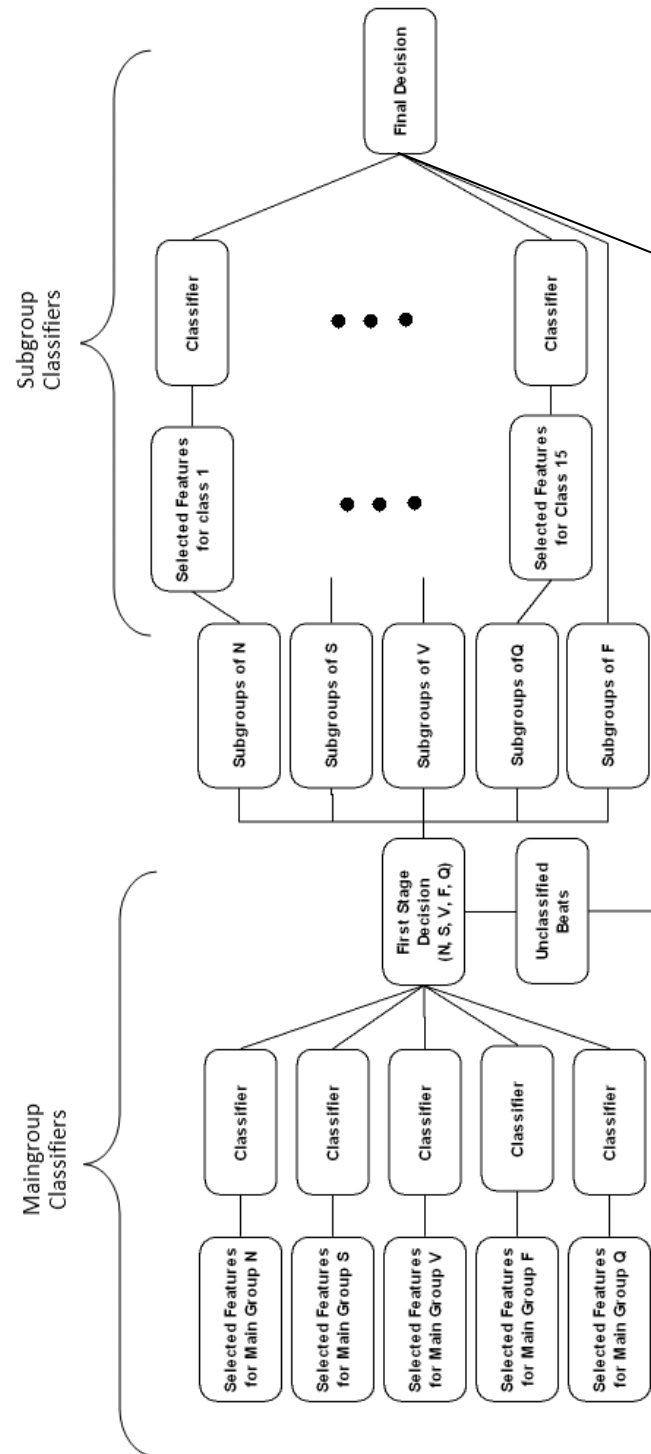
Figure 5.23 Detail of two stage Two stage classification system.

The purpose of this research is to explore methods for improving the performance of automated ECG beat classifiers. Therefore, the constructed system has two stages: in the first stage, it separates all beats to few classes which are called main groups then in the second stage, each main group is further separated into subclasses. But, combination of classifier may cause a contradiction and these beats from each stage are labeled as unclassified beat. This approach decreases the performance measures of classification model. Therefore, the third stage is added in the system to classified only unclassified beats into 16 beat types. KNN classifier and raw data as input vector is used in this stage. The constructed three stage system is shown in Figure 5.24.
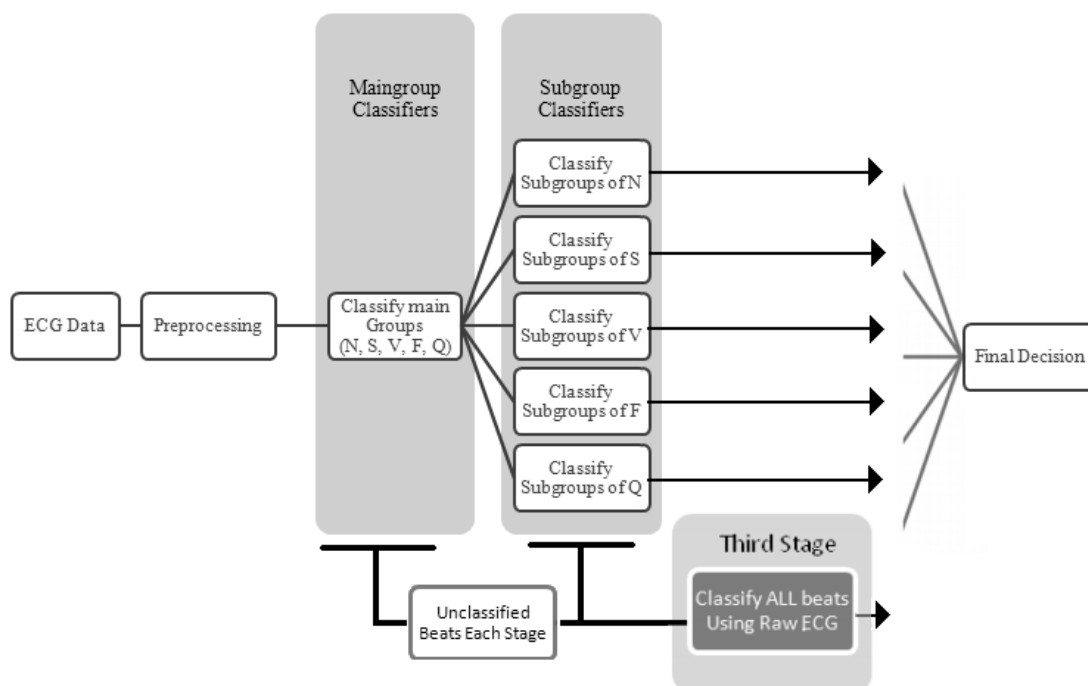
Figure 5.24 Three stage classification system.

5.8 Results and Discussion

In this implementation, a classification system is constructed for automatic heartbeat recognition. For this purpose, in feature extraction stage, total of 150 features were extracted from HOS measures, HOSofWPT measures, morphology measures, and frequency domain measures. After feature extraction stage, beat-based

feature selection is performed in two stages. In the first stage of classification, two different selection algorithms are used: GA based feature selection algorithm and SFSM base feature selection algorithm. Overall accuracy and ROC area measure of classifier are used as performance measure for selection algorithm. According to the results of these algorithms, the features which are obtained by GA algorithm provide better classification performance. Results of the selection algorithms are given in detail in Table 5.9. In addition, dimension of extracted features is reduced from 150 to different new dimensions using an MLP network. Results are discussed in detailed in section 5.6. Previously a single MLP classifier and combined MLP classifiers are used to classify five main groups of the first stage of classification using reduced dimension feature sets. The results show that the features obtained with GA and combined KNN classifiers have better results than MLP based classification process.

Consequently, the desired system is constructed with combined KNN classifier using the obtained features by GA in the first classifier stage. In the second classifier stage, also combined KNN classifiers using features obtained by SFSM are used.

Using all of the extracted features, the classification performance of the constructed two stage system is %79.50 for overall accuracy, and the other performance measures are given in Table 5.28.

Table 5.28 Performance measures of two stage system for all features

Labels	SEN	SEL	SPE	ROC Area
1	95.20	95.58	99.52	0.97
2	96.00	96.77	99.65	0.98
3	67.20	97.11	99.78	0.83
4	94.69	76.98	98.68	0.97
5	87.50	77.78	99.92	0.94
6	28.38	33.33	98.30	0.63
7	0.00	0.00	100.00	0.50
8	92.40	60.16	93.32	0.93
9	53.66	35.48	98.40	0.76
10	75.42	89.45	99.09	0.87
11	92.45	89.09	99.76	0.96
12	79.60	56.37	93.28	0.86
13	27.20	70.83	98.78	0.63
14	99.60	96.89	99.65	1.00
15	0.00	0.00	99.96	0.50
16	97.60	96.06	99.56	0.99
Average	**67.93**	**66.99**	**98.60**	**0.83**

Using only selected features with GA, the classification performance of the constructed two stage system is %91.97 for overall accuracy, and the other performance measures are shown in Table 5.29.

Table 5.29 Performance measures of the constructed two stage
system for selected features with GA

Labels	SEN	SEL	SPE	ROC Area
1	97.20	100.00	100.00	0.99
2	96.80	100.00	100.00	0.98
3	91.60	100.00	100.00	0.96
4	99.12	95.73	99.79	0.99
5	37.50	100.00	100.00	0.69
6	43.24	91.43	99.88	0.72
7	100.00	100.00	100.00	1.00
8	95.20	91.89	99.08	0.97
9	85.37	97.22	99.96	0.93
10	86.86	98.56	99.87	0.93
11	88.68	100.00	100.00	0.94
12	86.00	94.30	99.43	0.93
13	92.00	95.44	99.52	0.96
14	100.00	98.43	99.83	1.00
15	37.50	8.70	97.51	0.68
16	100.00	75.76	96.51	0.98
Average	**83.57**	**90.47**	**99.46**	**0.92**

The third stage is added in the system to increase the performance of the multi-stage classification system. Using raw data as feature set and KNN classifier for the third stage, the classification performance of the proposed three stage system is %93.59 for overall accuracy, and the other performance measures are shown in Table 5.30.

Table 5.30 Performance measures of the constructed three stage system using selected features

Labels	SEN	SEL	SPE	ROC Area
1	97.20	100.0	100.0	0.99
2	98.00	99.19	99.91	0.99
3	95.20	100.0	100.0	0.98
4	99.12	91.06	99.55	0.99
5	50.00	100.0	100.0	0.75
6	48.65	90.00	99.84	0.74
7	100.0	100.0	100.0	1.00
8	95.60	91.22	99.00	0.97
9	97.56	97.56	99.96	0.99
10	88.98	97.22	99.74	0.94
11	90.57	100.0	100.0	0.95
12	90.80	93.03	99.26	0.95
13	92.80	93.93	99.35	0.96
14	100.0	98.43	99.83	1.00
15	25.00	100.0	100.0	0.63
16	100.0	75.76	96.51	0.98
Average	**85.59**	**95.46**	**99.56**	**0.93**

The proposed system can work as a real time recognition system since it is beat based. It means that system requires features which are obtained only from the segmented beat and all ECG record is not necessary. This may be a limitation, because time domain measures such as R-R intervals, which are significant discriminative features for some beat types (ventricular beats, supraventricular beats), are not used in this system. Time domain measures, 30 R-R intervals (15 previous and 15 next), are prepared from the ECG signals and used as additional features. However, in this approach, 30 second record time is needed to start the recognition algorithm. When feature selection steps are restarted with new feature set, which includes time domain measures, the overall accuracy of the three stage classification system becomes 97.76% and other performance measures are given in Table 5.31.

Table 5.31 Performance measures of the three stage system
using new feature set (including time domain measures)

Labels	SEN	SEL	SPE	ROC Area
1	100.0	100.0	100.0	1.00
2	100.0	99.60	99.96	1.00
3	98.40	100.0	100.0	0.99
4	97.35	93.22	99.67	0.99
5	100.0	100.0	100.0	1.00
6	74.32	100.0	100.0	0.87
7	100.0	100.0	100.0	1.00
8	99.60	93.61	99.26	0.99
9	100.0	93.18	99.88	1.00
10	100.0	98.33	99.83	1.00
11	92.45	100.0	100.0	0.96
12	99.20	94.66	99.39	0.99
13	95.60	97.55	99.74	0.98
14	100.0	99.60	99.96	1.00
15	18.75	100.0	100.0	0.59
16	100.0	98.81	99.87	1.00
Average	**92.23**	**98.04**	**99.85**	**0.96**

Comparison of proposed system with other systems reported in the literature is really difficult because of the varieties in the classification techniques and data properties (e.g. different number of beat types belonging to different patients). But still, results show that the proposed novel system based on two stage combined KNN classifier for discriminating a broad range of heartbeats performs very good with average sensitivity, average selectivity, average specificity, average ROC area and overall accuracy of 85.59%, 95.46%, 99.56%, 0.93, and 93.59% respectively.

Hosseini, Reynolds, & Powers, (2001) used three stage ANN based system to classify six arrhythmias using QRS area, PR, QT, RS interval, ST segment area, QRS energy, ST slope, ST level, Autocorrelation coefficient, 52 sample QRS window that resample as 4:1. In first stage of system, using MLP and classify three classes and others (used in second stage), second stage: using MLP and classify three classes and

others (used in third stage) and third stage: using SOM for increase performance and classify the unknown beats which come from stage 2. Overall recognition rate is 0.883.

Nasiri et al. (2009) have presented a recognition system to discriminate four beat types. Only morphological measures of the QRS complexes are extracted as features. They applied all features, principle components and selected features by genetic-SVM method to classify heartbeats. Genetic algorithm with SVM method is presented as feature selection. The overall accuracy of 93.46% has reported using SVM classifier with selected features by GA.

Osowski & Linh, (2001) used two stage system to classify seven arrhythmias using 60 sample window, cumulants (second, third, and fourth), RR interval, QRS width, average RR interval (last 10). Fuzzy self organizing layer was used for pre-classifier, and MLP was used for final classifier in two stage classification system. The output of Fuzzy NN was used as the input of MLP and output of MLP is 7 that is class size. The efficiency of the system is 96.06%, misclassification rate of the system is 3.94%.

Chazal, O'Dwyer, & Reilly, (2004) used a linear discriminant based classification system. Morphology (QRS duration in time, T wave duration in time, P wave flag, QRS complex (10 sample), T-wave segment (9 sample), RR interval: Pre RR, Post RR interval, average RR interval, Local average RR interval (10 sample), Normalized (std), and combination of all were used as feature sets. They reported sensitivity of 75.9%, positive prediction of 38.5%, false prediction of 4.7% for performance of supraventricular beats and sensitivity of 77.7%, and positive prediction of 81.9%, false prediction of 1.2% for performance of ventricular beats. The system has been constructed to classify five classes (normal beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q)) according to AAMI. Average sensitivity, selectivity and specificity have been calculated from Table V in the work of Chazal et al. The measures were 73%, 45%, and 96% for sensitivity, selectivity and specificity, respectively.

Song et al, (2005) used ANN, Fuzzy Inference System, Support Vector Machine to classify six classes using pre RR and post RR intervals, 144 sample window, WT coefficient (4, 5, 6, 7 order). Using LDA, PCA features was reduced 4 features. The average performance of classifiers are sensitivity 98.9% , specificity %93.7 accuracy %98.1 for original feature, sensitivity 98.2%, specificity 92.5%, accuracy 97.5% for PCA feature, sensitivity %99.6, specificity 95.1%, accuracy 98.9% for LDA feature. According to feature set LDA features have best accuracy. Using LDA feature average sensitivity 92.2%, specificity %98.1, accuracy 98.6% for MLP, sensitivity 89.6%, specificity 98.0%, accuracy %98.9 for FIS, sensitivity %92.3, specificity 99.1%, accuracy 99.3% for SVM.

Tsipourasa, Fotiadisa, & Sideris (2005) used decision tree and simple threshold technique to classify four class arrhythmias using RR intervals, Pre RR, Current RR, and Post RR intervals. Sensitivity, positive prediction and accuracy are 94.9%, 96.1% and 98.2%, respectively.

Alliche & Mokrani (2003) used distance measurement method to classify three arrhythmias using higher order statistic (HOS). The misclassification rate is 5.5%.

Acır (2005) used MLP and least Square SVM system to classify six class of arrhythmias   using Raw data, DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), DWT (Discrete Wavelet Transform), AAR (15th order). Using Dynamic programming feature redundant and for each group 15 features was taken for each set. Sensitivity, specificity and accuracy are 97.6%, 93.8%, and 95.2%, respectively, for LS-SVM, 94.4%, 89.2%, and 91.2%, respectively, for MLP.

The expert system methods are presented by Hu, Palreddy, & Tompkins (1997). They applied Global Expert (GE), Local Expert (LE), and the mixture of expert (MOE) for classification into four main groups (normal beats (N), ventricular ectopic beats (V), fusion beats (F), and unknown beats (Q)). The average classification rate obtained by MOE was 94% and the average error obtained by LE was 4.1%.

Besrour, Lachiri, & Ellouze (2008) applied two ECG feature sets reported sensitivity of 98.3% and specificity of 94.4% for 12-rhythm types, applying HOS (2, 3, 4 cumulant obtained and used only five point different lag(15,30,45,60,75)) and Morphological features (min max of QRS, QRS width, area of QRS, slope) and RR interval (RR interval between a given heartbeat and the previous heartbeat and post RR interval between a given heartbeat and the following heartbeat) with SVM classifier.

Lagerholm, Peterson, & Braccini (2000) presented a recognition system using 16 beat types existed in the MIT-BIH database. Hermite function representation of the QRS complexes and self-organized clustering method are presented. It is reported a total classification accuracy of 98.5%. Other performance measures of 16 heartbeat types from the confusion matrix in Table VI in the work of (Lagerholm, Peterson, & Braccini, 2000) have been calculated as average sensitivity of 67.3%, average selectivity of 81.8%, and average specificity of 99.8%.

Chazal & Reilly (2006) constructed system to classify five classes (normal beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q)). It is reported that sensitivity of 94.34%, selectivity of 94.30%, and specificity of 98.63% for only performance of ventricular beats. But the constructed system was used to classify five classes (normal beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q)). Average sensitivity, selectivity and specificity have been calculated from Table V in the work of Chazal & Reilly (2006). The measures were 70%, 53%, and 97% for average sensitivity, average selectivity and average specificity, respectively.

Melgani & Yakoub (2008) presented a particle swarm optimization and SVM based systems for six types of ECG beats (normal beat, atrial premature beat, ventricular premature beat, right bundle branch block, left bundle branch block, and paced beat). It is reported a total classification accuracy of 91.67%.

Chia-Hung Lin et al. (2009) used probabilistic neural network to classify five classes using fractal properties as features. They reported 96%-97% of overall accuracy.

Arif, Akram, & Afsar (2009) reported that overall recognition rate is 97.33% using a Fuzzy K-NN based system to classify six arrhythmias using statistical properties of wavelet transform and R-R interval.

These results show that the classifier designed in this study has very good performance with an average sensitivity of 83.57%, average selectivity of 90.46%, average specificity of 99.46%, and ROC area of 0.92 for 16 heartbeat types. However, it should be noted that, in each study in the literature, different number of beat types belonging to different patients have been recognized, thus, it is really difficult to compare the results in a fair and objective way.

CHAPTER SIX
CONCLUSIONS

Automatic pattern recognition systems for EEG and ECG patterns have been proposed in this thesis. Pattern recognition techniques, ECG signals, EEG signals, performance measures and combined classifier models have been introduced for pattern classification.

The first phase of the study focuses on the automatic spike detection from EEG records. Interictal spike detection plays a crucial role in the diagnosis of epilepsy. Epilepsy is a very common neurological disorder leading to disturbing seizures. In the interictal period; i.e., in between seizures, epileptic transients, in the form of spikes and sharp waves, are typically observed in the EEG recordings.

True spike waves are aimed to be detected from EEG records by different nonlinear classifiers constructed using single MLP structures utilizing different number of hidden neurons. For training the MLP networks, early stopping versions of backpropagation, backpropagation with adaptive learning rate, Levenberg-Marquardt algorithms and regularization methods are used. In this study, two approaches, the parameterized and raw forms of data, are used as input in MLP-based detection systems. In the first approach, the raw EEG data is directly used as the input to the MLP. In the second approach, some features are extracted from the EEG records and fed into the MLP as the input. In addition, different linear and nonlinear transformations are applied to the available data set in order to obtain different representations of them. The best classifier designed in this study has better performance (with an average sensitivity of 94.1%, an average selectivity of 87.5%, and an average specificity of 95.8%) than all studies in literature. The second system is constructed as a multi-stage classification system. In this approach two classifiers are used in the first stage and one classifier used in the second stage. The definite

spike and definite non-spikes are determined in the first stage and possible spikes and possible non-spikes are used in the second stage. The best performance measures of constructed multi-stage classification system are sensitivity of 94.8%, selectivity of 92.5%, specificity of 96.3%, and overall accuracy of 95.8%. Due to high accuracy rates, the proposed systems could be used to help the physicians in diagnosing epileptic activity in clinical environments.

The second part of this study aims to construct a robust arrhythmia classification system. The detection of different types of arrhythmias from an electrocardiogram has been a very important subject. This is due to the fact that the accurate recognition and classification of various types of arrhythmias are essential for the correct treatment of the patient. This study introduces an automatic classification system based on a diverse set of features for the automatic detection of 16 heartbeat classes. Nearest neighbor based multi-stage classification system is constructed in this respect.

The studies in literature show that there is no unique robust feature set which could successfully discriminate all 16 arrhythmic type beats from each other. Therefore, diverse range of features should be examined to represent each beat in the classification system to achieve a robust performance. A beat-based feature selection is performed to determine the most discriminative feature set for each beat type. The genetic algorithm is used for finding the optimal or near-optimal combination of features for discrimination. Experimental results indicate that the feature selection step not only eliminates a large number of redundant features but also helps to avoid the curse of dimensionality problem.

Most of the similar work reported in the literature seems to deal with the classification of only a few beat types. A major reason for this situation is the lack of available data for some arrhythmic beats so that it is almost impossible to train a classifier such as neural networks for these beat types. Another reason is related to the fact that no unique robust feature set has yet to be found to successfully discriminate all beat types. Regarding these facts, the proposed system shows very

satisfactory performance in discriminating 16 different beat types. The success of the proposed system is due to its beat-based feature selection from a diverse set of various features such as higher order statistics, Fourier transform components, the higher order statistics of wavelet packet decomposition, and morphological features. In the proposed system, the nearest neighbor rule is used as the classifier. There are two reasons for using this algorithm. Wrapper type feature selection algorithms require extensive time during the selection process. Having a classification algorithm, which does not demand a training step such as nearest neighbor algorithm, reduces the processing time in this step. The other reason for preferring nearest neighbor rule as classifier is the fact that it is provides robust results compared to other algorithms that have more parameters and training step.

In conclusion, the proposed method can be put into practice easily in any computer-based monitoring system. The proposed system is consists of three classification stages. At the first stage, heartbeats are classified into 5 main groups using optimal feature sets for each main group. Then at the second stage, main groups are classified into subgroups using optimal features for each subgroup. A diverse set of features including higher order statistics, morphological features, Fourier transform coefficients, higher order statistics of the wavelet package coefficients are extracted for each different type of ECG beat. At the first stage optimal features for main groups are determined by using a wrapper type feature selection algorithm. Then at the second stage optimal features are similarly selected for discriminating each subgroup of the main groups. In both stages, the classifiers are based on the nearest neighbor algorithm. Then at the third stage unclassified beats from both first and second stages are classified into 16 heartbeats using raw ECG data. The results show that the proposed novel system based on multi-stage classifier for discriminating a broad range of heartbeats performs very good with average sensitivity, average selectivity, average specificity, average ROC area and overall accuracy of 85.59%, 95.46%, 99.56%, 0.93, and 93.59% respectively when it is compared with the studies in literature.

Obviously, the performance of the system depends on the classifiers since feature selection steps is based on wrapper approach. In the proposed system, KNN and MLP classifiers are used. Furthermore, other feature extraction techniques which are not used in this thesis, and other classifiers such as support vector machines, radial basis function networks, could be explored for improving the performance of automated ECG beat classifiers as a future work.

REFERENCES

Acharya, R. U., Suri, J. S., Spaan, J.A.E., & Krishnan, S.M. (2007). *Advances in Cardiac Signal Processing.* Springer

Acır, N. (2004). *Automatic pattern recognition in EEG by using artificial neural networks*. Dokuz Eylül University, The Graduate School of Natural and Applied Sciences, Ph.D. Thesis.

Acır, N. (2005). Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm", *Neural Computing and Applications,* 14(4), 299–309.

Acır, N., & Güzeliş, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Computers in Biology and Medicine*, 34, 561–575.

Acır, N., Öztura, I., Kuntalp, M., Baklan, B., & Güzeliş, C. (2005).Automatic Detection of Epileptiform Events in EEG by a Three-Stage Procedure Based on Artificial Neural Networks, *IEEE Transactions On Biomedical Engineering,* 52(1),30-40.

Activation sequence, (2009). Activation sequence of sinus rhythm, Retrieved July 3, 2009, from http://www.bioen.utah.edu/faculty/sri/Lab8_humancardiovascular.htm

Adjouadi, M., Sanchez, D., Cabrerizo, M., Ayala, M., Jayakar, P., Yaylali, İ., et al. (2004). Interictal Spike Detection Using the Walsh Transform*, IEEE Transactions On Biomedical Engineering,* vol. 51, no. 5.

Afonso, V.X., & Tompkins, W.J. (1995). Detecting ventricular fibrillation. *IEEE Eng Med Biol*., 14, 152-159.

Alberg, A.J., Park, J. W., Hager, B. W., BA, Malcolm V Brock, MD, & Marie Diener-West, (2004). The Use of "Overall Accuracy" to Evaluate the Validity of Screening or Diagnostic Tests, *J Gen Intern Med*., pp.460–465.

Al-Fahoum, A.S., & Howitt, I. (1999). Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias. *Med Biol Eng Comp.,* 37, 566-573.

Alliche, A., Mokrani, K. (2003). *Higher order statistics and ECG arrhythmia classification,;Signal Processing and Information Technology*, 2003. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on :641-643.

Alpert, M.A. (1980). *Cardiac arrhythmias*. Chicago: Year Book Medical.

Amari, S. (1995).Training Error, Generalization Error and Learning Curves in Neural Learning, *Artificial Neural Networks and Expert Systems, Proceedings., Second New Zealand International Two-Stream Conference*, 4–5.

Arian R. van Erkel, & Peter M. Th. Pattynama, (1998). Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, *European Journal of Radiology,* Volume 27, Issue 2, Pages 88-94.

Arif, M., Akram, M.U., & Afsar, F.A., (2009). *Arrhythmia Beat Classification Using Pruned Fuzzy K-Nearest Neighbor Classifier Soft Computing and Pattern Recognition*. SOCPAR '09. International Conference, p37-42 .

Besrour, R., Lachiri, Z., & Ellouze, N. (2008) ECG Beat Classi1fier Using Support Vector Machine. *3rd International Conference on Information and Communication Technologies: From Theory to Applications*, ICTTA 2008, 1-5.

Bhaskar, H., Hoyle, D.C., & Singh, S. (2006). Machine learning in bioinformatics: A

brief survey and recommendations for practitioners. *Computer in Biology and Medicine*, 36, 1104-1125,

Bishop, C.M. (1995).*Neural Networks For Pattern Recognition*, NY, Oxford Unv. Press.

Bottoni, P., Cigada, M., De Giuli, A., Di Cristofaro, B., & Mussio, P. (1990). Shape Description as a Key to ECG Analysis. *IEEE Conference: Computers in Cardiology, IEEE Computer Society Press*, Los Alamitos, CA, 443-6

Braccini, G., Edenbrandt, L., Lagerholm, M., Peterson, C., Rauer, O., Rittner, R. et al. (1997). Self-organizing maps and Hermite functions for classification of ECG complexes. *Comp Cardiol*, 425-428.

Chan, Z.S.H., Ngan, H.W., Rad, A.B., & Ho, T.K. (2002). Regularization Alleviation Overfitting Via Genetically Regularized Neural Network, *Electronics Letters*, 38(15), 809-810.

Chatrian, E., Bergamini, L., Dondey, M., Klas, D.W., Lennox-Buchthal, M., & Petersen, I. (1974). A glossary of terms most commonly used by clinical electroencephalographs, *Electroencephalogr. Clinical Neurophysiology*, 37, 538-548.

Chazal, P. de, O'Dwyer, M., & Reilly, R. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, 1196–1206.

Chazal, P. de, & Reilly, R. (2006). A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features, *IEEE Transactions On Biomedical Engineering*, 2535-2543.

Chen, S.W., Clarkson, P.M., & Fan, Q. (1996). A robust sequential detection algorithm for cardiac arrhythmia classification. *IEEE Trans Biomed Eng,* 43, 1120-1125.

Chen, T.C., Han, D.J., Au, F.T.K., & Tham, L.G. (2003). Acceleration of Levenberg-Marquardt Training of Neural Networks with Variable Decay Rate, *Neural Networks, Proceedings of the International Joint Conference,* 3, 1873-1878.

Chia-Hung Lin, Chao-Lin Kuo, Jian-Liung Chen, & Wei-Der Chang, (2009). Fractal features for cardiac arrhythmias recognition using neural network based classifier, *Networking, Sensing and Control, 2009. ICNSC '09. International Conference,* pp:930 – 935.

Clayton, R.H., Murray, A. & Campbell, R.W.F. (1993). Comparison of four techniques for recognition of ventricular fibrillation of the surface ECG. *Med. Biol. Eng. Comp.,* 31, 111-117.

Clayton, R.H., Murray, A., & Campbell, R.W.F. (1994). Recognition of ventricular fibrillation using neural networks. *Med Biol Eng Comp,*32,217-220.

Cover, T. M. & Hart, P. E. (1968). Nearest neighbor pattern classification, *IEEE Trans. Inform. Theory*, 21–27.

Crawford, M.H. (2004) *Crawford kardiyoloji,* London, Mosby press.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory, 36(5), 961–1005.

Demuth, H., & Beale, M. (1998). *Neural Network Toolbox for use with MATLAB*, The MathWorks, Inc.

Dingle, A.A., Jones, R.D., Carroll, G.J., & Fright, W.R. (1993). A multistage system to detect epileptiform activity in the EEG, *IEEE Trans. Biomed. Eng.,* 40(12), 1260-1268.

Dokur, Z., & Olmez, T. (2001). ECG beat classification by a hybrid neural network. *Comp Meth Prog Biomed,* 66, 167- 181.

Duda, R.O., Hart, P.E., & Stock, D.G. (2001). *Pattern Classification*, John Wiley&Sons.

Eberhart, R.C. & Dobbins, R.W. (1990). *Neural network PC tools*. San Diego: Academic Press.

Erkel, A.R. van, & Pattynama, P. M. Th. (1998). Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, *European Journal of Radiology,* 27(2), 88-94.

Exarchos, T.P., Tzallas, A.T., & Fotiadis, D.I. (2006). EEG Transient Event Detection and Classification Using Association Rules. *IEEE Transactions on Information Technology in Biomedicine,* vol. 10, no 3.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.

Feng, D. (2007). *Biomedical Information Technology*. Elsevier.

Fisch, BJ. (1991). *Spehlmann's EEG primer*. Amsterdam: Elsevier Publication.

Gabor, A.J., & Seyal, M. (1992). Automated inter-ictal EEG spike detection using artificial neural networks, *Electroenceph. Clinical Neurophysiology*, 83, 271-280.

Gertsch, M. (2003). *The ECG A Two Step Approach to Diagnosis*. Springer.

Gibbons, R.J., Balady, G.J., Beasley, J.W., Bricker, J.T., Duvernoy, W.F., Froelicher, V.F., et al. (1997). ACC/AHA guidelines for exercise testing, a report of the American College of Cardiology/ American Heart Association Task Force on Practice guidelines (committee on exercise testing). *Journal of the American College of Cardiology*, 30(1), 260-311.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning,* Kluwer Academic Publishers, Boston, MA,

Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P. Ch, Mark, R.G. et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220   Circulation   Electronic   Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215.

Göksan, B. (1998). Epilepside Tani Yöntemleri. *Epilepsilerde Tani ve Tedavi Sempozyumu*, Istanbul, 39-50.

Haddad, S.A.P., & Serdijn, W. A. (2009). *Ultra Low-Power Biomedical Signal Processing*, Springer.

Hagan, M. T., & Menhaj, M. B. (1994). Training feed forward networks with the Marquardt algorithm. *IEEE Trans. On Neural Net.*, 6. 989-993.

Hagiwara, K., & Kuno, K. (2000). Regularization Learning And Early Stopping In Linear Networks. *Neural Networks, IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference*, 4, 511-516. (Haykin, 1999) Haykin, S. (1999). *Neural Networks*, Prentice Hall.

Ham, F.M., & Han, S. (1996). Classification of cardiac arrhythmias using fuzzy ARTMAP. *IEEE Trans Biomed Eng*, 43, 425-430. (Hecht-Nielsen, 1989) Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network, *Neural Networks. IJCNN., International Joint Conference*, 1, 593-605.

Hamilton, P.S., & Tompkins, W.J. (1986). Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmiac database, *IEEE Trans. Biomed. Eng*. 33, 1157-1165.

Heart Structure, (2009). *Heart Structure*, Retrieved July 3, 2009, from http://en.wikibooks.org/wiki/Wikijunior:Human_Body/Heart

Heidari, H., Shahidi, A.V., Aminian, K., & Sadati, N. (1998). *Analysis of the Sustained Ventricular Arrhythmias from SAECG Using Artificial Neural Network and Fuzzy Clustering Algorithm*. Proceedings of the 20th Annual International Conference of the ZEEE Engineering in Medicine and Biology Society, vol. 20, no 1.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*, first ed., University of Michigan Press, Ann Arbor, MI,

Hosseini, H.G., Reynolds, K.J., & Powers, D. (2001). *A Multi-stage Neural Network Classifier for ECG Events*, Proceedings of the 23rd International Conference of the IEEE Engineering in Medicine and Biology Society, 25-28.

Hu, Y.Z., Palreddy, S., & Tompkins, W.J. (1997). A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Trans Biomed Eng,*44:891-900.

İsler, Y., & Kuntalp, M. (2007). Combining Classical HRV Indices With Wavelet Entropy Measures Improves to Performance in Diagnosing Congestive Heart Failure, *Computers in Biology and Medicine,* vol.37(10), 1502-1510.

İnan, Z.H., & Kuntalp, M. (2007). A study on fuzzy C-means clustering-based systems in automatic spike detection. *Computers in Biology and Medicine* 37, 1160 – 1166.

Ingle, V.K. & Proakis, J.G. (2000). *Digital Signal Process,ng Using Matlab*. Pacific Grove, USA: Brooks/Cole Publishing Company.

Jain, A.K., Duin, R.P.W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1), 4-37.

James, C.J., Jones, R.D., Bones, P.J., & Carroll, G.J. (1999). Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages, *Clinical Neurophysiology*, 110, 2049–2063.

John, G. Kohavi, R. & Pfleger, K. (1994). Irrelevant features and the subset selection problem. *In Machine Learning: Proceedings of the Eleventh International Conference*.

Kalayci, T., & Özdamar, Ö. (1995). Wavelet Processing For Automated Neural Network Detection of EEG Spikes, *IEEE Eng. Med. Biol. Mag*., 13, 160-166.

Karimifard, S., Ahmadian, A., & Khoshnevisan, M. (2006). Morphological heart arrhythmia detection using hermitian basis functions and kNN classifier. *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vols 1-15, 4489-4492.

Khadra, L., Al-Fahoum, A.S., & Al-Nashash, H. (1997). Detection of life-threatening cardiac arrhythmias using wavelet transformation. *Med Biol Eng Comp*, 35, 626-632.

Khadra, L., Al-Fahoum, A. & Binajjaj, S. (2005). A Quantitative Analysis Approach for Cardiac Arrhythmia Classi- fication Using Higher Order Spectral Techniques. *IEEE Transactions on Biomedical Engineering.*

Kiloh, L.G., McComas, A.J., Osselton, J.W., & Upton, A. (1981). *Clinical electroencephalography (4th ed).* London UK: Butterworths.

Ko, C.W., & Chung, H.W. (2000). Automatic spike detection via an artificial neural network using raw EEG data: effect of data preparatio n and implications in the limitations of online recognition, *Clinical Neurophysiology*, 111, 477-481.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence,* 97, 273-324

Kohonen, T. (1982). Self-Organized formation of topologically correct feature maps", *Biological Cybernetics*, 43, 59-69.

Kohonen, T. (2001). *Self-Organizing Maps*, 3rd edn., Springer.

Ktonas, P.Y. (1987). *Methods of analysis of brain electrical and magnetic signals*, EEG, Elsevier; Amsterdam: Handbook.

Kutlu, Y., İşler, Y., Kuntalp, D., & Kuntalp, M. (2006). *Çok katmanlı Ağ Yapıları Kullanılarak Diken Dalgaların Algılanması,* IEEE 14th Signal Processing and Communications Applications Conference (SIU2006), Antalya / Turkey, 17-19 April 2006.

Kutlu, Y., Kuntalp, D., & Kuntalp, M. (2006). *Gürültü Eklenmiş EEG Verisi Kullanılarak Çok Katmanlı Ağ Yapılarıyla Diken Dalgaların Ayıklanması,* Biyomedikal Mühendisligi Ulusal Toplantisi, (BIYOMUT2006), İstanbul / Turkey, 125-28 Mayıs 2006

Kutlu, Y., Kuntalp, M. & Kuntalp, D. (2007). *Çok katmanlı Ağ Yapıları Kullanılarak Aritmi Algılanmasında Pencere Boyutu etkisi,* IEEE 15th Signal Processing and Communications Applications Conference (SIU2007), Eskisehir / Turkey, 11-13 June 2007

Kutlu, Y., Kuntalp, M. & Kuntalp, D. (2008a). *Öz Düzenleyici Haritalar Kullanilarak Diken Dalgalarin Analizi,* Genç Bilim İnsanları ile Beyin Biyofiziği II. Çalıştayı, Izmir / Turkey, 21-23 Feb 2008

Kutlu, Y., Kuntalp, M. & Kuntalp, D. (2008b). *Yüksek Mertebeden Istatistik kullanilarak Aritmi Siniflandirilmasi*, IEEE 16th Signal Processing and Communications Applications Conference (SIU2008), Didim / Aydin / Turkey, 20-22 April 2008,

Kutlu, Y., Kuntalp, M. & Kuntalp, D. (2008c). *Öz Düzenleyici Haritalar Kullanılarak Aritmik Ekg Vurularının Topografik Analizi,* XIII.Biyomedikal Mühendisliği Ulusal toplantısı BIYOMUT 2008, Ankara / Turkey, 29-31 May 2008

Kutlu, Y. & Kuntalp, D. (2009a). *Feature Reduction Method Using Self Organizing Maps*, 6th International Conference on Electrical and Electronics Engineering,5-8 November 2009, Bursa, TURKEY.

Kutlu, Y. & Kuntalp, D. (2009b). Feature Extraction for ECG Heartbeats Using Higher Order Statistics of Wavelet Packet Decomposition Coefficients, *Pattern recognition*, under review, since june 6, 2009.

Kutlu, Y., Kuntalp, M., & Kuntalp, D. (2009a). *EKG Vurularının Morfolojik Özniteliklerin Topografik Analizi,* XIV. Biyomedikal Mühendisliği Ulusal toplantısı BIYOMUT 2009, İzmir / Turkey, 20-23 May 2009,

Kutlu, Y., Kuntalp, M., & Kuntalp, D. (2009b). Optimizing the performance of an MLP classifier for the automatic detection of epileptic spikes, *Expert Systems with Applications,* Volume 36, Issue 4, May 2009, Pages 7567-7575

Kutlu, Y., Kuntalp, M., & Kuntalp, D. (2009c). *The Applied Brain Biophysics, chapter: Visualization of SOM as a U-matrix*, pages 309-319, 2009, Dokuz Eylül Üniversitesi Tıp Fakültesi Yayınları, ISBN: 978-975-441-259-8

Kutlu, Y. & Kuntalp, D. (2010). Multi-Stage Automatic Heartbeat Recognition, ready to submition.

Lagerholm, M., Peterson, C., & Braccini, G. (2000). Ebendrandt L, Sornmo L. Clustering ECG complexes using hermite functions and self-organizing maps. *IEEE Trans Biomed Eng.,*47, 838-848.

Learned, R.E., & Willsky, A.S. (1995). A wavelet packet approach to transient signal classification. *Applied and Computational Harmonic Analysis*, 2. 265–278.

Livarinen, J., Kohonen, T., Kangas, J., & Kaski, S. (1994). Visualizing the clusters on the self-organizing map. *Proceedings of the Conference on Artificial Intelligence Research in Finland*, 122-126.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.,* 11, 674-693.

Maroño, N. S., Betanzos, A.A., & Sanromán, M. T. (2007). Filter methods for selection − A comparative study. *Intelligent Data Engineering and Automated*

Learning - IDEAL 2007, 178-187.

Melgani, F, & Yakoub, B. (2008). Classification of ECG signals with Support Vector Machine and Particle Swarm Optimization. 3rd . *IEEE Transactions on Information Technologyin Biomedicine,* 12, 667-677.

Melo, S.L., Caloba, L.P. & Nadal, J. (2000). Arrhythmia analysis using artificial neural network and decimated electrocardiographic data. *Comp Cardiol*, 27, 73-76.

Mendel, J.M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications, *Proceedings of the IEEE,* 79(3), 278 - 305

Meyer, Y. (1993). *Wavelet algorithm and application*. Philadephia,PN: SIAM,

Minami, K., Nakajima, H. & Toyoshima, T. (1997). Arrhythmia diagnosis with discrimination of rhythm origin and measurement of heart-rate variation. *Comp Cardiol, 243–246.*

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Trans Biomed Eng,* 46,179-185.

Minnix, J.I. (1991). An analysis of the effects of the noisy training sets on the fault tolerance of neural networks, Decision Aiding for Complex Systems, Conference Proceedings., *IEEE International Conference on*, 2, 713-718.

Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. M. (2004). *Wavelet toolbox for use with Matlab, User's Guide,* Ver. 3.

Moazzen, I., Ahmadzadeh, M.R., Doost-Hoseini, A.M., & Omidi, M.J., (2009). An intelligent classifier for cardiac arrhythmias recognition, *Wireless Communications & Signal Processing, 2009. WCSP 2009. International Conference,* pp 1-5.

Nadal, J. & Bossan, M. (1993). Classification of cardiac arrhythmia based on principal components analysis and feedforward neural Networks", *Comput Cardiol*, , 341-344.

Nasiri, J.A., Naghibzadeh, M., Yazdi, H.S., Naghibzadeh, B., (2009). ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm, *Computer Modeling and Simulation, 2009. EMS '09. Third UKSim European Symposium,* Page 187 – 192.

Nicholson, A. (2002). *Generalization Error Estimates and Training Data Valuation*, thesis, California Institute of Technology.

Niedermeyer, E., & Silva, F.L.D. (1993). *Electroencephalography, basic principles, clinical applications and related fields,* (3rd ed), Baltimore, MD: Williams & Wilkins.

Nuh, M., Jazidie, A., & Muslim, M.A. (2002). Automatic detection of epileptic spikes based on wavelet neural network, *Circuits and Systems, APCCAS '02. Asia-Pacific Conference on,* 483 - 486.

Oja, M., Nikkila, J., Toronen, P., Wong, G., Castr´en, E., & Kaski, S. (2002). *Exploratory clustering of gene expression profiles of mutated yeast strains*. In W. Zhang & I. Shmulevich (Eds.), Computational and Statistical Approaches to Genomics, Kluwer Academic Publishers.

Osowski, S., & Linh, T.H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Trans Biomed Eng*;48:1265-71.

Osowski, S., Hoai, L.T. & Markiewicz, T. (2004). Support Vector Machine-Based Expert System for Reliable Heartbeat Reliable Heartbeat Recognition", *IEEE Trans. on BME*, 51(4).

Osowski, S., Siroic, R. & Siwek, K., (2009). *Genetic algorithm for integration of ensemble of classifiers in arrhythmia* recognition, Instrumentation and Measurement Technology Conference, 2009. I2MTC '09. IEEE, Page(s): 1496 – 1500.

Owis, M.I., Abou-Zied, A.H., Youssef, A.M., & Kadah, Y.M. (2002). Study of features based on nonlinear dynamical modelling in ECG arrhythmia detection and classification. *IEEE Trans Biomed Eng*, 49, 733-736.

Özdamar, Ö., & Kalayci, T. (1998). Detection of spikes with artificial neural networks using raw EEG, *Computers and Biomedical Research*, 31, 122–142.

Özdamar, Ö., Yaylali, I., Jayaker, P., & Lopez, C.N. (1991). Multilevel neural network system for EEG spike detection, *Computer-Based Medical Systems, Proceedings of the Fourth Annual IEEE Symposium,* 272–279.

Pan, J., & Tompkins, W.J. (1985). A Real-Time QRS Detection Algorithm, *IEEE Trans. Biomed. Eng.*, 32(3), 230-236.

Pang, C.C., Upton, A.R.M., Shine, G., & Kamath, M.V. (2003). A comparison of algorithms for detection of spikes in the electroencephalogram, *IEEE Transactions Biomedical Engineering*, 50(4), 521-526.

Pudil, P., Novovicova, J., & Kittler, J. (1994). Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15, 1119- 1125.

Raghav, S., & Mishra, A.K., (2008). Fractal feature based ECG arrhythmia classification, *TENCON 2008 IEEE Region 10 Conference*, p 1 − 5.

Rangaraj, M.R. (2002). *Biomedical signal analysis*. IEEE Press.

Sawhney, G.S. (*2007*). *Fundamentals of biomedical engineering.* New Age Publishers.

Somol, P., Pudil, P., Novovicova, J., & Paclik, P. (1999). Adaptive Floating search methods in feature selection. *Pattern Recognition Letters,* 20, 1157-1163,

Song, M.H., Lee, J., Cho, S.P., Lee, K.J., & Yoo, S.K. (2005). Support Vector Machine Based Arrhythmia Classification Using Reduced Features. *International Journal of Control, Automation, and Systems*, 3(4), 571–579.

Tang, E.K., Suganthan, P.N., Yao, X., & Qin, A.K. (2005). Linear dimensionality reduction using relevance weighted LDA *, Pattern Recognition*, 38(4), 485-493.

Tarassenko, L., Khan, Y.U., & Holt, M.R.G. (1998). Identification of inter- ictal spikes in the EEG using neural network analysis, *Inst. Elect. Eng.-Proc. Sci. Meas. Technol.,* 145(6), 270–278.

Thakor, N.V., Natarajan, A.,& Tomaselli, G. (1994). Multiway sequential hypothesis testing for tachyarrhythmia discrimination. *IEEE Trans Biomed Eng,* 41,480-487.

Thakor, N.V., Zhu, Y.S., & Pan, K.Y. (1990). Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Trans Biomed Eng*, 37,837-43.

Tompkins, W. J. (1993). *Biomedical Digital Signal Processing: C-Language Examples .and Laboratory Experiments for the IBM Pc*. Prentice Hal.

Torun, M.U., İşler, Y., Kuntalp, D. & Kuntalp, M. (2006). *Dal Bloğu Vuruların Yüksek Mertebeden İzgel Çözümleme ve Yapay Sinir Ağları ile Sınıflandırılması*. IEEE 14. Sinyal İşleme ve İletişim Uygulamaları Kurultayı.

Tsipourasa, M.G., Fotiadisa, D.I., & Sideris, D. (2005). An arrhythmia classification system based on the RR-interval signal, *Artificial Intelligence in Medicine* 33, 237-250.

Tsukuda, Y., Kurokawa, H., & Mori, S. (1995). Investigation of generalization ability by using noise to enhance MLP performance, *Neural Networks, Proceedings., IEEE International Conference on,* 5, 2795-2798.

Ultsch, A. (2003). Maps for the Visalization of high-dimensional Data Spaces, in Proc. *Workshop on Self organizing Maps*, 225 - 230.

Ultsch, A. (1992). Self-Organizing Neural Networks for Visualization and Classification, *Proc. Conf. Soc. for Information and Classification.*

Ultsch, A. (1993). Self-organizing neural networks for visualization and classification. *Information and Classification, eds. O.Opitz, B.Lausen and R.Klar*, Springer-Verlag, Berlin, 307- 313.

Unser, M., & Aldroubi, A. (1996). A review of wavelets in biomedical applications. *Proceedings of the IEEE,* 84(4), 626–638.

Übeyli, E.D. (2007). ECG beats classification using multiclass support vector machines with error correcting output codes. *Digital Signal Processing*, vol 17, 675-684

Wagner, G.S. (2001). *Marriott'un Pratik Elektrokardiyografisi* (10th ed). Lippincott W&W.

Wang, Y., Zhu, Y.S., Thakor, N.V., & Xu, Y.H. (2001). A short-time multifractal approach for arrhythmia detection based on fuzzy neural network. *IEEE Trans Biomed. Eng.* 48, 989-995.

Watanabe, S. (1985). *Pattern recognition: human and mechanical.* New York: John Wiley & Sons Inc.

Webber, W.R.S., Litt, B., Wilson, K., & Lesser, R. (1994). Practical detection of epileptiform discharges (ED's) in the EEG using an artificial neural network: a comparison of raw and parameterized data, *Electroencephalography and Clinical Neurophysiology*, 91, 194-204.

Webster, J.G. (1993). *Design of Cardiac Pacemakers*. IEEE Press.

Webster, J.G. (1998) *Medical instrumentation application and design (3rd ed)*. John Wiley&sons.

Wiskott, L. (2004). *Principal Component Analysis*, Humboldt-University Berlin.

Yang, T.F., Device, B., & Macfarlane, P.W. (1994). Artificial neural networks for the diagnosis of atrial fibrillation. *Med Biol Eng Comp.*32,615-619.

Yong, Z., Wenxue, H., & Yonghong, X. (2009). *ECG Beats Feature Extraction Based on Geometric Algebra,* Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference*,* pp. 1 − 3.

Yu, C.C., & Liu, B.D. (2002). A backpropagation algorithm with adaptive learning rate and momentum coefficient. *Neural Networks, IJCNN '02. Proceedings of the 2002 International Joint Conference,* 2, 1218-1223.

Yu, X., Efe, M.O., & Kaynak, O. (2002). A general backpropagation algorithm for feedforward Neural Networks learning, *Neural Networks, IEEE Transactions on,* 13(1), 251–254.

Zhang, X.S., Zhu, Y.S., Thakor, N.V., & Wang, Z.Z. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Trans Biomed Eng,* 45,548-555.