DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEVELOPMENT OF A SYSTEM TO DIAGNOSE PAROXYSMAL ATRIAL FIBRILLATION PATIENTS FROM ARRHYTHMIA FREE ECG RECORDS

by
İrem HİLAVİN

April, 2016
İZMİR

DEVELOPMENT OF A SYSTEM TO DIAGNOSE PAROXYSMAL ATRIAL FIBRILLATION PATIENTS FROM ARRHYTHMIA FREE ECG RECORDS

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

by
İrem HİLAVİN

April, 2016
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**DEVELOPMENT OF A SYSTEM TO DIAGNOSE PAROXYSMAL ATRIAL FIBRILLATION PATIENTS FROM ARRHYTHMIA FREE ECG RECORDS**" completed by **İREM HİLAVİN** under supervision of **PROF. DR. MEHMET 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.

Prof. Dr. Mehmet KUNTALP

Supervisor

Assoc. Prof. Dr. Olcay AKAY	Assoc. Prof. Dr. Adil ALPKOÇAK
Thesis Committee Member	Thesis Committee Member
Prof. Dr. Nurettin ACIR	Assoc. Prof. Dr. Türker İNCE
Examining Committee Member	Examining Committee Member

Prof. Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Prof. Dr. Mehmet KUNTALP for his guidance through this thesis work, and for enlightening me with advices and knowledge.

I would like to thank my thesis examining committee members; Assoc. Prof. Dr. Olcay AKAY and Assoc. Prof. Dr. Adil ALPKOÇAK who pointed out missing points and helped me to improve my thesis.

I also would like to thank Turkish Scientific and Technological Research Council for supporting me in the framework of BIDEP 2211 Scholarship Program.

Finally, I thank my parents Mehmet-Asiye Düven, my husband Sezgin and my daughter İpek for their support and patience.

I also want to thank academic and administrative staff of the Graduate School of Natural and Applied Sciences of Dokuz Eylül University, especially to Mrs. Filiz Gürsan for their valuable support.

İrem HİLAVİN

DEVELOPMENT OF A SYSTEM TO DIAGNOSE PAROXYSMAL ATRIAL FIBRILLATION PATIENTS FROM ARRHYTHMIA FREE ECG RECORDS

ABSTRACT

Atrial fibrillation (AF) patients could easily be determined based on their electrocardiogram (ECG) records. However, in paroxysmal atrial fibrillation (PAF) case, AF episodes occur randomly and mostly the cardiac rhythm returns to normal sinus rhythm before the subjects reach a health care facility. This makes it very difficult to obtain ECG records during a PAF attack. Therefore, there is a need for a method that could diagnose PAF based on the ECG recordings taken during normal sinus rhythm.

In this thesis, a system to diagnose PAF patients from their ECG records taken during normal sinus rhythm was proposed. The ECG records of PAF patients were selected at least 45-minute away from any PAF attack to ensure the signal is independent from any AF effect to simulate real clinical case. A combination of time domain, frequency domain and nonlinear heart rate variability (HRV) features were offered to discriminate PAF patients and non-PAF subjects. The most distinguishing features were selected by genetic algorithm among thirty three features. Discriminative ability of the selected features was visualized with self organizing maps. Different classifiers were examined to find out which one separates two groups better in multidimensional feature space. The best sensitivity, specificity and accuracy values, which were obtained with support vector machine classifier, were found to be 93%, 95% and 95%, respectively, using the selected features. Hereby, the proposed system can be used to detect PAF patients from their 5-minute AF free records in clinical use.

Keywords: Paroxysmal atrial fibrillation, classification, heart rate variability analysis, genetic algorithm, support vector machine

ARİTMİSİZ EKG KAYITLARINDAN PAROKSİSMAL ATRİYAL FİBRİLASYON HASTALARINI TEŞHİŞ EDİCİ SİSTEM GELİŞTİRİLMESİ

ÖZ

Atriyal fibrilasyon (AF) hastaları elektrokardiyagram (EKG) kayıtlarına dayanarak kolayca belirlenebilir. Ancak paroksismal atriyal fibrilasyon (PAF) durumunda AF atakları rasgeledir ve genellikle hasta bir sağlık kuruluşuna gitmeden kalp ritmi kendiliğinden normal sinus ritmine geri döner. Bu durum PAF atağı sırasında EKG kaydı almayı çok zorlaştırır. Bu yüzden normal sinus ritminde alınan EKG kayıtlarından PAF hastalığını teşhis edici bir metoda ihtiyaç vardır.

Bu tezde, normal sinus ritminde alınan EKG kayıtlarından PAF hastalarını teşhis edici bir sistem önerilmiştir. Gerçek klinik durumu simüle etmek için PAF hastalarının EKG kayıtları herhangi bir AF atağından en azından 45 dakika uzakta alınarak sinyalin herhangi bir AF etkisinden bağımsız olması sağlanmıştır. PAF hastaları ve PAF olmayan kişileri ayırmak için zaman düzlemi, frekans düzlemi ve lineer olmayan kalp hızı değişikliği özniteliklerinin bir kombinasyonu önerilmiştir. Otuzüç öznitelik arasından en ayırıcı olan öznitelikler genetik algoritma ile seçilmiştir. Seçilen özniteliklerin ayırma yeteneği özdüzenleyici haritalar ile görselleştirilmiştir. Çok boyutlu öznitelik uzayında iki grubu daha başarılı ayıran sınıflandırıcıyı bulmak amacıyla farklı sıınıflandırıcılar incelenmiştir. Vektör destek makinesi sınıflandırıcı ile sırasıyla %93, %95, %95 duyarlılık, özgüllük ve doğruluk sonuçları elde edilmiştir. Böylelikle önerilen system ile klinik uygulamalarda PAF hastaları atriyal fibrilasyonsuz 5 dakikalık kayıtlarından teşhis edilebilir.

Anahtar kelimeler: Paroksismal atriyal fibrilasyon, sınıflandırma, kalp hızı değişkenliği analizi, genetik algoritma, vektör destek makinaları

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Human body is a life-sustaining structure which is composed of 11 body systems; circulatory, digestive, respiratory, urinary, skeletal, muscular, integumentary, immune, nervous, endocrine and reproductive (Sherwood, 2015). Body systems function in cooperation with each other to achieve homeostasis, which is keeping the internal environment of the body in steady state with respect to changing external conditions.

Circulatory system is the transport system of the body to maintain homeostasis. It carries the vital materials such as nutrients, oxygen, carbon dioxide, wastes, and hormones from one part of the body to another part. The circulatory system, in other words the cardiovascular system, consists of the heart, the blood vessels and the blood. The heart produces necessary pressure to the blood for flowing through whole body. The blood vessels are the passageways through which the blood is directed. The blood is the transport medium of dissolved or suspended materials to reach all parts of the body. Any problem in one of these parts strongly affects the proper functioning of other body systems.

The term arrhythmia refers to any change of the normal electrical impulses in the heart. The impulses might be too slow, too fast or erratic which causes the heart to pump blood insufficiently. Atrial fibrillation (AF) is an arrhythmia which is caused by the disordered electrical impulses in the atria. These disordered impulses prevail the normal sinus rhythm and thus atria start to contract in a disordered way. As a result, the atria cannot empty the blood to ventricles completely and clot formation may occur in atria. If this clot joins the circulation, it may lead to stroke. In fact, AF is the leading cause of 15% of all stroke cases (Go, Hylek, Phillips, & Chang, 2001). AF is the most common sustained arrhythmia (Camm et al., 2010; Miyasaka et al., 2006) and recognized as an increasing health-care burden due to the aging population and survival from cardiac disorders. Prevalence of AF is estimated 1-2% in general population (Camm et al., 2010; Stewart, Hart, Hole, & McMurray, 2001) whereas this ratio increases with age from 0.5% at 40-50 years to 5% over 65 years, and to 14% in subjects over 85 years (Majeed, Moser, & Carroll, 2001; Naccarelli, Varker, Lin, & Schulman, 2009). However, studies show that the real prevalence is much higher (Gladstone et al., 2014; Sanna et al., 2014). The reason why the real numbers are unclear is that approximately one third of AF cases is asymptomatic (Furberg et al., 1994; Savelieva & Camm, 2000) and even in symptomatic cases, the symptoms are ascribed to other illnesses.

AF may be self-terminating or non-self-terminating. If the episodes of AF self-terminate within 48 hours, it is called paroxysmal atrial fibrillation (PAF). Even if the definition is 48 hours, AF episodes generally terminate within minutes (Hoshino, Ishizuka, Nagao, Shimizu, & Uchiyama, 2013; Page, Wilkinson, Clair, McCarthy, & Pritchett, 1994). If an episode of AF lasts more than 7 days, this kind of AF is referred to as persistent AF. Electrical or pharmacological cardioversion is necessary to recover the normal rhythm. AF is called permanent if AF exists for some time and return to normal rhythm fails with cardioversion or AF returns within 24h after successful cardioversion (Levy et al., 2003; Lip & Hee, 2001).

Studies show that a large number of PAF patients would develop permanent AF with time (de Vos et al., 2010; Kato, Yamashita, Sagara, Iinuma, & Fu, 2004; Kerr et al., 2005; Van Gelder & Hemels, 2006). Also the risk of stroke in PAF patients is similar to the risk of persistent/permanent AF patients (Friberg, Hammar, & Rosenqvist, 2010; Hart et al., 2000). Therefore, it is crucial that PAF patients are diagnosed early and correctly. If PAF patients are detected in early stages of the illness, the progression to more sustained AF can be avoided. Also, the stroke risk can be reduced considerably with suitable antithrombotic treatment in PAF patients.

However, it is difficult to detect AF patients since there are no specific symptoms assigned to AF. If it is asymptomatic AF, there is no obvious complaint of the patient. The complaints of the symptomatic AF patients are breathlessness/dyspnea, palpitations, syncope/dizziness, chest discomfort and stroke/transient ischemic attack. The first step in diagnosis is to perform manual examination of the rhythm to find if there is any irregular pulse. If an irregular pulse has been detected, an electrocardiogram is performed. If the patient has PAF, it is difficult to detect the arrhythmia with a standard ECG since the episodes are random and there might be even days within paroxysms. William Evans describes the difficulty of diagnosing paroxysmal atrial fibrillation as “A fugitive illness that only visits a patient periodically may be more difficult to bear than one whose effects are durable and persistent. Paroxysmal atrial fibrillation is such an illness.” (Evans, 1959).

In suspected PAF case,

- a 24-hour ambulatory ECG monitoring is used in patients with suspected asymptomatic episodes or symptomatic episodes less than 24 hours apart
- an event recorder ECG is used in patients with symptomatic episodes more than 24 hours apart (Cowan et al., 2014).

Therefore, there is a need for a new and efficient method that could correctly diagnose PAF patients based on the ECG recordings taken during normal sinus rhythm periods. There are many papers published on detection of PAF patients from ECG records. Most of them are the throughputs of Computers in Cardiology Challenge 2001: Predicting the Onset of Paroxysmal Atrial Fibrillation. This competition had two events. The first event was PAF screening, which is to classify the subjects into PAF and non-PAF groups. The second event was PAF prediction, which is identifying the record that immediately precedes PAF, by using the records of detected patients in the first event (Moody, Goldberger, McClennen, & Swiryn, 2001). The subject of this thesis is similar to the first event and a part of the database of this competition was used. However, there are fundamental differences between the competition and this work, which will be mentioned in the next chapters.

In the literature, most of the successful works are based on morphological features obtained from ECG records. Martinez et al. used morphological features of P-waves to discriminate ECG segments of healthy subjects and patients suffering from PAF (Martínez, Alcaraz, & Rieta, 2012, 2014). Ros et al. studied 22 parameters obtained from P-wave analysis to correctly classify PAF patients (Ros, Mota, Fernández, Toro, & Bernier, 2004). Thuraisingham examined wavelet decomposition of ECG signals (Thuraisingham, 2007). Schreier et al. used a correlation based assessment of the P-wave morphology of both regular and premature heartbeats from supraventricular origin (Schreier, Kastner, & Marko, 2001). Zong et al. developed an algorithm based upon the number and timing of the atrial premature complexes in the ECG (Zong, Mukkamala, & Mark, 2001). Maier et al. used different features obtained from heart rate variability analysis describing the magnitude as well as the regularity of heart rate fluctuations and the number of supraventricular and ventricular premature beats (Maier, Bauch, & Dickhaus, 2001). Lynn and Chiang created feature vectors from return and difference maps of 30 minute ECG signals. They divided the maps into lattices and found the number of samples in each lattice. Then they created vectors from those numbers and fed into a k-nearest neighbor classifier (Lynn & Chiang, 2001). Yang and Yin coded successive RR intervals as equal, accelerated or decelerated and mapped to a single integer. Then, the histogram of those numbers was constructed for PAF patients and non-PAF subjects. The best cutoff frequency was found with receiver operator characteristics (ROC) analysis (Yang & Yin, 2001). Chazal and Heneghan examined features from the interval based power spectral density of RR intervals, time domain measures, P-wave amplitude features and frequency representation of the P-wave. The effect of the length of the signal was also controlled by using 30-minute, 10-minute and 5-minute windows of the ECG signals. Their best performance was obtained with power spectral density (de Chazal & Heneghan, 2001).

The main goal of this thesis is to develop a system to diagnose PAF patients from arrhythmia free ECG records. The records belonging to PAF patients were taken at least 45-minutes away from any AF episode because the proposed method in this thesis is based on whether an ECG taken from a PAF patient at a normal time reveals any information for the existence of the illness. Thirty three features were derived from RR series. Thirty one features come from short-time heart rate variability (HRV) analysis and the other two features are atrial and ventricular ectopic beat numbers. The best discriminating features were selected by using genetic algorithms.

PAF and non-PAF patients were discriminated with Bayes’, k-nearest neighbor (kNN), artificial neural network (ANN) and support vector machine (SVM) classifiers. The most successful discrimination was achieved by the SVM classifier when eight features selected by genetic algorithms were used as the input of the classifier.

Chapter 2 of the thesis presents the physiological background of circulatory system and arrhythmias. Preprocessing of cardiovascular time series and derivation of heart rate variability features are given in Chapter 3. Feature selection with genetic algorithm is explained in Chapter 4. The classification and evaluation methods are explained in Chapter 5. Results are presented and discussed in Chapter 6. Finally, a conclusion is given in Chapter 7.

CHAPTER TWO
PHYSIOLOGICAL BACKGROUND

There is a great organization from cell to systems in multicellular organisms. The cell is the smallest unit of life. Cells which have similar structure and function are organized into tissues. Two or more types of primary tissues form organs whose function is to perform a specific function. Different organs function in cooperation which each other to achieve a common activity and form body systems. Finally, the body, which is an independently living individual, is composed of different body systems. Human body consists of circulatory, digestive, respiratory, urinary, skeletal, muscular, integumentary, immune, nervous, endocrine and reproductive systems which function in close cooperation in order to maintain the body survival.

The circulatory system is responsible for maintaining blood flow in the body. In this section, basic components of the circulatory system, the anatomy and regulation system of the heart, measurement of cardiac signals and arrhythmias are described together with related literature review.

2.1 Circulatory System

Circulatory system is responsible for transporting nutrients, oxygen, carbondioxide, hormones throughout whole body. A basic representation of the human circulatory system is given in Figure 2.1. Circulatory system has three components: the heart, the blood vessels and the blood. The system can be considered as a closed loop hydraulic system (Webster, 1998). The heart is the driving force of the system. It provides necessary pressure to the system for the circulation of blood within blood vessels. The blood vessels are the channels for blood flowing. Arteries are the blood vessels that carry blood away from the heart whereas veins are the blood vessels that carry blood toward the heart. Since circulatory system is a closed loop, there are capillaries between arteries and veins which are very narrow vessels located within tissues. The exchange of materials between blood and cells occur at capillaries.

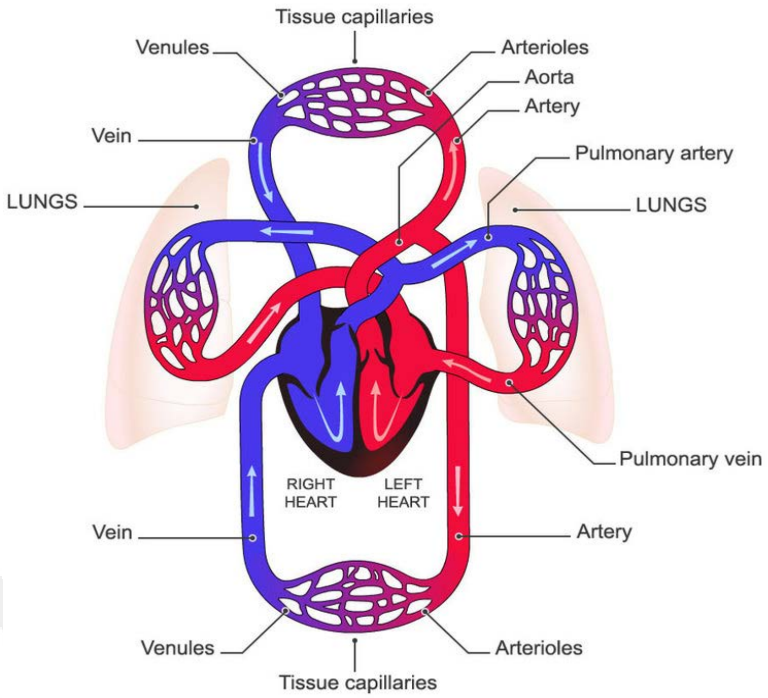


Figure 2.1 Basic representation of human circulatory system (Circulatory system, (n.d.)).

2.2 Anatomy of the Heart

The heart is a hollow muscular tube that consists of four chambers; two upper chambers called atria and two lower chambers called ventricles. These chambers are organized in a way that right atrium coworks with right ventricle to get carbon dioxide rich blood from body and pump it to lungs whereas left atrium coworks with left ventricle to get oxygen rich blood from lungs and pump it to the whole body. The basic structure of the heart is given in Figure 2.2.

The blood flow inside the heart is always unidirectional which is achieved by the organized contraction of the heart and the orientation of the cardiac valves. There are four valves within the heart:

- Tricuspid valve: Lies between right atrium and right ventricle
- Mitral valve: Lies between left atrium and left ventricle

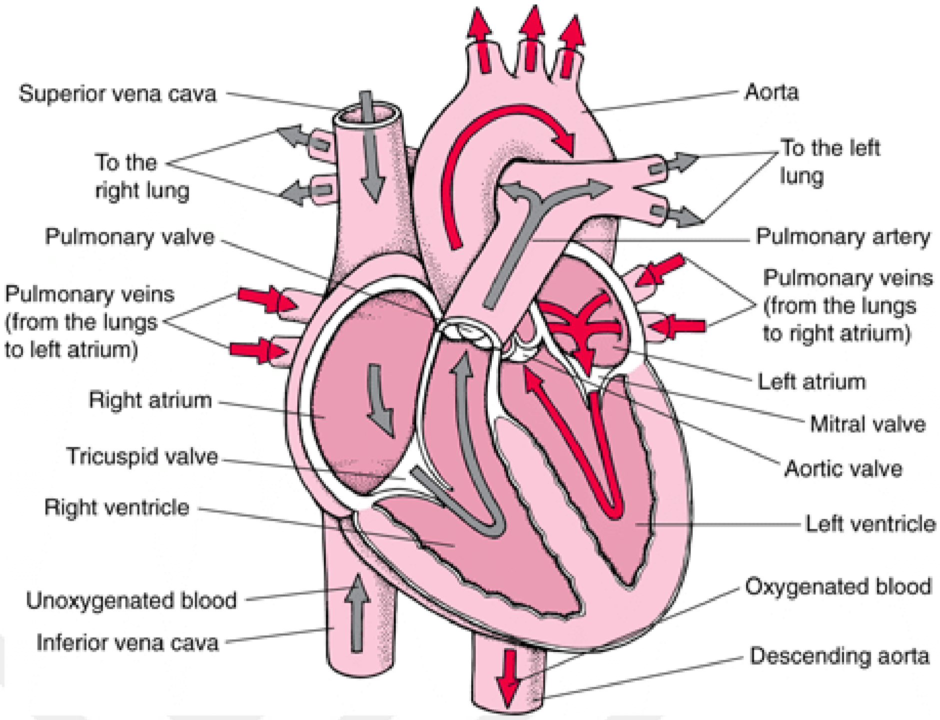


Figure 2.2 Anatomic structure of the heart and the direction of the blood flow (Webster, 1998).

- Pulmonic valve: Lies between right ventricle and pulmonary vein to lungs
- Aortic valve: Lies between left ventricle and aorta

As the blood leaves each chamber of the heart, it passes through a valve. The heart valves make sure that blood flows in only one direction through the heart. The valve is made of strong, thin pieces or flaps of tissue called leaflets. The leaflets are attached to and supported by a ring of though fibrous tissue called annulus. The annulus helps to provide support and maintain the proper shape of the valve. The valve leaflets can be compared to doors opening and closing while the annulus functions as the door frame. The leaflets of the valves which are between atria and ventricle are also supported by though, fibrous strings called chordae tendineae. The chordae tendineae extend from the valve leaflets to small muscles, called papillary muscles, which are part of the inside walls of the ventricles. The chordae tendineae and papillary muscle keep the leaflets stable against any backward flow of blood (Berne & Levy, 1997).

The right and left sides of the heart work simultaneously. The flow of the blood inside the heart during a beat can be summarized as in Table 2.1.

Table 2.1 Blood flow through the heart in one beat cycle.

Stage	Chamber	Action
Filling	Right Side	Blood enters the heart through two large veins, the inferior and superior vena cava, emptying oxygen-poor blood from body into the right atrium.
	Left Side	The pulmonary vein empties the oxygen-rich blood, from the lung into the left atrium.
Atrial contraction	Right Side	Blood flows from the right atrium into the right ventricle through the open tricuspid valve
	Left Side	Blood flows from the left atrium into the left ventricle through the open mitral valve.
Ventricular contraction	Right Side	Blood leaves the heart through the pulmonic valve, into the pulmonary artery and to the lungs.
	Left Side	Blood leaves the heart through the aortic valve, into the aorta and to the body.

2.2.1 Electroconduction System of the Heart

The heart tissue is an extremely specialized form of tissue. Most of the cardiac muscle (nearly 99%) is the contractile cells which do the mechanical work of pumping and do not initiate action potentials normally. The remainder of the cells (about 1%) is autorhythmic cells which initiate and conduct action potentials for the contraction of working cells. Although the heart is innervated by the autonomic nervous system, the heart does not require the nervous system to function because of these autorhtyhmic cells. These cells are found in sinoatrial (SA) node, atrioventricular (AV) node, bundle of His (or common bundle), the right and left bundle branches and Purkinje fiber (see Figure 2.3) (Webster, 1995).

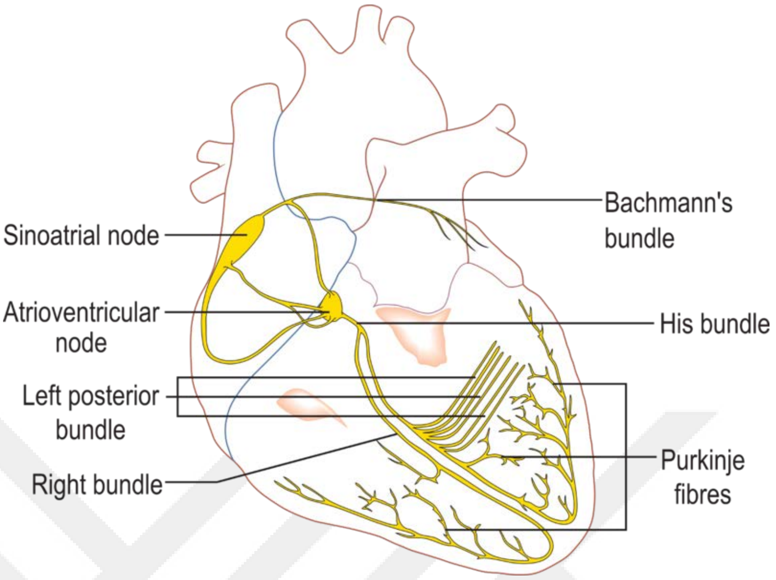


Figure 2.3 The electroconduction system of the heart (Todd, 2013).

The SA node is a small, specialized region in the right atrial wall near the opening of the superior vena cava. It can produce 70-80 action potentials per minute and has the fastest rate of action potential initiation. It dominates all other cells and known as the pacemaker of the heart. Once the cardiac impulse originates in SA node, it spreads through atria and reaches AV node via specialized ways called intermodal tracts.

The AV node is a small bundle of specialized cardiac muscle cells located at the base of the right atrium near the septum, just above the junction of atria and ventricles. It electrically connects the atria and ventricles. It can produce 40-60 action potentials per minute. One important task of the AV node is to delay electrical signals to pass to ventricles before the atria is fully empty. After the delay, the impulse continues into the bundle of His, the right and left bundle branches and the Purkinje network.

The bundle of His is a tract of specialized cells that originate at AV node and enter the septum between the ventricles. Here, it divides to right and left bundle branches that travel down the septum.

The Purkinje fibers are small terminal fibers that extend from bundle branches and are spread throughout the ventricular myocardium.

The bundle of His and Purkinje fibers can produce 20-40 action potentials per minute.

2.2.2 Heart Rate Regulation

The cardiovascular system must be able to adapt to changing circumstances to provide necessary blood to tissues. This is achieved by the autonomic nervous system (ANS) and circulating hormones (Klabunde, 2011). The effect of ANS on cardiac system is modifying the cardiac cycle length (hence the heart rate) and the speed of conduction of the electrical activity through the heart. Autonomic regulation of cardiovascular function is controlled by the two divisions of the nervous system: parasympathetic and sympathetic nerves (Berne & Levy, 1997).

The parasympathetic nerves innervating the heart originate from the cell bodies located within medulla of the brainstem. On the other hand, cardiac sympathetic fibers originate in the inter-mediolateral columns of the upper five or six thoracic and lower one or two cervical segments of the spinal cord and alter the cardiac cycle through adrenergic neurotransmitters. Parasympathetic nervous system is dominant under quiet, relaxed situations when the body is not demanding an enhanced cardiac output whereas sympathetic nervous system controls heart action in emergency or exercise situations when there is a need for increased blood flow (Sherwood, 2015). The effects of autonomic nervous system on the heart and other structures which influence the heart are summarized in Table 2.2.

Table 2.2 Effects of autonomic nervous system on the heart and other structures (Sherwood, 2015).

Affected Area	**Effect of Parasympathetic Nervous System**	**Effect of Sympathetic Nervous System**
SA node	Decreases rate of depolarization	Increases rate of depolarization
AV node	Increases AV nodal delay	Decreases AV nodal delay
Ventricular conduction pathway	No effect	Increases the speed of conduction through bundle of His and Purkinje fibers
Atrial Muscle	Decreases contractility	Increases contractility
Ventricular Muscle	No effect	Increases contractility
Adrenal Medulla	No effect	Promotes epinephrine secretion which augments the sympathetic nervous system's action on the heart
Veins	No effect	Increases venous return, which strengthens the cardiac contraction

If there was no intervention to heart from ANS, the heart would beat at 90-100 bpm which is called the intrinsic rate of the heart. However, in resting conditions, the heart is slower than the intrinsic rate due to the inhibitory influence of the parasympathetic nervous system. The monitoring for this critical homeostatic process entails firstly mechanical information about pressure in the arterial system and, secondarily, chemical information about the level of oxygen and carbon dioxide in the blood. The parasympathetic and sympathetic activity relevant to cardiovascular control is determined by the information supplied by these sensors.

The mechanoreceptors, also called baroreceptors, are located in the heart and major blood vessels; i.e. aortic arch and carotid sinuses. The nerve endings in baroreceptors are activated by deformation as the elastic elements of the vessel walls expand and contract. The chemoreceptors in the carotid bodies and aorta respond directly to the partial pressure of oxygen and carbondioxide in the blood. Both afferent systems convey their status via the vagus nerve to the nucleus of the solitary tract which relays this information to the hypothalamus and the relevant brainstem tegmental nuclei (Purves et al., 2012) .

The afferent information from changes in arterial pressure and blood gas levels reflexively modulates the activity of the relevant visceral motor pathways and, ultimately, of target smooth and cardiac muscles and other more specialized structures. For example, a rise in blood pressure activates baroreceptors that inhibit the tonic activity of sympathetic nerves and stimulates parasympathetic activity in parallel. As a result of this shift in the balance of sympathetic and parasympathetic activity, heart rate and the effectiveness of atrial and ventricular myocardial contraction are reduced and peripheral arterioles dilate, the blood pressure decreases.

The heart rate regulation mechanism is summarized in Figure 2.4. The control is made with the direct intervention of sympathetic and parasympathetic nerves to the cardiac muscle to change heart rate. Sympathetic activity also has control over adrenal medulla, arterioles, veins and strength of contraction to increase both heart rate and stroke volume.

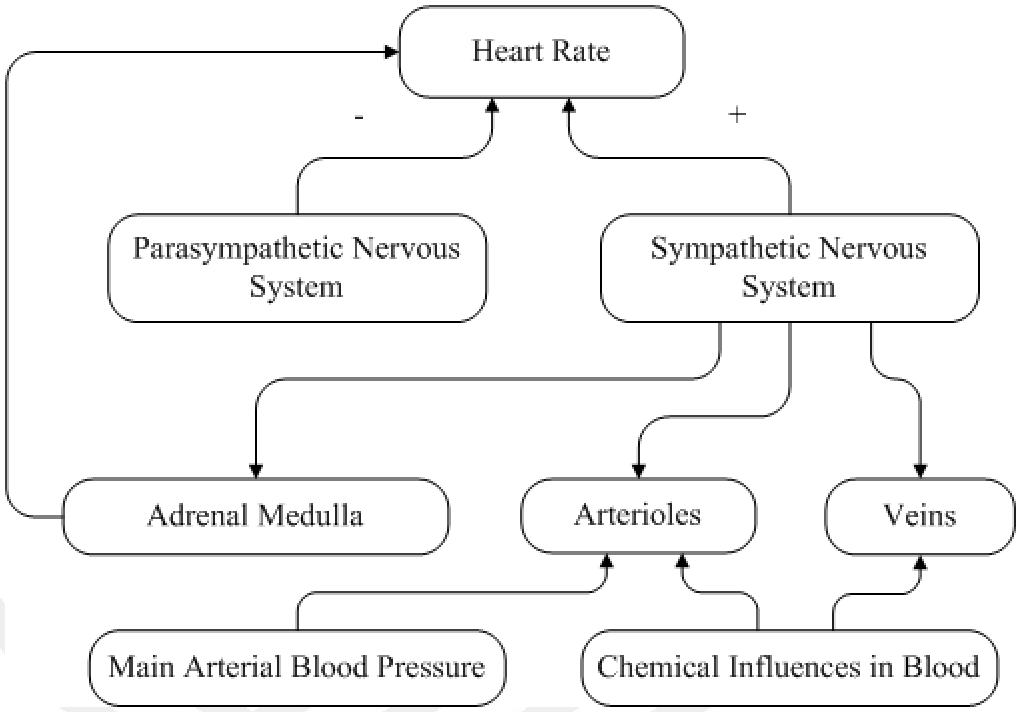


Figure 2.4 Regulation of heart rate by autonomic nervous system.

2.3 Electrocardiography

In a normal heart, each beat begins with the stimulation of the SA node which is spontaneous and regular, i.e. autorhtyhmic as mentioned in the previous section. This stimulation spreads through atria and depolarize the two upper chambers. The electrical signal from SA node reaches the AV node through specialized pathways. An impulse can reach the ventricles via the AV node only since the rest of the myocardium is separated by a non-conducting fibrous ring between atria and ventricles.

As the impulse reaches the AV node, it is momentarily delayed by the AV node as a precaution to prevent rapid atrial impulses from spreading to ventricles at the same rate. If AV node fails to receive signals from SA node, it will take the responsibility of being the pacemaker of the heart with a slower rate (40-60 bpm). Normally, SA node inhibits the impulses from AV node with its higher frequency.

Once the impulse has passed the AV node, it enters the bundle of His and spreads through right and left bundle branches then Purkinje fibers which cause the depolarization of the ventricles. If there is no stimulation from AV node, the ventricular pathways take over as the main pacemaker with 20-40 bpm frequency.

After the depolarization of the ventricles, there is a transient period where no further ionic current can flow through myocardium which is known as refractory period and lasts about 200 ms. Then, a new heart beat starts.

The most common and easy method to monitor the electrical activity of the heart is to place electrodes on the skin and record the measured electrical signal versus time. The resultant waveform is called electrocardiogram (ECG).

The choice of the electrode configuration on the thorax to record the ECG is dictated by the type of clinical information required. Since the voltage difference between a pair of electrodes (called as lead) is only representative of variations along one axis from the heart (see Figure 2.5), there is no three-dimensional activity information in single lead configuration.

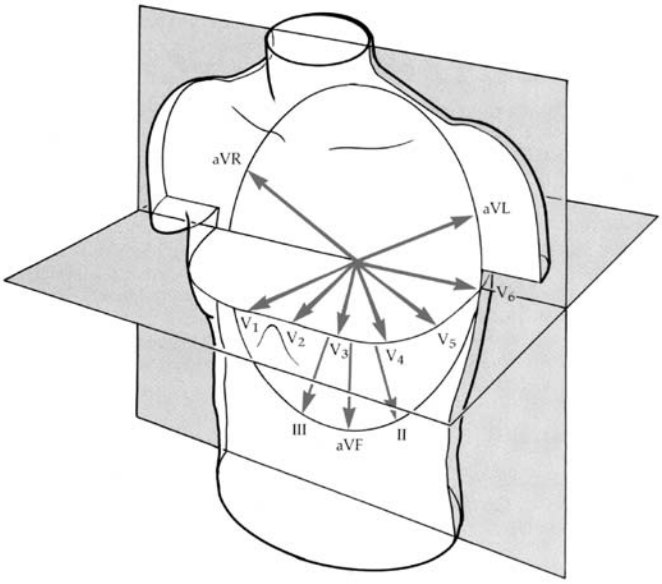


Figure 2.5 Electrical vectors for the standard ECG lead configurations (Clifford, 2002).

In order to get a detailed picture of a patient's cardiac activity, 12-lead ECG recording is made. Figure 2.6 shows the standard positions for 12 lead ECG recordings and Figure 2.7 shows the standard vectors that are visualized by these leads.

Leads I, II and III are bipolar leads which have one positive and one negative pole (see Figure 2.7). They make up the frontal-plane ECG (the plane of your body that is perpendicular to ground when you are standing erect). The three electrodes are connected to left arm (LA), right arm (RA) and left leg (LL). Also, an electrode can be attached to right leg and grounded or connected to special circuits. The resultant leads are lead I from LA to RA; lead II LL to RA and lead III from LL to LA. These three leads can be approximated by an equilateral triangle called as Einthoven's triangle as shown in Figure 2.7. Since the scalar signal on each lead can be represented as a voltage source, the Kirchoff's voltage law for three leads can be written as follows:

$$I - II + III = 0 \tag{2.1}$$

Three additional leads in frontal plane are augmented leads. These leads are unipolar leads whose negative pole is a composite pole made up of signals from multiple other electrodes, which is called Wilson's central terminal (V_w). Wilson's central terminal is produced by connecting the electrodes RA, LA and LL together through a resistive network and give an average potential across the body as in Equation (2.2). The value of the resistors should be at least 5 MΩ so that the loading will be minimum (Webster, 1998).

$$V_w = \frac{1}{3}(RA + LA + LL) \tag{2.2}$$

The signal between the central point and LA is called as aVL, and RA as aVR and LL as aVF (see Figure 2.5).

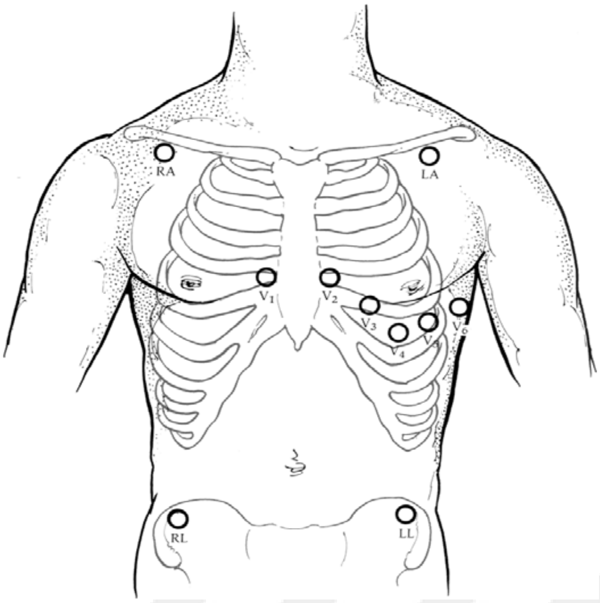


Figure 2.6 Standard electrode positions for 12-lead ECG configuration (Clifford, 2002).

In order to check the electrical activity in transverse plane (the plane of your body that is parallel to ground when you are standing erect), precordial (chest leads) are used. These six leads (V1 to V6) are also unipolar leads as seen in Figure 2.5 and their placement on the chest is shown in Figure 2.6.

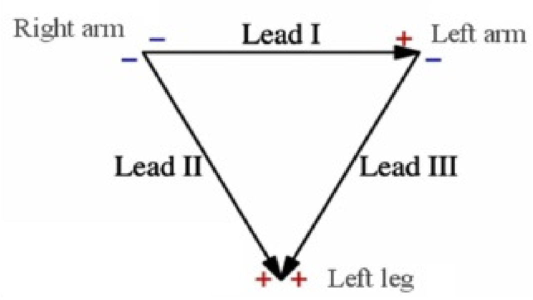


Figure 2.7 Einthoven’s triangle.

An example 12 lead ECG sheet is given in Figure 2.8. Each small square has 1 mm width and height. Also, 5mm divisions are denoted with bold lines in horizontal and vertical direction. It is standard to represent 1 mV on the y-axis as 1 cm and each second as 25 mm on the x-axis. In standard 12 lead ECG, short segments of each lead is shown with labels (see Figure 2.8) for the same measurement time. After that cycle, the new measurement cycle results are shown.

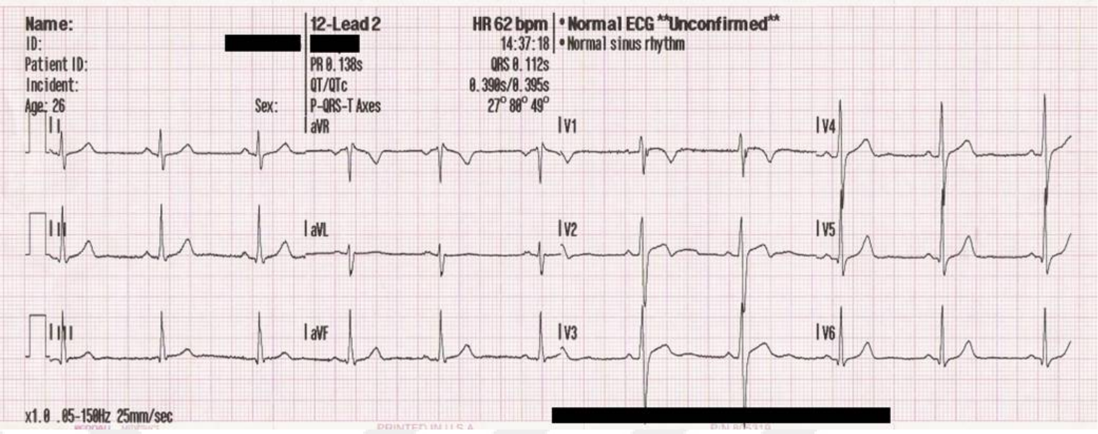


Figure 2.8 A 12 lead ECG sheet.

The ECG of a typical heart beat is shown in Figure 2.9. A beat normally comprises of an initial P-wave followed by the main QRS complex and a trailing T-wave. These waves are described as follows:

- P-wave: The low voltage fluctuation caused by the depolarization of the atria prior to contraction. The atria muscle is small, and then the voltage change is small. The duration is less than 0.12 s and amplitude is less than 0.25 mV.

- QRS complex: The largest amplitude portion of the ECG caused by the depolarization of ventricles. The time during which the ventricular contraction occurs is referred to as the systole. Although the atrial repolarization occurs simultaneously, it is not seen due to the low amplitude of the signal generated by this process. The duration of the complex is less than 0.1 s and amplitude varies in different lead configurations but upper limit is 2.5-3 mV.

- T-wave: Caused by ventricular repolarization and identification of a discrete T wave is difficult (Yanowitz, 2012).

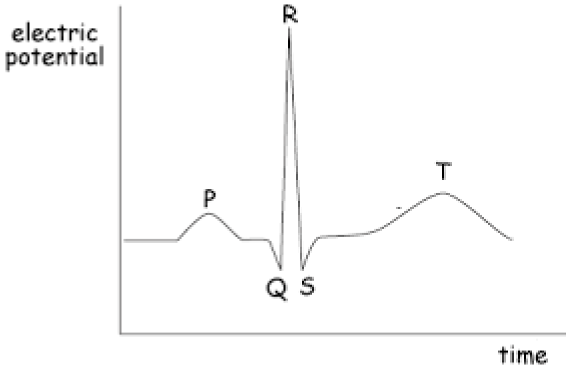


Figure 2.9 A typical ECG waveform for one heart beat. The vertical axis represents the mV fluctuations whereas the horizontal axis shows the time.

2.4 Arrhythmias

Arrhythmia can be defined as any deviation of heart’s rhythm from normal operation. The result of arrhythmias may change from nothing to death. During an arrhythmia, the heart can beat too fast, too slow or irregularly. Arrhythmias can be classified according to the underlying mechanism or the origin of the arrhythmia. Three underlying mechanism of arrhythmias are abnormal impulse initiation, abnormalities of impulse propagation and combination of both (Gertsch, 2003; Webster, 1995). Arrhythmias can also be identified according to where they occur in the heart as supraventricular or ventricular arrhythmias. Supraventricular arrhythmias include atrial tissue or AV node originated disorders whereas ventricular arrhythmias originate from ventricular tissue.

2.4.1 Ventricular Based Arrhythmias

Premature Ventricular Contractions (PVC): It is also called ventricular ectopic beats (VEB). It is an early contraction resulting from abnormal electrical activation originating in the ventricles before a normal heartbeat would occur (see Figure 2.10). Rate is variable and rhythm is irregular. Skipped beats and palpitations may be felt. Medication is needed for treatment.

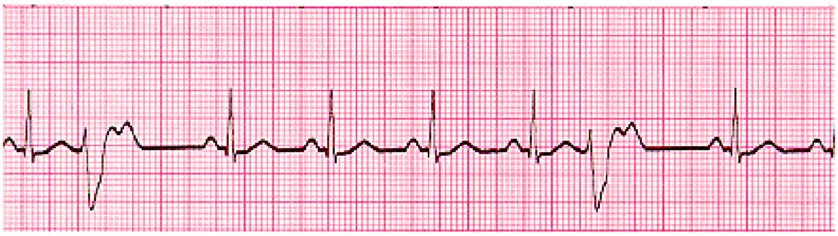


Figure 2.10 An example ECG strip with two premature ventricular contractions. Inverted peaks can be noticed between QRS complexes.

Ventricular Flutter: Fast discharging of heart from an ectopic focus in ventricles. Waveform on ECG is similar to sine wave and no identifiable P, T-waves and QRS complex (see Figure 2.11). Rate can be faster than 250 bpm and very dangerous because ventricles cannot fill completely, resulting in very low cardiac output.

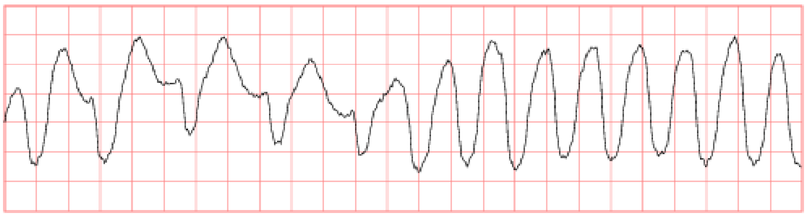


Figure 2.11 An example ECG strip with ventricular flutter. No identifiable peaks.

The patient is close to unconsciousness and defibrillation is needed to return to normal rhythm.

Ventricular Tachycardia: It is a fast arrhythmia initiated within the ventricles, characterized by 3 or more consecutive ventricular ectopic beats. Rate is over 100 bpm, the width of the QRS-complexes increases and dissociated P-waves occur (see Figure 2.12). Palpitations, paroxysmal dyspnea and syncope can be seen. Medication is needed to return to normal rhythm.

Figure 2.12 An example ECG strip with ventricular tachycardia.

Ventricular fibrillation: Uncoordinated contractions of lower chambers due to depolarization of ventricles repeatedly in an erratic, uncoordinated manner. The ECG waveform is chaotic and no identifiable P, T-waves and QRS complex (see Figure 2.13). Ventricular fibrillation is 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 patient loses consciousness rapidly. Defibrillation is needed to return to normal rhythm.

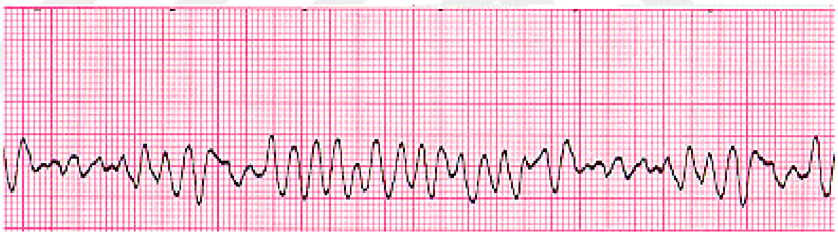


Figure 2.13 An example ECG strip with ventricular fibrillation. Waveform is completely irregular.

Ventricular Escape Beats: Beat production in ventricles due to the absence of a trigger from upper sites. Rate is slow (20-40 bpm) and rhythm is irregular. QRS complexes are broad (see Figure 2.14). This is caused by the absence of sinus node inhibitation. Skipped beats may be felt and cardiac output is low. Medication is needed to increase the rhythm.

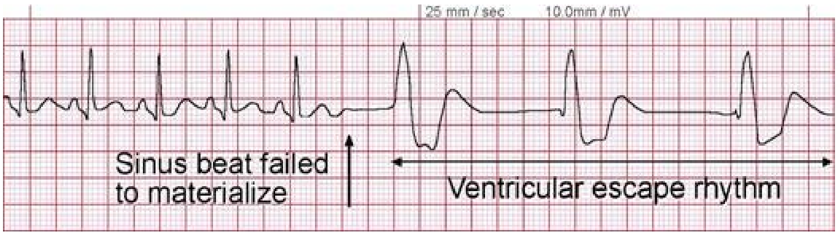


Figure 2.14 An example ECG strip with ventricular escape beats. QRS complexes are wider and rate is slower.

Bundle Branch Block: It is a conduction block in left or right bundle branches and can reduce pumping efficiency of ventricles. Abnormalities in QRS complex (m-shaped QRS complex, blurred S-wave, notched R-wave) exist (see Figure 2.15). Syncope can be seen in patients. The treatment should be solving the underlying mechanism of the block.

Figure 2.15 An example ECG strip with right bundle branch block with wide, blurred S-wave.

2.4.2 Supraventricular Based Arrhythmias

Sinus Bradycardia: The initiation of impulses from SA node is slower than normal. The heart rate is lower than 60 bpm (see Figure 2.16) and results in decreased cardiac output. The underlying mechanism is increased parasympathetic activity and a pacemaker is used for treatment.

Figure 2.16 An example ECG strip with sinus bradycardia.

Sinus Tachycardia: The initiation of impulses from SA node is faster than normal. The heart rate is above 100 bpm (see Figure 2.17) and results in decreased cardiac output. The underlying mechanism is external factors, hormones, infarction. Medication is necessary for treatment.

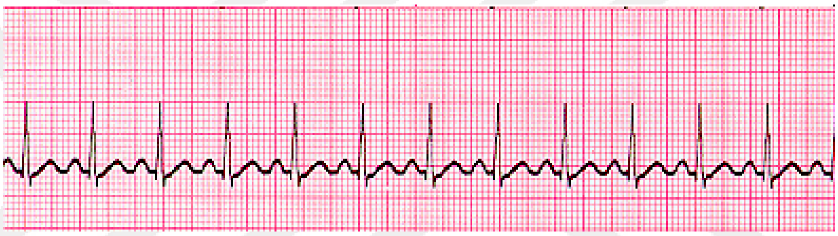


Figure 2.17 An example ECG strip with sinus tachycardia.

SA Block: SA block is grouped into different classes according to severity. 1st degree SA block is the conduction delay between SA node and atria. Rate and rhythm is normal. 2nd degree SA block type I is the conduction delay and block between SA node and atria. It includes the conduction delay and sometimes the conduction block. Thus, PP intervals shorten. 2nd degree SA block type II is the conduction block between SA node and atria intermittently (see Figure 2.18). Rate is slower because some pulses are not conducted to atria. 3rd degree SA block is the complete conduction block between SA node and atria. Rate and rhythm are irregular because no pulse is transmitted to atria from SA node. Escape rhythm is the pacemaker. The underlying mechanisms may be medication, myocardial infarct, change in anisotropy or gap junction resistance. SA block patients feel escape beats and if the delay is longer than 3s, dizziness, weakness, syncope, chest pressure is felt. Medication to increase SA node activity or pacemaker is needed for treatment.

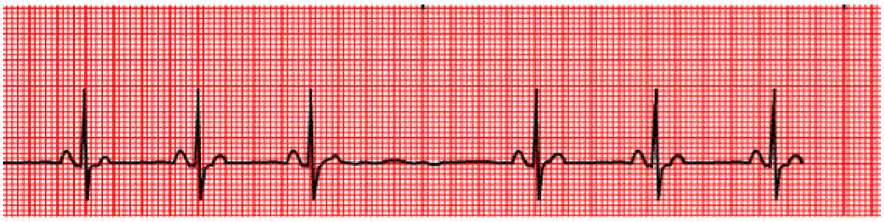


Figure 2.18 An example ECG strip with 2nd degree Type II SA block.

Sinus Arrest: SA node stops producing impulses (see Figure 2.19). Rate varies due to pauses and rhythm is irregular. Sinus node disease or increases in parasympathetic activity may cause sinus arrest. If the pauses are between 3-9 s and frequent, low cardiac output symptoms are seen. Medication to increase SA node activity is needed. If the underlying reason is myocardial infarct, pacemaker is necessary.

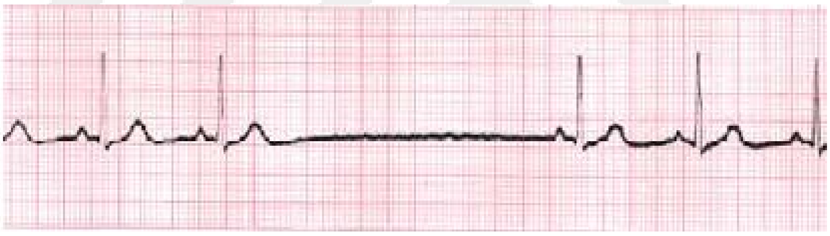


Figure 2.19 An example ECG strip with sinus arrest.

Sick Sinus Syndrome: This arrhythmia is also called sinus node dysfunction. Slow and fast arrhythmias alternate (brady-tachy) in sick sinus syndrome resulting in irregular rate and rhythm (see Figure 2.20). A damage or scar in the conduction

Figure 2.20 An example ECG strip with sick sinus syndrome.

system may result in sick sinus syndrome. Both pacemaker and medication are needed for treatment of slow and fast arrhythmias.

AV Block: AV block is grouped into different classes similar to SA block. 1st degree AV block is the conduction delay between SA node and AV node. PR interval is longer than 210 ms (see Figure 2.21). No symptom is felt. 2nd degree AV block type I, also called Wenckebach, is the conduction delay in AV node. Delays in PR interval increase progressively and a beat is dropped finally resulting in low cardiac output (see Figure 2.22). 2nd degree AV block type II, also called Mobitz, is the intermittent conduction block in His bundle or bundle branches. Some SA node impulses are not conducted to ventricles resulting in some P waves not followed by QRS complexes (see Figure 2.23). This type of AV block may progress to complete heart block, which may result in cardiac arrest or sudden cardiac death. 3rd degree AV block is the complete conduction block in AV node. There is no conduction between atria and ventricles. Ventricular escape rhythm is dominant and rate is slower than 60 bpm (see Figure 2.24). Coronary ischemia or degeneration of the electrical conduction system may cause complete block and results in low cardiac output. Medication to increase SA node activity or pacemaker is needed for treatment.

Figure 2.21 An example ECG strip with 1st degree AV block. Notice PR interval is longer.

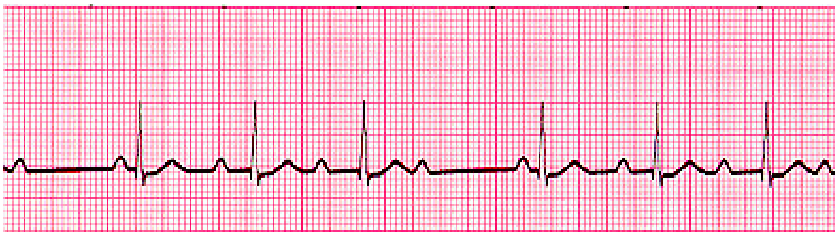


Figure 2.22 An example ECG strip with 2nd degree AV block type 1 (Wenckebach). PR interval increases and finally a beat is dropped.

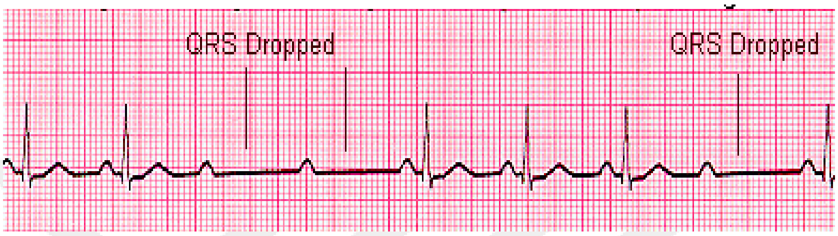


Figure 2.23 An example ECG strip with 2nd degree AV block type 2 (Mobitz). Notice P-waves are not followed by QRS complexes.

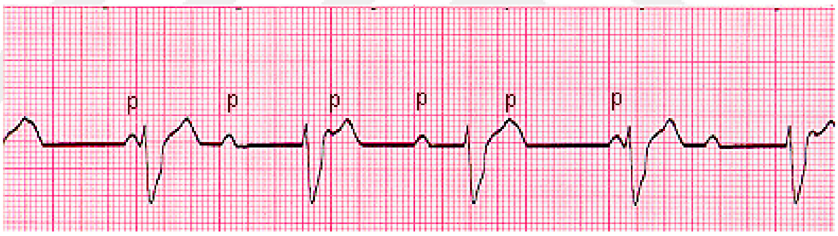


Figure 2.24 An example ECG strip with 3rd degree AV block. Notice the dissociation between atrial and ventricular contractions.

Premature Atrial Contractions (PAC): It is also called atrial ectopic beats (AEB). Normally SA node regulates the heart rate. However, if another region of the atria (ectopic focus) depolarizes faster than SA node, premature atrial contraction is triggered. Rate is normal, but rhythm is irregular due to premature contractions (see Figure 2.25). Skipped beats may be felt but generally no treatment is needed.

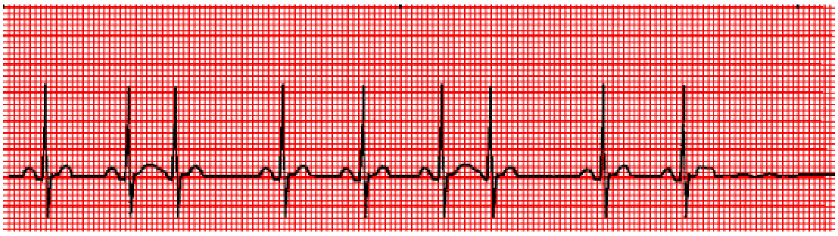


Figure 2.25 An example ECG strip with premature atrial contractions. The 3rd and 7th beats are PACs.

Atrial Tachycardia: Beat production in atria is from an ectopic focus and faster than SA node. The rate is between 140-250 bpm. Rhythm is irregular (see Figure 2.26). Patient feels dizziness, syncope, chest pain or pressure, shortness of breath and chest palpitations.

Figure 2.26 An example ECG strip with atrial tachycardia. Atrial rate is faster and P-wave morphology is abnormal.

Atrial Flutter: Beat production in atria is faster than 250 bpm and out of sync with ventricular rhythm (see Figure 2.27). Patients feel heart palpitations, shortness of breath, discomfort in chest and dizziness. Cardioversion to return to normal rhythm and ablation is suggested for treatment.

Figure 2.27 An example ECG strip with atrial flutter. Sawtooth pattern between QRS complexes can be seen. It is a 3:1 block.

2.4.3 Atrial Fibrillation

Atrial fibrillation is also a supraventricular based arrhythmia. Since it is the main interest of this thesis, a separate section is dedicated to this arrhythmia. The mechanism behind AF is focal activation or multiple wavelets. Atrial fibrillation is based on the generation of random impulses at multiple areas of the atria. This results in uncoordinated atrial activity, resulting in deterioration of mechanical function and insufficient filling/emptying. The rate of the atrial impulses is 300-650 bpm and they cannot be conducted to the ventricles because of the recovery period of AV node. Thus the ventricles have an irregular depolarizing rate (Webster, 1995) and ventricular rate, i.e. heart rate, might be 100-175 bpm. The symptoms vary from nothing to breathlessness, palpitations, syncope, dizziness, chest discomfort and stroke/transient ischemic attack (Cowan et al., 2014). This arrhythmia can be detected with disorganized electrical activity between QRS complexes and absent P waves (see Figure 2.28).

Figure 2.28 An example ECG strip with atrial fibrillation. Uncoordinated electrical activity can be noticed between QRS complexes. QRS complexes are irregular.

Classification of atrial fibrillation is necessary because different cases need different treatments. Atrial fibrillation is classified into 3 groups according to termination characteristics (Camm et al., 2010; Fuster et al., 2006; Gallagher & Camm, 1998; Lévy et al., 2003; Lip & Tse, 2007):

1. Paroxysmal atrial fibrillation terminates spontaneously; usually within 48 h. Although the termination is defined 48 h, the reality is that the episodes generally lasts shorter (even minutes) and patients present with sinus rhythm on their ECGs (Hoshino et al., 2013; Page et al., 1994).

2. Persistent atrial fibrillation is present when AF episode lasts more than 7 days. Cardioversion with drug or direct current is needed to return to normal rhythm.

3. Permanent atrial fibrillation exists when AF episode cannot be reverted to normal rhythm.

The major risk of paroxysmal atrial fibrillation is that it increases the predisposition to stroke due to the clot formation in left atria. Uncoordinated contractions in atria reduce the transfer of all blood to ventricles and there is some remaining blood inside which may form clots. Studies show that the stroke risk of PAF patients is similar to persistent/permanent AF patients (Hohnloser et al., 2007; Lip & Li Saw Hee, 2001).

The treatment of PAF has three goals (Lip, 1999):

1. Suppressing paroxysms of AF and maintaining long term sinus rhythm with suitable medication
2. Controlling heart rate during paroxysms of AF by increasing the refractoriness of AV node with medication. Pacemaker implantation and ablation are also used to control ventricular rate (Heist, Mansour, & Ruskin, 2011)
3. Preventing the complications associated with PAF; i.e. stroke with antithrombotic therapy.

CHAPTER THREE
HEART RATE VARIABILITY ANALYSIS

3.1 Background

Heart rate variability (HRV) is the evaluation of the fluctuations in the time intervals between heart beats, known as RR intervals. The importance of HRV is that it can reveal information about the autonomic nervous function, sympathetic-parasympathetic balance and cardiovascular health (Berntson, 1997; Camm, Malik, Bigger, & Breithardt, 1996; Malik & Camm, 1995). In the last three decade, HRV has been popular for the investigation of cardiovascular physiology. Before HRV, complex invasive techniques in animal models or imprecise reflex based tests in humans were required for the investigation of autonomic physiology. HRV analysis has provided a simple reproducible and non-invasive method for autonomic assessment (Pumprla, Howorka, Groves, Chester, & Nolan, 2002).

HRV analysis is the key point of this thesis. The feature space, which was used for the input of the classifiers, was constructed with the outcomes of the HRV analysis.

3.1.1 Clinical Applications of HRV

The first clinical use of HRV is in 1965 when Hon and Lee noticed that fetal distress was accompanied by the changes in beat-to-beat variation of the fetal heart (Hon & Lee, 1963). In the 1970s, Ewing et al. showed that short-term HRV measurement is a marker of diabetic autonomic neuropathy (Ewing, Martyn, Young, & Clarke, 1985). In 1977, Wolf et al. found that patients with reduced HRV after a myocardial infarction (MI) had an increased mortality (Wolf, Varigos, Hunt, & Sloman, 1978). Later, other work supported that persistent low HRV after MI is a strong predictor of mortality post MI (Bigger et al., 1992; Bigger, Fleiss, Rolnitzky, Steinman, & Schneider, 1991; Kleiger, Miller, Bigger, & Moss, 1987; Malik, Farrell, Cripps, & Camm, 1989). Since then, HRV analysis has become a frequently visited research area in relation to several cardiovascular diseases, physical exercise, stress, gender, age, sleep, smoking and bio-feedback (Acharya, Joseph, Kannathal, Lim, & Suri, 2006; Achten & Jeukendrup, 2003; Camm et al., 1996; Malik & Camm, 1993; Pumprla et al., 2002; van Ravenswaaij-Arts, Kollee, Hopman, Stoelinga, & van Geijn, 1993). In more recent work, Zuern et al. (Zuern, Barthel, & Bauer, 2011) and Huikuri & Stein (Huikuri & Stein, 2012) evaluated HRV and heart rate turbulence (HRT) as a tool for risk assessment for the patients recovering from MI. Perkiömäki has reported that HRV indices that quantifies the non-linear dynamics of HR may have a greater prognostic value to identify patients with greatest risk for adverse cardiovascular events than do conventional HRV indices (Perkiömäki, 2011). Papaioannou et al. investigated the association between changes in HRV and the inflammatory response in patients with cardiovascular diseases (Papaioannou, Pneumatikos, & Maglaveras, 2013). Different changes in HRV produced by physiological and pathological stress were explored by Bravi et al. (Bravi et al., 2013). Hinojosa-Laborde et al. investigated whether any HRV index could accurately distinguish between individuals with high and low tolerances to simulated hemorrhage (Hinojosa-Laborde, Rickards, Ryan, & Convertino, 2011). Tobaldini et al. used linear and non-linear HRV analysis to assess autonomic changes during sleep under physiological and pathological conditions such as sleep-related breathing disorders or insomnia (Tobaldini et al., 2015). İşler & Kuntalp found that combining classical HRV indices with wavelet entropy measures improves the performance in diagnosing congestive heart failure patients (Işler & Kuntalp, 2007).

Although HRV has found a wide application area as mentioned, clinical implication of HRV analysis has been recognized in two areas: (1) as predictor of arrhythmic events or sudden cardiac death after acute myocardial infarction (MI) (Acharya et al., 2006; Camm et al., 1996; Laitio, Jalonen, Kuusela, & Scheinin, 2007) and (2) as an early warning sign of diabetic neuropathy (Acharya et al., 2006; Camm et al., 1996).

3.1.2 Physiological Origins of HRV

The cardiovascular system achieves dynamical stability by autonomically mediated control of heart rate, blood pressure and other factors which respond to internal and external stimuli such as acute ischemia, metabolic imbalance and changes in physical or mental activity (Pumprla et al., 2002).

The main pacemaker of the heart, the SA node, is highly innervated by both sympathetic and parasympathetic nervous system and reflects their modulating effect. Parasympathetic activity slows the heart rate by synaptic release of acetylcholine, which has a very short latency and high turnover rate. This enables the parasympathetic nervous system to regulate cardiac function on a beat to beat basis. Sympathetic activation increases the heart rate, conduction speed and contractility. This is achieved by the release of noradrenalin which is reabsorbed and metabolised relatively slowly. Therefore, the changes in cardiac function under sympathetic activity have a slower time course. Because of the different behavior of these neurotransmitters, sympathetic and parasympathetic nervous system operates at different frequencies. Thus, their activity can be identified and quantified by the variation of heart rate (Greenwood, Batin, & Nolan, 1997; Greenwood, Durham, & Nolan, 1998). Thus, analysis of HRV data in frequency domain provide the basis for non-invasive semi-quantitive assessment of autonomic activity (Camm et al., 1996; Eckberg, 1997).

In normal individuals, there are cyclic oscillations as a result of respiration (Pagani et al., 1986). This respiratory related variation occurs typically around 0.25 Hz and can be abolished by vagal blockade (Keselbrener & Akselrod, 1998). These two factors suggest that this particular type of high frequency cyclic HRV is parasympathetically mediated (Pumprla et al., 2002).

Another cyclic oscillation occurs with the changes in baroreceptor activity due to fluctuations in blood pressure (Sleight et al., 1995). This baroreceptor mediated activity occurs at a lower frequency, typically around 0.1 Hz and can be significantly modified by sympathetic blockade (Keselbrener & Akselrod, 1998). There is also close relationship between this low frequency variation in heart rate and direct measures of muscle sympathetic nerve activity (Pagani et al., 1997). These factors suggest that sympathetic activity is a strong mediator of this low frequency. However, some studies showed that vagal blockade also produces some changes in this low frequency. Therefore, measurement of low frequency cyclic HRV is not a direct quantitive index of sympathetic activity (Keselbrener & Akselrod, 1998). However, simultaneous measurement of high and low frequency can be used to investigate the sympathovagal balance (Bootsma et al., 1994; Montano et al., 1994).

There are also very slow cyclic variations equal to or less than 0.01 Hz. The mechanism behind this modulation is not well defined but thought to be related to changes in autonomic activity associated with thermoregulatory mechanism (Fleisher et al., 1996), changes in peripheral chemoreceptor activity (Francis et al., 2000; Ponikowski et al., 1997) and fluctuations in the activity of the renin-angiotension and parasympathetic systems (Duprez et al., 1995; Taylor, Carr, Myers, & Eckberg, 1998).

3.2 Preprocessing

Signals obtained from real life measurements always include unwanted parts such as interference from other sources, artifacts or simply noise. Before the processing of the signals, these undesirable components must be eliminated for accurate results. This is called preprocessing of data. In this study, the preprocessing starts with derivation of cardiovascular time series; i.e. derivation of RR interval series from ECG signals. Then, the artifacts caused by the measurement and ectopic beats are eliminated. The next step is dividing RR series into 5 min segments for short time HRV analysis. Preprocessing steps are explained with details in the following sections.

3.2.1 Derivation of Cardiovascular Time Series

HRV analysis is the examination of the time series which are constructed from the durations between consecutive heart beats. QRS complex is the most dominant pattern in ECG and used as the marker of a heart beat. Several methods have been proposed for the detection of QRS complex.

Several methods have been investigated by researchers for accurate QRS detection. Wavelet transform (Kadambe, Murray, & Boudreaux-Bartels, 1999; Li, Zheng, & Tai, 1995), Hilbert transform (Benitez, Gaydecki, Zaidi, & Fitzpatrick, 2001), artificial neural networks (Dokur, Olmez, Yazgan, & Ersoy, 1997; Vijaya, Kumar, & Verma, 1998), digital filters (Afonso, Tompkins, Nguyen, & Luo, 1999; Keselbrener, Keselbrener, & Akselrod, 1997), geometrical matching (Suárez, Silva, Berthoumieu, Gomis, & Najim, 2007), slope vector waveform (Xu & Liu, 2004) and difference operation method (Yeh & Wang, 2008) are some of the suggested algorithms.

The algorithm suggested by Pan & Tompkins has been widely accepted (Pan & Tompkins, 1985). The algorithm includes the steps in Figure 3.1. The energy of QRS complex is concentrated approximately between 5-15 Hz (Pan & Tompkins, 1985). Then the signal is filtered by a cascaded low-pass and high-pass filter firstly.

The next step in QRS detection is differentiation which makes R points more clear in ECG signal and also attenuates the P and T-waves. Squaring the signal is a nonlinear transformation and increases the output of the differentiation process. Since there might be large amplitude and long duration QRS complexes which cannot be detected with the slope of R-waves, moving window integration is necessary for further feature extraction. The last step in QRS detection is applying an adaptive threshold to detect R-peak locations.

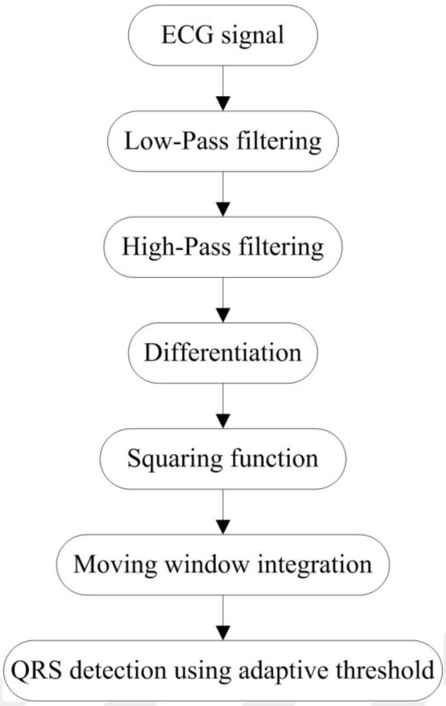


Figure 3.1 Block diagram of QRS detection algorithm.

Once the locations of R peaks are detected, RR interval time series can be derived. Firstly, the time differences between consecutive R peaks are calculated, i.e. n'th RR interval is calculated by $T_n = t_n - t_{n-1}$ (see Figure 3.2a). Then, three methods are offered in literature for the representation of RR interval series. The simplest approach is to represent RR intervals as a function of the beat number which is also called RR tachogram (see Figure 3.2b) (Baselli et al., 1987). However, this assumption causes distortion in spectral analysis (Mateo & Laguna, 2000). Another approach, which is used in this study, is to represent RR intervals as a function of RR durations (see Figure 3.2c). In this representation, the series is not equidistantly sampled and must be interpolated and resampled for correct frequency domain analysis (Camm et al., 1996). The third approach is spectrum of counts method which places impulse functions at R peak occurrence times (see Figure 3.2d). This method is based on integral pulse frequency modulator (IPFM) which aims to reflect the neural modulation of SA node (Rompelman, 1993). In this method, the modulating signal is integrated until a threshold is achieved which an impulse is emitted and the integrator is set to zero. Then, the spectrum of the series can be calculated by first lowpass filtering the series and then calculating the spectrum (Deboer, Karemaker, & Strackee, 1984).

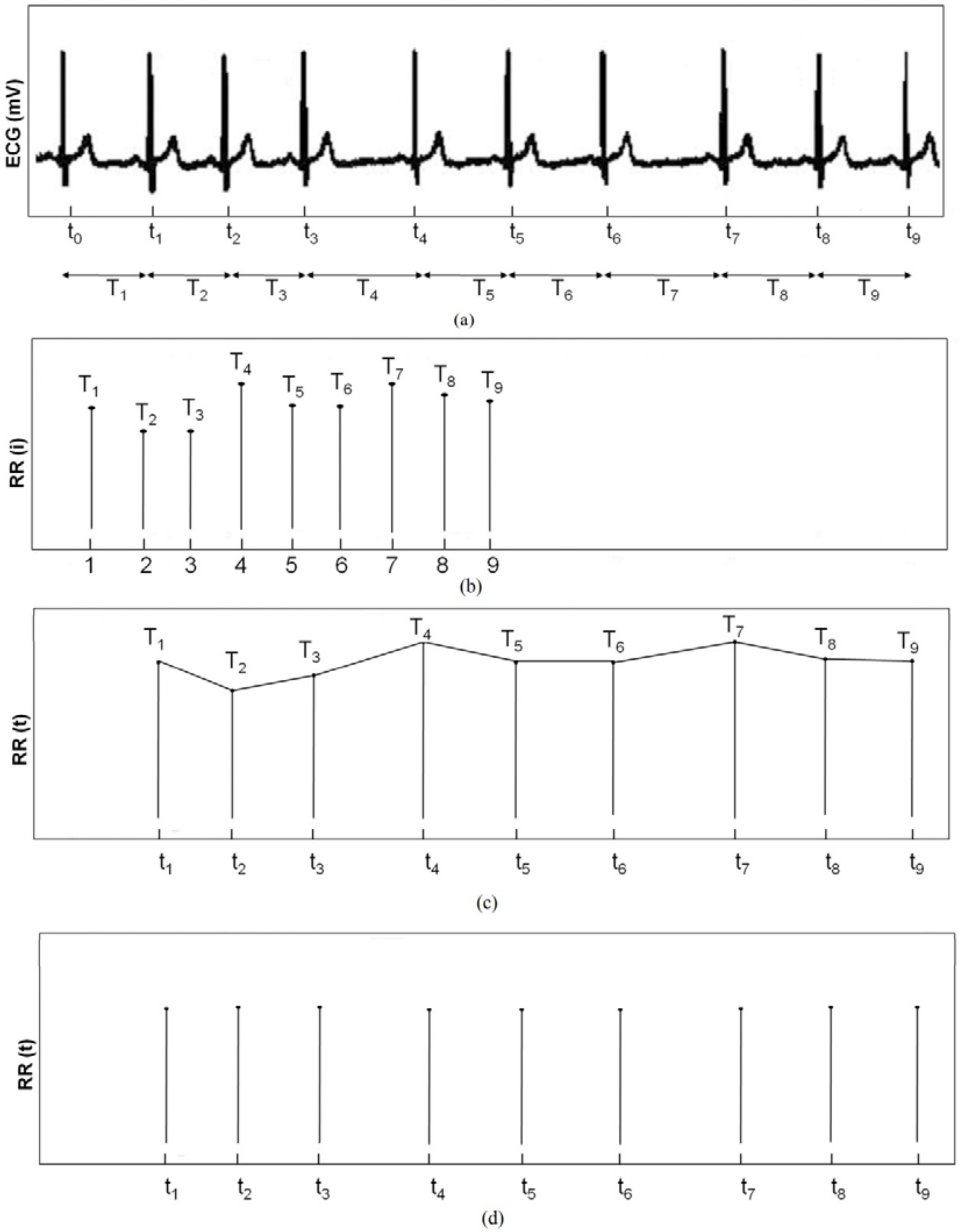


Figure 3.2 (a) ECG signal with beat occurrence times (t_n) and RR intervals (T_N). (b) RR intervals as a function of beat number. (c) RR intervals as a function of beat occurrence times. (d) RR intervals represented with impulses in IPFM.

3.2.2 Segmentation

Some HRV metrics are affected by the length of the RR interval data used for the analysis. Thus, it is inappropriate to compare the HRV measures obtained from different duration signals. For this reason, the HRV analysis uses short-term 5 min or long-term 24 h recordings (Camm et al., 1996). In this study, short–term HRV analysis was preferred because of the available dataset, whose the recordings were 30 min long, and for the ease of use in real application of the method. Then, six 5 min data segments were obtained from each recoding without any overlap. Totally 288 five minute segments from PAF patients and 510 five minute segments from non-PAF subjects were obtained. The segments belonging to the same subject were labeled in such a way that the data belonging the same subject never be in test and train group at the same time during validation in order to avoid any false bias due to correlation.

3.2.3 Artifact Removal

In HRV analysis, it is desirable to examine the autonomic function; i.e. the function of sympathetic and parasympathetic nervous system (Camm et al., 1996). However, RR interval time series include physiological or technical artifacts. Physiological artifacts are ectopic beats or fibrillation waves. Technical artifacts can be poorly fastened electrodes or movement of the subject which result in abnormalities in ECG and thus problems in QRS detection. All artifacts should be eliminated for accurate HRV analysis.

Ectopic beats originate from heart tissue other than autonomic cells and are not under the control of autonomic nervous system. Further, they give an impulse type appearance in the RR interval time series (see Figure 3.3a) and can give erroneous statistical measures in HRV analysis if not removed (Thuraisingham, 2006).

Deletion and interpolation are two common methods for the correction of ectopic beats (Peltola, 2012). In deletion method, abnormal RR intervals are just omitted from the series and remaining data is concatenated (see Figure 3.3b). In interpolation method, abnormal RR intervals are replaced with new interpolated RR intervals. Interpolation of degree zero, linear, spline and non-linear predictive interpolation are among various interpolation algorithms (see Figure 3.3c) (Peltola, 2012). Both methods have different side effects on HRV analysis and different studies have been made to find out how different editing methods affect HRV analysis (Albrecht & Cohen, 1988; Lippman, Stein, & Lerman, 1994; Tarkiainen et al., 2007). However, the studies are on different populations and the lengths of RR interval time series used are different. Then, it is difficult to say one method is superior to other method. General acceptance is that deletion method is suitable for time domain analysis whereas interpolation method is more suitable for frequency domain analysis (Salo, Huikuri, & Seppanen, 2001).

While selecting the ectopic beat correction method in this study, RR interval time series were examined to see the patterns of the ectopic beats. Two patterns are observed: single ectopic beats and recurrent ectopic beats. In single ectopic beat case, the preceding and the following beats of the abnormal beats are around the average RR value. Then, deleting the ectopic beat and related abnormal beats does not cause any problems in frequency domain analysis. In recurrent ectopic beats case, numerous ectopic beats are excluded as suggested by other work (Kamath & Fallen, 1995). In both cases, the duration of the remaining signal was checked and the signals longer than 270 s were accepted adequate for short-term HRV analysis.

The algorithm suggested by Langley et al. was used to detect and eliminate atrial and ventricular ectopic beats (Langley et al., 2001). The method detects ectopic beats by identifying short RR intervals relative to average of the surrounding RR intervals (RR_{av}). The number of RR intervals in moving average window and thresholds were optimized for the RR interval series used in this study. The algorithm is given in Figure 3.4. Ectopic beat detection starts with taking the average of four consecutive beats; i.e. RR_{av}. If the next beat (RR_k) is below some percentage of the average, it is a suspected beat. Langley et al. took this threshold about 80% of the average.

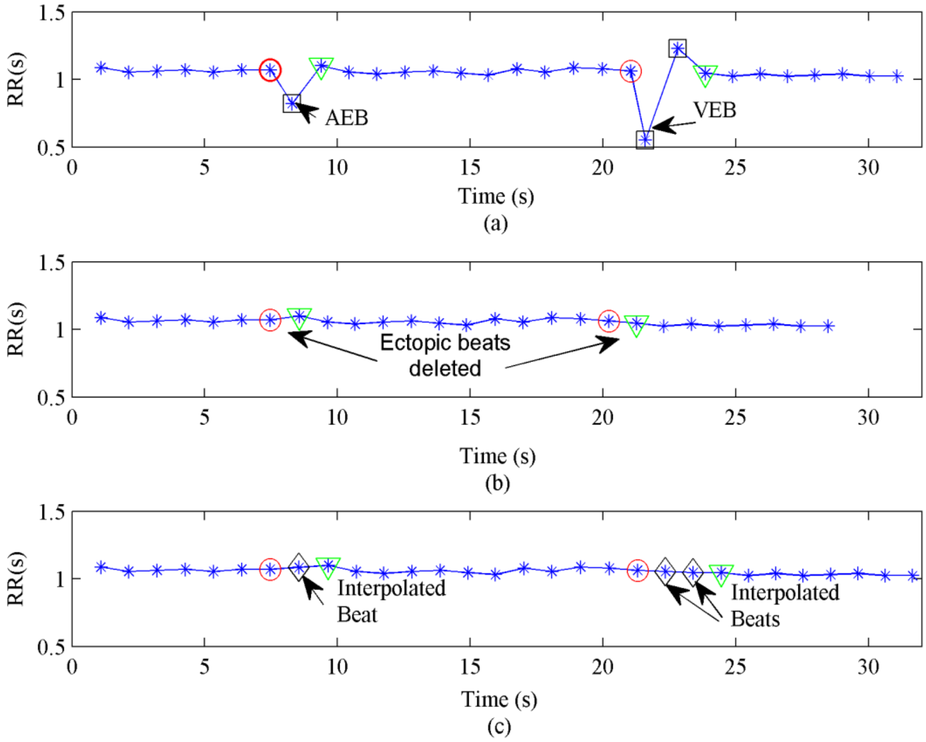


Figure 3.3 Two methods for ectopic beat elimination (a) RR series with ectopic beats (AEB: Atrial Ectopic Beat, VEB: Ventricular Ectopic Beat (b) Deletion method (c) Interpolation method.

When this threshold was used in RR time interval series in this thesis, some obvious ectopic beats were not eliminated. After some trials, the best elimination performance was found as 92% of the average. If RR_k is below 92% of the average (RR_{av}), the next beat (RR_{k+1}) is checked. If it is +/-10% range of the average, RR_k is an atrial ectopic beat and RR_k is omitted from the series. If the change between RR_{k+1} and the average is 30% greater than the change between RR_k and the average, RR_k is a ventricular ectopic beat. Then RR_k and RR_{k+1} are omitted from the series. If the sum of the RR_k and RR_{k+1} is within 10% of the average, it is accepted as wrong detection of T wave as an R wave and both RR_k and RR_{k+1} are omitted. During examination of the data, it was observed that some RR intervals were very long with respect to average. The ECG records of that RR series were checked and seen that there were some artifacts in the signal such as saturation or electrode movement. These components were eliminated from the series by omitting the RR intervals which were 120% of the average.

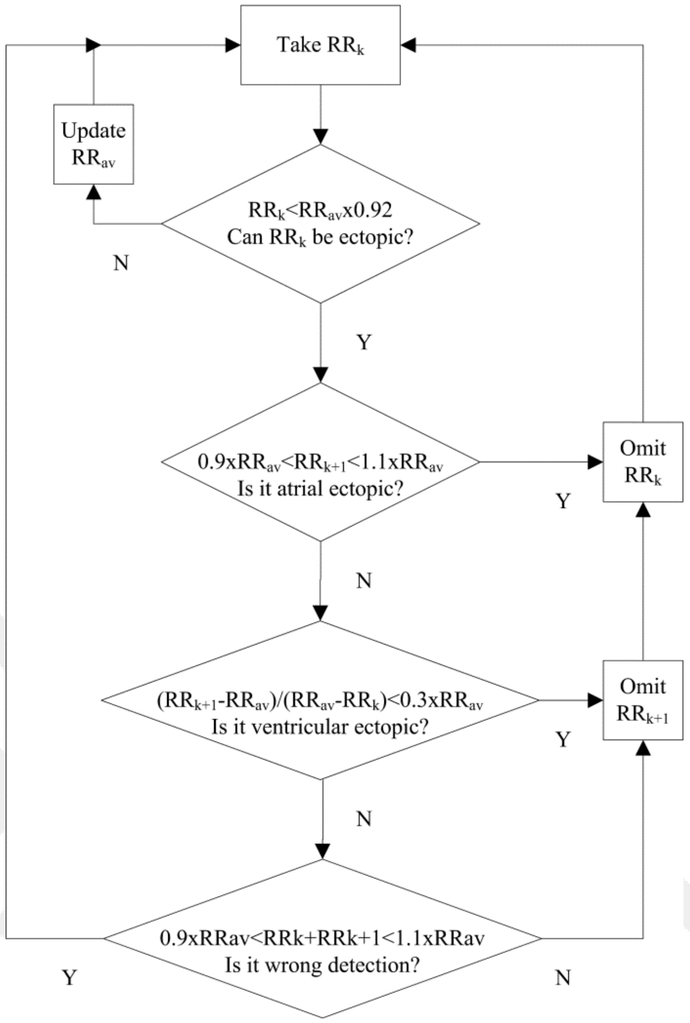


Figure 3.4 Ectopic beat elimination algorithm.

3.3 Feature Extraction from RR Interval Data

In this section, the derivation of HRV features from short-term RR series was explained. The analysis methods are based on guideline (Camm et al., 1996). HRV analysis can be grouped as time domain, frequency domain and non-linear methods. The features were extracted with Kubios HRV software (Tarvainen, Niskanen, Lipponen, Ranta-Aho, & Karjalainen, 2014). Derivation of features and the selection of parameters used for the calculation of each feature are given in the following sections.

3.4 Feature Normalization

The values of the features obtained from RR interval series varies widely and may lead to wrong operation of the classifiers. Also, some studies show that normalization improves the performance of the classifiers (Kotsiantis, Kanellopoulos, & Pintelas, 2006). In this thesis, min-max normalization was used. In this method, all the samples are normalized to the range [0,1] by

$$f_{i,N} = \frac{f_i - \min(f_i)}{\max(f_i) - \min(f_i)} \quad , \qquad i = 1, 2, \ldots, \mathrm{d} \tag{3.1}$$

where d is the number of features, $f_{i,N}$ is the normalized i-th feature, f_i is the i-th feature, min(f_i) and max(f_i) are the minimum and maximum values of the i-th feature respectively.

3.4.1 Time Domain Analysis

Time domain analysis is the statistical examination of the fluctuations in RR intervals and commonly used because of easy calculation. Statistical analysis and geometrical analysis are two branches of time domain HRV analysis. Geometrical methods require long-term RR interval data and not used in this thesis (Camm et al., 1996).

In some sources, normal-to-normal (NN) interval phrase is used to indicate that the intervals are obtained from QRS complexes originated from SA node depolarization (Camm et al., 1996). In practice, the NN and RR intervals are the same and RR is preferred in this thesis.

The statistical measures and their derivation are summarized below:

Mean NN (s): Arithmetic mean value of all RR intervals. For a series with length N, the mean is calculated as

$$Mean\ NN = \overline{NN} = \frac{1}{N}\sum\nolimits_{i=1}^{N} NN_i \tag{3.2}$$

SDNN (s): Standard deviation of RR intervals which reflect overall variation and defined as

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(NN_j - \overline{NN}\right)^2} \tag{3.3}$$

SDHR (Hz): Standard deviation of instantaneous heart rate. Instantaneous heart rate is calculated with $HR = 60/RR$ and standard deviation of instantaneous heart rate is

$$SDHR = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(HR_j - \overline{HR}\right)^2} \tag{3.4}$$

RMSSD (s): Root mean square of successive differences between RR intervals and calculated as (Berntson, 1997)

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}\left(RR_{j+1} - RR_j\right)^2} \tag{3.5}$$

NN50 (n.u.): Number of successive NN interval pairs that differ by more than 50 ms (Bigger et al., 1989).

pNN50 (%): Relative amount of successive NN interval pairs that differ by more than 50 ms:

$$pNN50 = \frac{NN50}{N-1} x100\% \tag{3.6}$$

3.4.2 Frequency Domain Analysis

Frequency domain methods decompose the total variation of the RR interval series into different frequency components, which can be considered as markers of different physiological effects (Camm et al., 1996; Nattel & Harada, 2014). The frequency bands for short-term HRV analysis and the regulators for these bands are summarized in Table 3.1. For long-term RR interval series ULF band is also introduced whose frequency range is 0-0.003. Since the data used in this thesis are short-term, this band is not used and VLF band includes its frequencies.

Table 3.1 Frequency bands defined for frequency domain HRV analysis and their regulators.

Band	**Frequency Range (Hz)**	**Regulation Mechanism**
Very Low Frequency (VLF)	0-0.04	Temperature & humoral systems (Braga, Lemos, Da Silva, Fontes, & Dos Santos, 2002; Porter & Rivkees, 2001; Williams, Chambers, Henderson, Rashotte, & Overton, 2002) and circadian rhythm (Barrett, Navakatikyan, & Malpas, 2001; Braga et al., 2002)
Low Frequency (LF)	0.04-0.15	Both sympathetic and parasympathetic nervous system (Goldstein et al., 1998; Malpas, 2002)
High Frequency (HF)	0.15-0.4	Parasympathetic (vagal) intervention and respiratory sinus arrhythmia (RSA) (Barbieri, Triedman, & Saul, 2002; Rentero et al., 2002)

Prior to frequency domain analysis, RR interval series must be interpolated and resampled because they are unevenly sampled. Also, detrending should be done to eliminate slowly varying trends from the signal. These preprocessing steps for frequency domain analysis are explained in the following sections.

3.4.2.1 Interpolation

The RR interval time series is an unevenly sampled data as shown in Figure 3.2c. Before spectral analysis, the series must be interpolated and evenly re-sampled. Cubic spline interpolation with 4 Hz sampling rate was applied to series similar to other work in literature (Berntson, 1997; Tarvainen et al., 2014). A short segment of the original and resampled data is given in Figure 3.5.

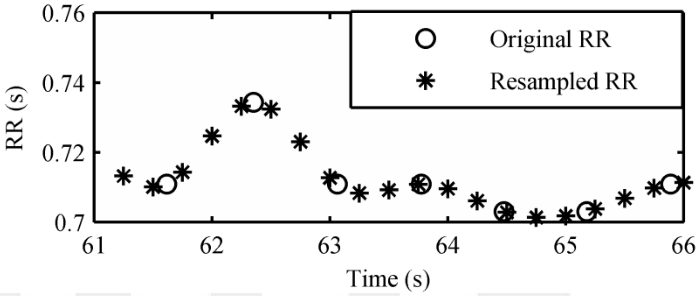


Figure 3.5 Original and resampled RR data. Cubic spline interpolation with 4 Hz sampling frequency was used.

3.4.2.2 Detrending

The RR interval time series are generally obtained from long duration recordings such as 24 hour Holter monitors and often includes baseline trends (Tarvainen, Ranta-Aho, & Karjalainen, 2002). These slowly varying trends must be eliminated in order to avoid any intervention to cardiovascular signals.

An advanced detrending method called smoothness priors detrending was used (Tarvainen et al., 2002). In this method, the signal is considered to have two parts; the first is the stationary signal and the second is the low frequency trend component. The task is to estimate low frequency component by some fitting procedure and get stationary signal by subtracting the trend component from the original signal. This method works as a high pass filter by which the frequency components below a selected frequency are eliminated. The method suggested by Tarvainen et al. is summarized below (Tarvainen et al., 2002).

Let x ∈ R^{N} denote the RR interval time series which includes two parts:

$$x = x_{stat} + x_{trend} \tag{3.7}$$

where x_{stat} is the nearly stationary component and x_{trend} is the low frequency aperiodic trend component. Suppose that the trend component can be modeled with a linear observation model as

$$x_{trend} = H\theta + e \tag{3.8}$$

where $H \in R^{Nxp}$ is the observation matrix, $\theta \in R^p$ are regression parameters, and e is the observation error. Then, the task is to estimate the parameters by some fitting procedure so that $\hat{x}_{trend} = H\hat{\theta}$ can be used as the estimate of trend. Regularized least squares solution is used for the estimation of $\hat{\theta}$

$$\hat{\theta}_{\lambda} = argmin_{\theta}\{\|x - H\theta\|^2 + \lambda\|D_d(H\theta)\|^2\} \tag{3.9}$$

where λ is the regularization parameter and D_d indicates the discrete approximation of d-th derivative operator. This is clearly a modification of the ordinary least squares solution to the direction in which the side norm $\|D_d(H\theta)\|$ gets smaller. In this way, prior information about the predicted trend *H*θ can be incorpoated to the estimation. The solution of Equation (3.9) can be written in the following form

$$\hat{\theta}_{\lambda} = (H^T H + \lambda H^T D_d^T D_d H)^{-1} H^T x \tag{3.10}$$

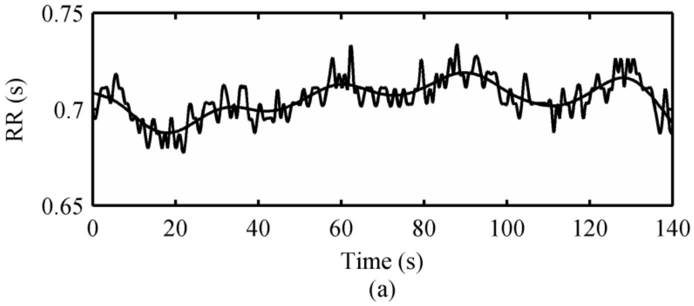
and the estimate of the trend is

$$\hat{x}_{trend} = H\hat{\theta}_{\lambda} \tag{3.11}$$

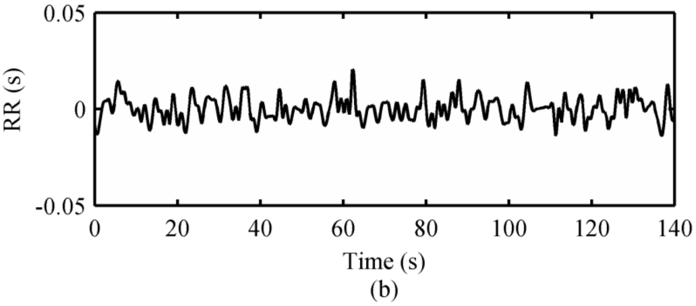
The selection of the observation matrix H can be implemented according to some known properties of the data *x*. Here, identity matrix $H = I \in R^{NxN}$ was used. Also second order difference (D_2) was used for the estimation of the trend. With these choices, the detrended nearly stationary RR series can be written as

$$\hat{x}_{stat} = x - H\hat{\theta}_{\lambda} = (1 - (1 + \lambda D_2^T D_2)^{-1})x \tag{3.12}$$

The regularization parameter λ was selected 500, which corresponds to a cutoff frequency of 0.035 Hz and preserves the information in VLF band (Tarvainen et al., 2002). An RR interval time series is shown in Figure 3.6 before and after detrending operation with selected parameters. The effect of detrending on frequency domain analysis can be seen in Figure 3.7. FFT spectrum of the nondetrended signal is given in Figure 3.7a and detrended signal with smoothness prior detrending method in Figure 3.7b. The power of the low frequency band decreases significantly which affect many frequency domain measures.

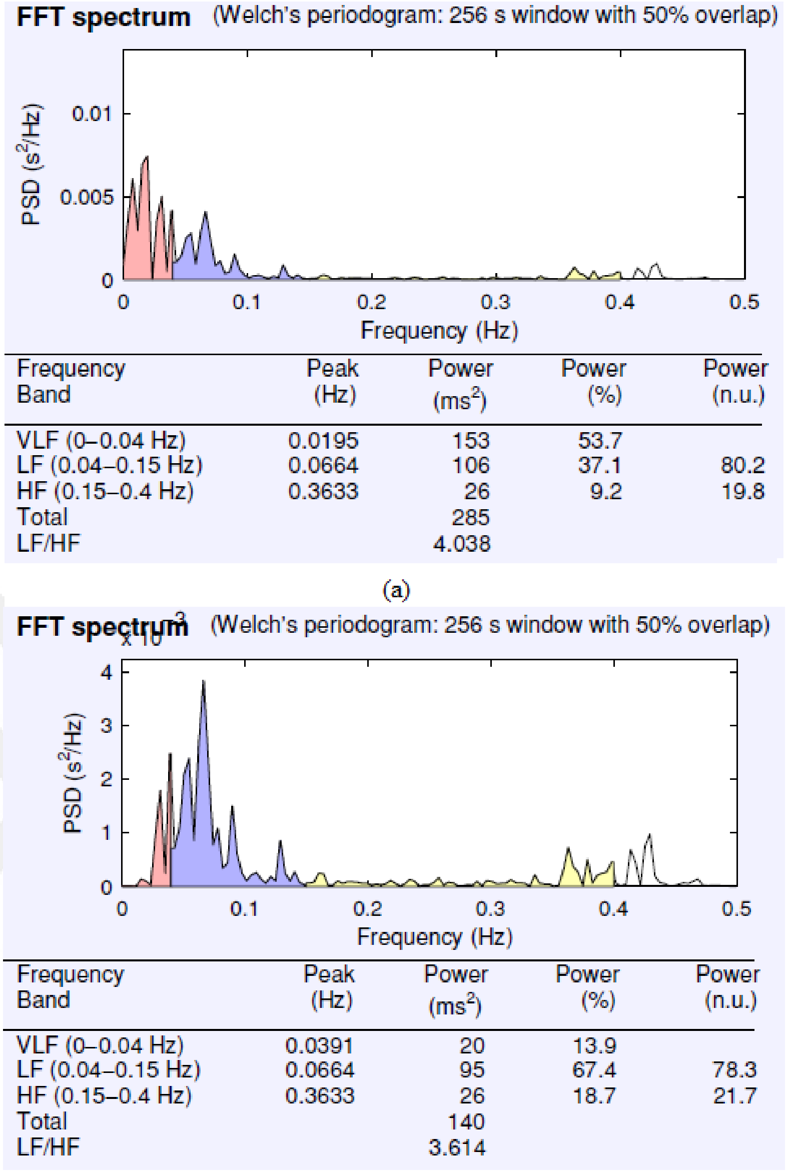


(a)



(b)

Figure 3.6 Original and (b) detrended RR interval time series with smoothness priors detrending method with λ=500.



Frequency Band	Peak (Hz)	Power (ms^2)	Power (%)	Power (n.u.)
VLF (0–0.04 Hz)	0.0195	153	53.7	
LF (0.04–0.15 Hz)	0.0664	106	37.1	80.2
HF (0.15–0.4 Hz)	0.3633	26	9.2	19.8
Total		285		
LF/HF		4.038		

(a)

Frequency Band	Peak (Hz)	Power (ms^2)	Power (%)	Power (n.u.)
VLF (0–0.04 Hz)	0.0391	20	13.9	
LF (0.04–0.15 Hz)	0.0664	95	67.4	78.3
HF (0.15–0.4 Hz)	0.3633	26	18.7	21.7
Total		140		
LF/HF		3.614		

(b)

Figure 3.7 Comparison of the spectrums of (a) nondetrended and (b) detrended signals.

3.4.2.3 Derivation of Frequency Domain HRV Features

There are two possible methods for spectrum calculation: Fast Fourier Transform (FFT) based Welch's periodogram and autoregressive (AR) modeling based method. However, there is not an exact rule for model order selection in AR. The main problems with AR model order selection are that only one AR component will result in each frequency band and, secondly, negative power values can result in closely spaced AR components (Tarvainen & Niskanen, 2008). The spectrums of an RR interval series obtained with two methods are given in Figure 3.8. When FFT and AR spectrums are compared, it can be seen that AR spectrum is smoother. Increasing the model order in AR model affects mostly VLF and LF bands.

Power spectrum density (PSD) estimates of the RR interval series were calculated by using the FFT based Welch's periodogram (Tarvainen et al., 2002). In using this method, the RR interval series were divided into 50% overlapping windows each with 256 samples and then the FFT was calculated for each window. The final spectrum was found by averaging the spectrums of the individual windows. The variance of the FFT spectrum was decreased by this averaging effect. Absolute powers of each frequency band were calculated by integrating the spectrum over the specified band limits. Frequency domain features used in this study are:

VLF peak (Hz): VLF band peak frequency.

LF peak (Hz): LF band peak frequency.

HF peak (Hz): HF band peak frequency.

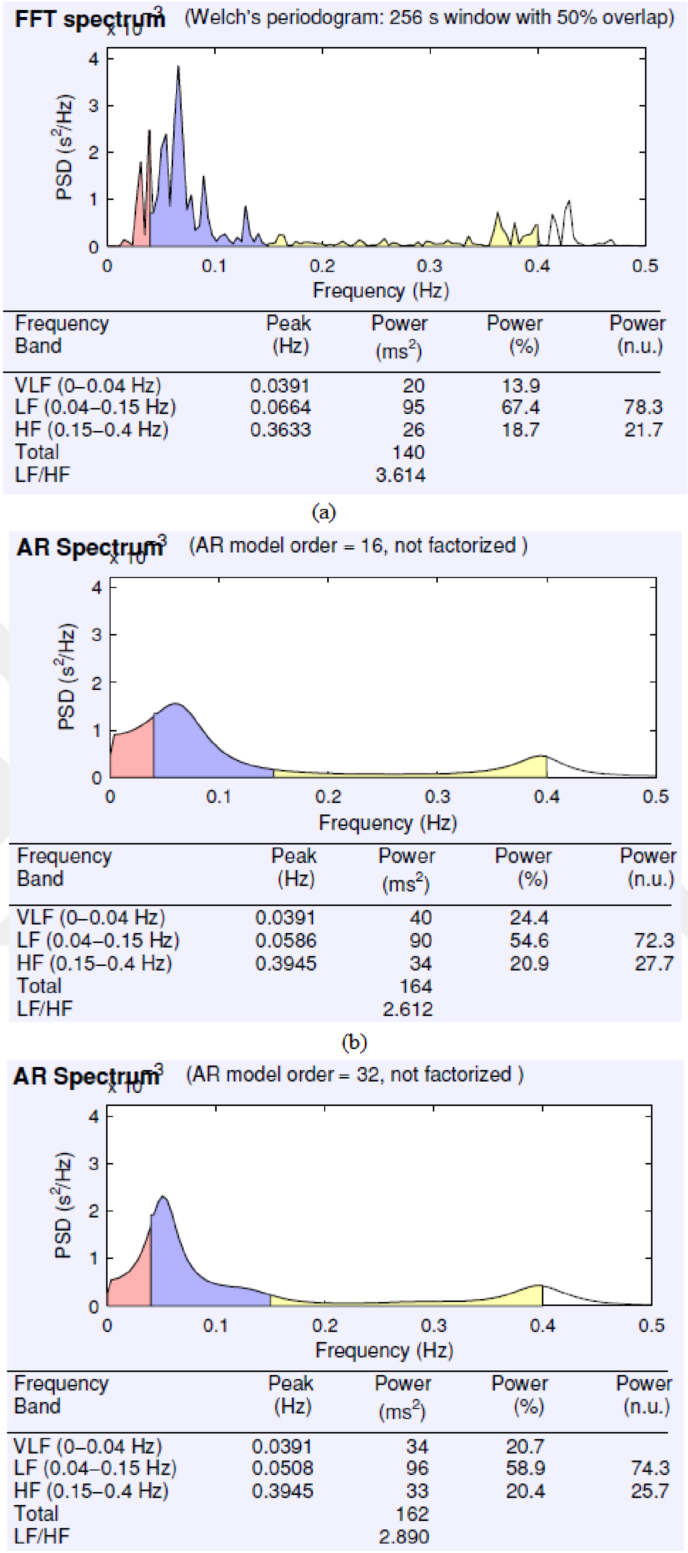
VLF power (ms^2): Absolute power of VLF band.

LF power (ms^2): Absolute power of LF band.

HF power (ms^2): Absolute power of HF band.

VLF power prc (%): Relative power of VLF band:

$$VLF\ power\ prc = \frac{VLF\ power}{Total\ power} x100\% \tag{3.13}$$



FFT spectrum (Welch's periodogram: 256 s window with 50% overlap)

Frequency Band	Peak (Hz)	Power (ms^2)	Power (%)	Power (n.u.)
VLF (0–0.04 Hz)	0.0391	20	13.9	
LF (0.04–0.15 Hz)	0.0664	95	67.4	78.3
HF (0.15–0.4 Hz)	0.3633	26	18.7	21.7
Total		140		
LF/HF		3.614		

(a)

AR Spectrum (AR model order = 16, not factorized)

Frequency Band	Peak (Hz)	Power (ms^2)	Power (%)	Power (n.u.)
VLF (0–0.04 Hz)	0.0391	40	24.4	
LF (0.04–0.15 Hz)	0.0586	90	54.6	72.3
HF (0.15–0.4 Hz)	0.3945	34	20.9	27.7
Total		164		
LF/HF		2.612		

(b)

AR Spectrum (AR model order = 32, not factorized)

Frequency Band	Peak (Hz)	Power (ms^2)	Power (%)	Power (n.u.)
VLF (0–0.04 Hz)	0.0391	34	20.7	
LF (0.04–0.15 Hz)	0.0508	96	58.9	74.3
HF (0.15–0.4 Hz)	0.3945	33	20.4	25.7
Total		162		
LF/HF		2.890		

(c)

Figure 3.8 Comparison of FFT and AR methods. (a) The spectrum obtained with FFT based Welch's periodogram (b) The spectrum obtained with AR method with order 16. (c) The spectrum obtained with AR method with order 32.

LF power prc (%): Relative power of LF band:

$$LF\ power\ prc = \frac{LF\ power}{Total\ power} x100\% \tag{3.14}$$

HF power prc (%): Relative power of HF band.

$$HF\ power\ prc = \frac{HF\ power}{Total\ power} x100\% \tag{3.15}$$

LF power norm (n.u): Power of LF band in normalized units.

$$LF\ power\ norm = \frac{LF\ power}{Total\ power - VLF\ power} x100\% \tag{3.16}$$

HF power norm (n.u.): Power of HF band in normalized units.

$$HF\ power\ norm = \frac{HF\ power}{Total\ power - VLF\ power} x100\% \tag{3.17}$$

LF/HF power (n.u.): Ratio of LF and HF power bands.

$$LF/HF = \frac{LF\ power}{HF\ power} x100\% \tag{3.18}$$

3.4.3 Nonlinear Analysis

Heart is under the control of autonomic nervous system which has a nonlinear nature and analyzing the nonlinear properties of the RR intervals may thus reveal some information about the complex and inherently nonlinear nature of these physiological mechanisms (Huikuri, Mäkikallio, & Perkiömäki, 2003). The number of studies using nonlinear HRV analysis is increasing (Alcaraz, Abásolo, Hornero, & Rieta, 2010; Costa, Goldberger, & Peng, 2002, 2005; Goldberger, Amaral, Glass, & Hausdorff, 2000; Hu, Ivanov, Chen, Carpena, & Stanley, 2001; Huikuri et al., 2003; Kamen, Krum, & Tonkin, 1996; Kantelhardt, Koscielny-Bunde, Rego, Havlin, &

Bunde, 2001; Karmakar, Khandoker, Gubbi, & Palaniswami, 2009; Martínez et al., 2014; Perkiömäki, 2011; Yeragani et al., 1998). However, the difficulty in nonlinear analysis is the physiological interpretation of the results. Nonlinear features used in this study are given below.

Approximate Entropy (ApEn): Approximate entropy measures the complexity or irregularity of the signal (Richman & Moorman, 2000). ApEn of a time series is a nonnegative number whose large values indicate high irregularity or complexity (Pincus, 1991). For an RR interval series with length N and an embedding space of R^m , ApEn is calculated by (Acharya et al., 2006):

$$ApEn(m,r,N) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} logC_i^m(r) - \frac{1}{N-m}\sum_{i=1}^{N-m} logC_i^{m+1}(r) \tag{3.19}$$

where $C_i^m(r) = \frac{1}{N-m+1}\sum_{j=1}^{N-m+1} \Theta\left(r - \left\|x_i - x_j\right\|\right)$ is the correlation integral, 'r' is the tolerance value and 'm' is the embedding dimension. 'r' has a strong effect on ApEn and should be selected as a fraction of the standard deviation of the data to enable the comparison of different data sets. In this thesis, 'm' was set as 2 and 'r' was selected as $0.2xSDNN$ based on the studies of Pincus (Pincus & Goldberger, 1994).

Sample Entropy (SampEn): Sample entropy is similar to ApEn but it does not include self-similarity patterns as ApEn and shows advantages such as data length independence (Richman & Moorman, 2000). In ApEn, the comparison of template vector with other vectors also includes the comparison with itself which guarantees that probabilities $C_i^m(r)$ are always nonzero and allows taking logarithms of probabilities. This lowers the ApEn values and the signals are considered more regular than as they are (Acharya et al., 2006). For SampEn, the vector comparison with itself is removed

$$C_i'^m(r) = \left\{ \begin{matrix} the\ number\ of\ j, j \neq i, j \leq N-m+1, \\ such\ that\ d[u(i),u(j)] \leq r \end{matrix} \right\} / (N-m+1) \tag{3.20}$$

where $u(j) = (RR_j, RR_{j+1}, \dots, RR_{j+m-1})$ and j = 1, 2, …, N-m+1.

Then, $\varphi'^{m}(r)$ can be defined as

$$\varphi'^{m}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i'^{m}(r) \tag{3.21}$$

and finally

$$SampEn(m, r, N) = -ln \left[\varphi'^{m}(r) \Big/ \varphi'^{m+1}(r) \right] \tag{3.22}$$

The parameters to be determined for SampEn are similar to ApEN. The embedding dimension *'m'* was again selected as 2 and the tolerance value *'r'* was selected as *0.2xSDNN* similar to ApEn as suggested in (Alcaraz et al., 2010).

Detrended Fluctuation Analysis (DFA): Detrended fluctuation analysis quantifies the fractal scaling properties of short-term RR intervals series. This method measures the root-mean-square fluctuation of an integrated and detrended time series and plots against the size of the observation window on a log-log scale (Acharya et al., 2006). Firstly, the RR interval series (with length N) is integrated using

$$y(k) = \sum_{i=1}^{k} [RR_i - RR_{av}] \tag{3.23}$$

where $y(k)$ is the k th value of the integrated series, RR_i is the i th RR interval and RR_{av} is the overall average of the entire RR interval series. Then, the integrated series is divided into windows with equal length, n. In each window, a least-squares line is fitted to the data which is denoted by $y_n(k)$. Then, the integrated series $y(k)$ is detrended by subtracting the local trend within each segment and the root-mean-square fluctuation of this integrated and detrended time series is calculated by

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y(k) - y_n(k)\right)^2} \tag{3.24}$$

This computation is repeated for different segment lengths and gives the index *F(n)* as a function of segment length *n*. Then, (*n*, *F(n)*) pairs are plotted on log-log scale (see Figure 3.9).

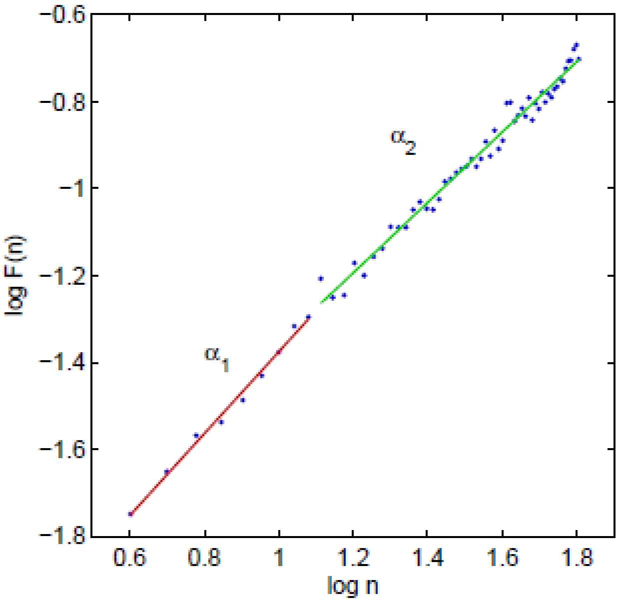


Figure 3.9 An example plot as the output of DFA. α_1 is the short term and α_2 is the long term fluctuation (Tarvainen & Niskanen, 2008).

A linear relationship on the graph indicates the fractal scaling and the fluctuations can be characterized by the scaling exponent α, which is the slope of the regression line relating log *F(n)* to log *n*. Typically, in DFA the correlations are divided into short-term (α_1) and long-term (α_2) fluctuations. In this thesis, α_1 was obtained from the range $4 \leq n \leq 16$ and α_2 was obtained from the range $16 \leq n \leq 64$ as default (Peng, Havlin, Stanley, & Goldberger, 1995).

Correlation Dimension (CorDim): Correlation dimension is another method for measuring the complexity or strangeness of a time series. It is expected to give information on the minimum number of dynamic variables needed to model the underlying system (Grassberger & Procaccia, 1983).

In order to calculate CorDim, the vectors u_j with length m were formed and the number of vectors u_k for which $d(u_j, u_k) \leq r$ were calculated similar to ApEn and SampEn. However, this time the distance function is defined as

$$d\left(u_j, u_k\right) = \sqrt{\sum_{l=1}^{m} \left(u_j(l) - u_k(l)\right)^2} \tag{3.25}$$

Then, an average of the term $C_j^m(r)$ is taken as

$$C^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} C_j^m(r) \tag{3.26}$$

which is called the correlation integral. Then, the correlation dimension (D_2) is defined as the limit value

$$D_2(m) = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C^m(r)}{\log \quad r} \tag{3.27}$$

In practice this limit value is approximated by the slope of the regression curve (log r, log$C^m(r)$) as in Figure 3.10 (Henry, Lovell, & Camacho, 2012). The slope is calculated from the linear part of the log-log plot. In this thesis, a default value of m = 10 was selected for the embedding (Tarvainen et al., 2014).

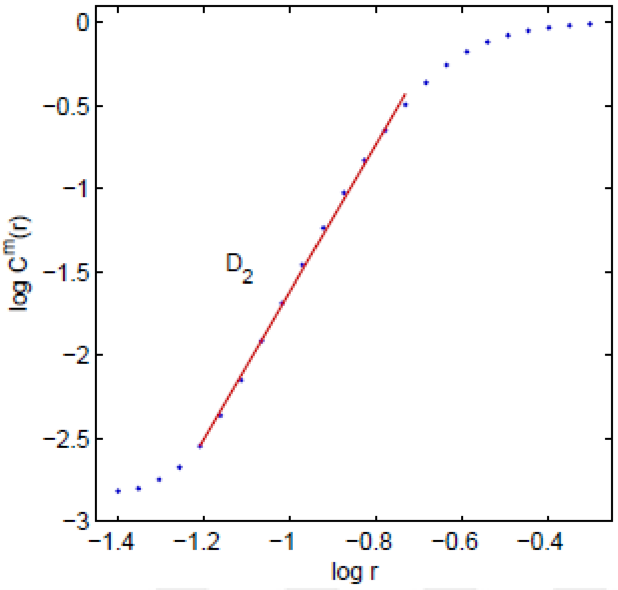


Figure 3.10 An approximation of the correlation dimension D_2 (Tarvainen & Niskanen, 2008).

Recurrence Plot Analysis (RPA): Recurrence plot is a visualization of a square matrix, in which the matrix elements correspond to those times at which a state of a dynamical system recurs. RPA reveals all the times when the phase space trajectory of the dynamical system visits roughly the same area in the phase space (Tarvainen et al., 2014).

To construct a recurrence plot of an RR series, vectors

$$u_j = (RR_j, RR_{j+1}, \dots, RR_{j+(m-1)\tau}) \qquad j=1, 2, \dots, N-(m-1)\tau \tag{3.28}$$

are formed. Here m is the embedding dimension and τ is the embedding lag. Then vectors u_j represent the RR interval series as a trajectory in m dimensional space. A recurrence plot is a symmetrical square matrix of ones and zeros. The element in j'th row and k'th column, i.e. RP(j,k) is 1 if the point u_j on the trajectory is close to point u_k:

$$RP(j,k) = \begin{cases} 1, & d(u_j - u_k) \le r \\ 0, & otherwise \end{cases} \tag{3.29}$$

where $d(u_j, u_k)$ is the Euclidean distance, and r is a fixed threshold. An example RP matrix is given in Figure 3.11. The structure of the RP matrix is symmetric and consists short line segments of ones parallel to the main diagonal which are represented in black in RP plot (zeros are represented in white). The length of these diagonal lines describe the duration of the two points which are close to each other (Tarvainen et al., 2014).

In this thesis the embedding dimension m was set as 10 and the threshold distance r was selected as $\sqrt{m}$xSDNN and τ was fixed to 1 similar to (Dabiré, Mestivier, Jarnet, Safar, & Chau, 1998).

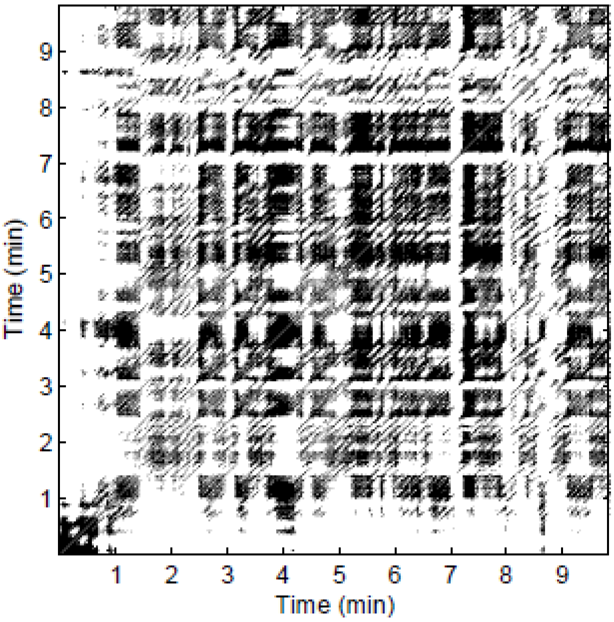


Figure 3.11 An example RP matrix of an RR series (Tarvainen & Niskanen, 2008).

The quantification methods of recurrence plots were proposed by Webber & Zbilut, which are recurrence rate (REC), maximum diagonal line length (l_{max}), average diagonal line length (l_{mean}), determinism (DET) and Shannon entropy (ShanEn) (Webber & Zbilut, 1994).

Recurrence rate is the ratio of ones and zeros in the RP matrix. The number of elements in RP matrix for τ=1 is N-m+1 and REC is

$$\mathrm{REC} = \frac{1}{N-m+1} \sum_{j,k=1}^{N-m+1} RP(j,k) \tag{3.30}$$

Average diagonal line length l_{mean} is obtained as

$$l_{mean} = \frac{\sum_{l=l_{min}}^{l_{max}} lN_l}{\sum_{l=l_{min}}^{l_{max}} N_l} \tag{3.31}$$

where N_l is the number of lines with length l.

The determinism (DET) is calculated by

$$\mathrm{DET} = \frac{\sum_{l=l_{min}}^{l_{max}} lN_l}{\sum_{j,k=1}^{N-m+1} RP(j,k)} \tag{3.32}$$

The Shannon information entropy (ShanEn) of the line length distribution is defined as

$$\mathrm{ShanEn} = -\sum_{l=l_{min}}^{l_{max}} n_l \ln n_l \tag{3.33}$$

Poincare Plot: Poincare plot is a method widely used for visually representing the heart rate variability obtained from ECG records (Brennan, Palaniswami, & Kamen, 2001). It shows the correlation between successive RR intervals as a plot of $\mathrm{RR}_{j+\tau}$ as a function of RR_j where τ is the time lag. An important issue when plotting Poincare plot is to determine time lag (τ). There must be a balance between the values of τ such as the coordinates RR_j and $\mathrm{RR}_{\mathrm{j}+\tau}$ will not be independent enough if τ is too short and will be random with respect to each other due nonlinear nature of the system if τ is too large. Shannon's mutual information can be used to determine the nonlinear dependence between the coordinates (Shannon & Weaver, 1949). Suppose that there are two measurements from the same system, which are a[n] and b[n]. Then, the amount of information (in bits) about measurement a[n] that is acquired by observing b[n] is

$$\log_2\left[\frac{\mathrm{P}(a,b)}{\mathrm{P}(a)\mathrm{P}(b)}\right] \tag{3.34}$$

In present case, a[n] and b[n] are the time delayed versions of the same time RR interval series. Average mutual information is the average of mutual information over all measurements:

$$\mathrm{I}(\tau) = \sum_{s(n),s(n+\tau)} \mathrm{P}\big(s(n),s(n+\tau)\big)\log_2\left[\frac{\mathrm{P}(s(n),s(n+\tau))}{\mathrm{P}(s(n))\mathrm{P}(s(n+\tau))}\right] \tag{3.35}$$

Then, the suitable time lag is the point where the lag versus mutual information graph first falls to its minimum (Abarbanel & Gollub, 1996). This value was found as 1 as shown in Figure 3.12 and then τ was fixed to 1 in Poincare plot analysis.

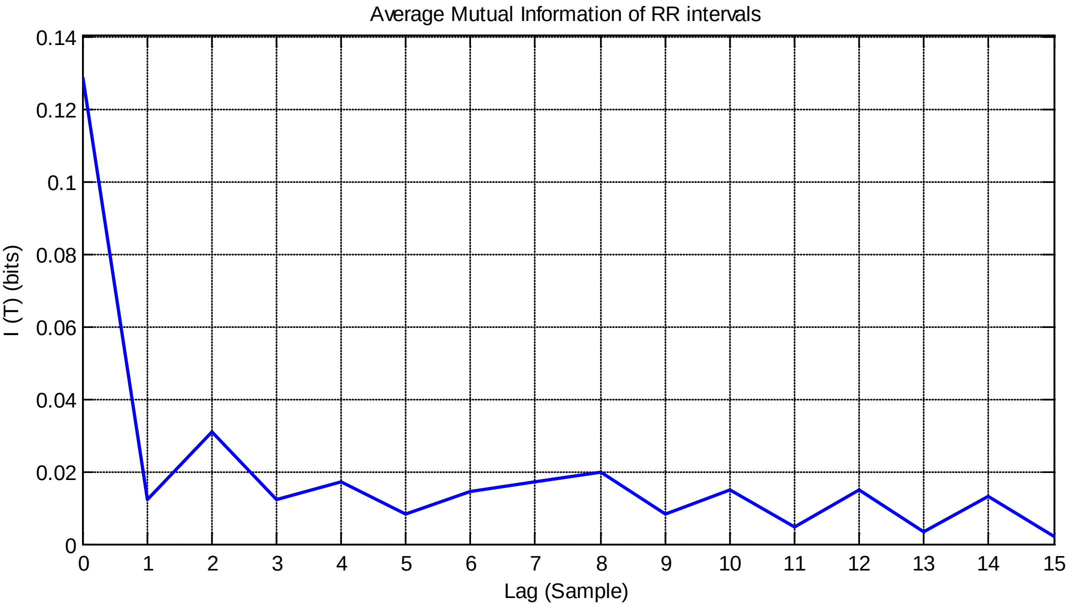


Figure 3.12 Lag versus mutual information plot of an RR interval series. The graph first falls to its minimum is lag 1.

An example Poincare plot of an RR interval series with lag 1 is given in Figure 3.13. The shape of the distribution of the data points is the essential feature. In order to quantify the geometry, an ellipse, which is oriented along the line of identity (RR_j=RR_{j+1}), is fitted to the plot. Then, the dispersion of the data points perpendicular to the line of identity (x_1 in Figure 3.13) is called SD1, which describes the short-term variability, and the dispersion of the data points along the line of identity (x_2 in Figure 3.13) is called SD2, describes the long-term variability (Brennan et al., 2001). Even SD1 nad SD2 are given as nonlinear HRV measures, it was shown that they can be expressed as the functions of time domain measures SDSD (standard deviation of successive differences) and SDRR as in Equations (3.36) and (3.37) (Brennan et al., 2001).

$$\mathrm{SD1}^2 = \frac{1}{2}\mathrm{SDSD}^2 \tag{3.36}$$

$$\mathrm{SD2}^2 = 2\mathrm{SDRR}^2 - \frac{1}{2}\mathrm{SDSD}^2 \tag{3.37}$$

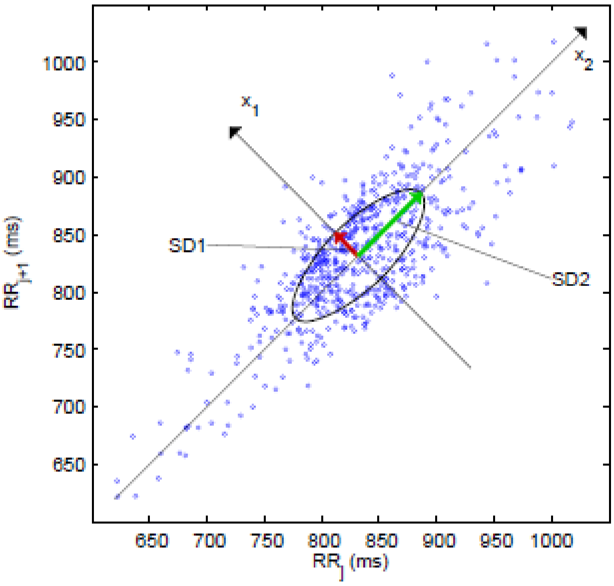


Figure 3.13 Poincare plot of an RR interval series with lag 1. An ellipse is fitted on the data points. SD1 describes short-term variability and SD2 describes long-term variability (Tarvainen & Niskanen, 2008).

Some other work also showed that SD1 is highly correlated with RMSSD and LF power while SD2 is highly correlated with SD and LF/HF power (Carrasco MJ Gaitán, 2001; Guzik et al., 2007).

Complex Correlation Measure (CCM): Standard descriptors of Poincare plot (SD1, SD2) measure only the gross variability of the time series data. In other words, it does not include temporal information. Thus, different RR interval series could have the same Poincare indices. To overcome this shortcoming and include temporal variations of the time series, a new measure called as complex correlation measure (CCM) was introduced (Karmakar et al., 2009). In contrast to SD1 and SD2, CCM is interested in point-to-point variation in the time series and thus includes temporal information. CCM is calculated by choosing a moving window of three consecutive points ((RR_k,RR_{k+1}), (RR_{k+1},RR_{k+2}), (RR_{k+2},RR_{k+3})) from the Poincare plot. Then the area of the i^{th} triangle A(i) formed by these three points is computed (see Figure 3.14) and repeated for all points in the series. If the Poincare plot is composed of N points with lag τ, then the temporal variation of the plot is composed of all overlapping three point windows and can be calculated by (Karmakar et al., 2009)

$$CCM(\tau) = \frac{1}{C_n(N-2)} \sum_{i=1}^{N-2} A(i) \tag{3.38}$$

where τ represents the lag of Poincare plot and C_n is the normalization constant. C_n is calculated by the formula $C_n = \pi \times SD1 \times SD2$ which actually represents the area of the fitted ellipse on the Poincare plot. Therefore, CCM is a multiple lag correlation of a time series. In this work, the lag was selected as 1 similar to Poincare plot.

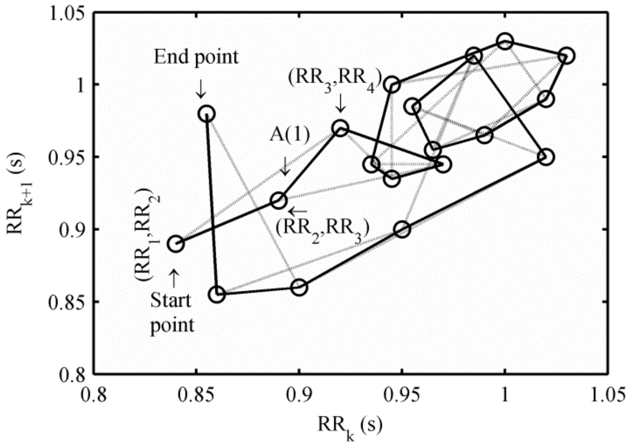


Figure 3.14 Construction of triangles for the calculation of CCM.

CHAPTER FOUR
FEATURE SELECTION WITH GENETIC ALGORITHMS

Real word data is generally high dimensional. Dimensionality reduction is the transformation of high dimensional data into a meaningful representation of reduced dimensionality. Ideally, the reduced representation has a dimensionality that corresponds to the intrinsic dimensionality of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data (Fukunaga, 2013). Dimensionality reduction is important in many domains, since it facilitates classification, visualization, and compression of high-dimensional data, by mitigating the curse of dimensionality and other undesired properties of high dimensional spaces (Jimenez & Landgrebe, 1998).

For each problem with some sample, there is a maximum number of features where performance degrades instead of improves, which is called the curse of dimensionality (see Figure 4.1). When the dimensionality increases, the volume of the space increases so fast that the available data become sparse. Also, the number of learning data should grow exponentially with the dimension. For example, if 10 data are reasonable to learn a 1-dimensional model, 100 are necessary for learning a 2-dimension and 1000 are necessary for learning a 3-dimensional model (Verleysen & François, 2005). An accurate mapping of lower-dimensional space of features is needed so no information is lost by discarding important and basic features (Hassanien, Kim, Kacprzyk, & Awad, 2014).

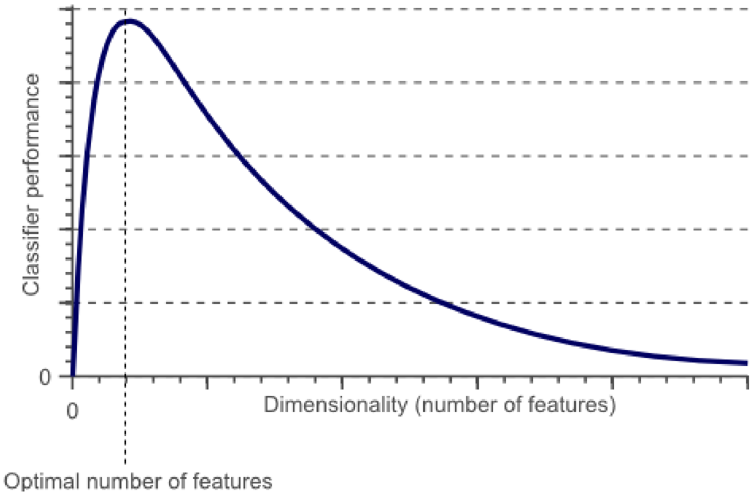


Figure 4.1 Increasing the dimension over optimal number of features degrades classification performance due to curse of dimensionality (Spruyt, 2014).

Benefits of the feature selection are (Guyon & Elisseeff, 2003)

- Facilitating data visualization and data understanding
- Reducing the measurement and storage requirements
- Reducing training and utilization times
- Defying the curse of dimensionality to improve prediction performance

There are two methods for dimensionality reduction: feature selection and feature extraction. Feature selection approaches try to find an optimal subset of original features whereas feature extraction transforms the features from a high dimensional space to a space of fewer dimensions.

Feature selection methods are broadly classified into filter, wrapper and embedded methods. Filter methods act as preprocessing to rank the features where the highly ranked features are selected and applied to a predictor. In wrapper methods, the feature selection criterion is the performance of the predictor, i.e. the predictor is wrapped on a search algorithm which will find a subset which gives the highest predictor performance. Embedding methods include variable selection as part of the training process without splitting the data into training and testing sets (Chandrashekar & Sahin, 2014).

Sequential forward/backward selection, exhaustive search, genetic algorithms, branch and bound methods are commonly used wrapper methods. Advantages of these methods are their implementation does not require much knowledge about the structural properties of the problem and algorithms can be easily coded.

Genetic algorithms (GA) is an optimization algorithm founded upon the principles of natural evolution discovered by Darwin (Pernkopf, 2007). In nature, individuals have to adapt to their environment in order to survive in a process of further development. It uses a stochastic, directed and highly parallel search based on principles of population genetics that artificially evolve solutions to a given problem (Goldberg, 1989; Holland, 1975).

GAs are stochastic optimization procedures which have been successfully applied in many feature selection tasks. The problem of dimensionality reduction is well suited to formulation as an optimization problem. Given a set of d-dimensional input patterns, the task of the GA is to find a transformed set of patterns in an m-dimensional space ($m < d$) that maximizes a set of optimization criteria (Raymer, Punch, Goodman, Kuhn, & Jain, 2000). Typically, the transformed patterns are evaluated based upon their dimensionality, and their class separation or the classification accuracy that can be obtained on the patterns with a given classifier.

The advantages of using genetic algorithms are (Goldberg, 1989)

- They work with a coding of the parameter set, not the parameters themselves
- They search from a population of points, not a single point
- They use payoff (objective function) information, not derivatives or other auxiliary knowledge
- They use probabilistic transition rules, not deterministic rules.

Genetic algorithm was used in this thesis because it is a biologically inspired algorithm, not trapped in local optimal solution with crossover and mutation and easy to implement.

Main elements of a GA mechanism, used in the feature selection, will be given in the following.

4.1 Representation of Individuals

A direct approach to using GA for feature selection was introduced by Siedlecki & Sklansky (Siedlecki & Sklansky, 1989). In their work, a GA is used to find an optimal binary vector, where each bit represents a feature (Figure 4.2). If the *i*-th bit of this vector equals 1, then the *i*-th feature is used in classification; otherwise, the corresponding feature is not used in classification. A set of binary vectors, which is called initial population, is constructed as the first step.

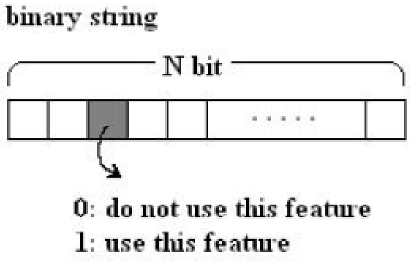


Figure 4.2 Representation of N features with a binary string (Siedlecki & Sklansky, 1989).

4.2 Fitness

The evolutionary process is driven by the fitness measure. The fitness function assigns a value to each string in the population according to some success criteria. For feature selection, the criterion is the classification accuracy or error rate.

4.3 Selection

The selection operation improves the average quality of the population by giving a higher probability to individuals with higher fitness to undertake any genetic operation. An important feature of the selection mechanism is its independence of the representation scheme, as only the fitness is taken into account. The probabilistic feature allocates to every individual a chance of being selected occasionally.

The most popular of the stochastic selection strategies is fitness proportionate selection, also called biased roulette wheel selection. It can be regarded as allocating pie slices on a roulette wheel, with each slice proportional to a string's fitness as in Figure 4.3. Selection of a string to be a parent can then be viewed as a spin of the wheel, with the winning slice being the one where the spin ends up. Although this is a random procedure, the chance of a string to be selected is directly proportional to its fitness and the least fit individuals will gradually be driven out of a population.

A second common strategy is called tournament selection (Goldberg & Deb, 1991). A subpopulation of individuals is chosen at random. The individual from this subpopulation with the highest fitness wins the tournament. Generally, tournaments are held between two individuals (binary tournament). However, this can be generalized to an arbitrary group whose size is called the tournament size. This algorithm can be implemented efficiently as no sorting of the population is required. More important, it guarantees diversity of the population. The most important feature of this selection scheme is that it does not use the value of the fitness function. It is only necessary to determine whether an individual is fitter than any other or not.

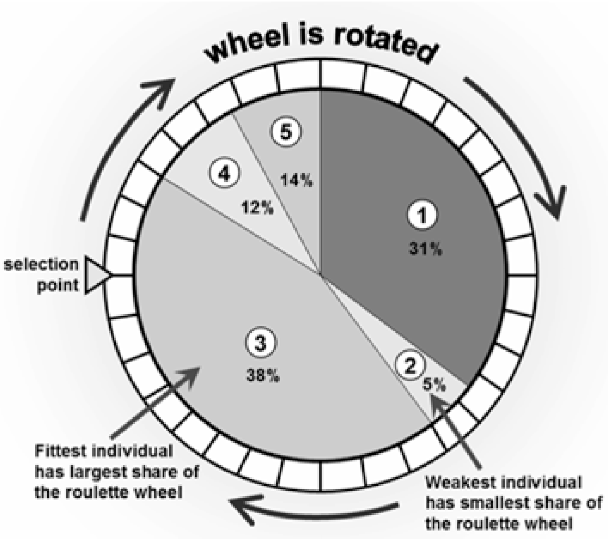


Figure 4.3 Roulette wheel selection. The chance of selecting the individual with higher fittness is also high which means fittest individuals continue genetic operations.

4.4 Crossover

The crossover operator is responsible for combining good information from two strings and for testing new points in the search space. The two off-springs are composed entirely of the genetic material from their two parents. By recombining randomly certain effective parts of a character string, there is a good chance of obtaining an even more fit string and making progress toward solving the optimization problem.

Several ways of performing crossover can be used. The simplest but very effective is the one-point crossover (Goldberg, 1989). Two individual strings are selected at random from the population. Next, a crossover point is selected at random along the string length, and two new strings are generated by exchanging the substrings that come after the crossover point in both parents. The mechanism is illustrated in Figure 4.4.

A more general case is the multi-point crossover in which parts of the information from the two parents are swapped among more string segments (de Jong, 1975). An example is the two-point crossover, where two crossover points are selected at random and the substring lying in between the points are swapped.

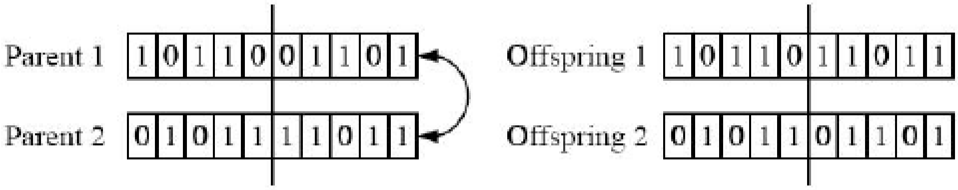


Figure 4.4 One point crossover. The bits after crossover point are exchanged between parents to create offsprings to jump new points in searchspace.

In uniform crossover, each bit of the offspring is created by copying the corresponding bit from one or the other parent selected at random with equal probability (Syswerda, 1989). Uniform crossover has the advantage that the ordering of bits is entirely irrelevant because there is no linkage between adjacent bits. Multi-point crossover takes half of the material from each parent in alternation, while uniform crossover decides independently which parent to choose. When the population has largely converged, the exchange between two similar parents leads to a very similar offspring. This is less likely to happen with uniform crossover particularly with small population size, and so, gives more robust performance.

4.5 Mutation

Mutation prevents the population from premature convergence or from having multiple copies of the same string. This feature refers to the phenomenon in which the GA loses population diversity because an individual that does not represent the global optimum becomes dominant. In such cases the algorithm would be unable to explore the possibility of a better solution.

Mutation consists of the random alteration of a string with low probability. It is implemented by randomly selecting a string location and changing its value from 0 to 1 or vice versa, as shown in Figure 4.5.

Figure 4.5 Mutation operation. The randomly selected bits are complemented.

The implementation of GA for feature selection in this thesis is as follows:

1. *Representation:* Initial population was constructed with 20 strings. Each string consisted of 33 bit representing 33 features. The bit values were determined by simulating a random coin flip. Bit 1's represented selected features and bit 0's not selected features.

2. *Fitness:* One nearest neighbor classification of PAF and non-PAF subjects was done with the selected features in each string. The fitness value was selected as the average of sensitivity and specificity of the classification which corresponds to the area under ROC curve.

3. *Selection:* According to fitness values, the parents for the next generation were selected with roulette wheel selection. The chance of selecting a string with higher fitness value was also higher with this selection.

4. C*rossover:* Crossover point was determined randomly and the bits after the crossover point were exchanged between parents.

5. *Mutation:* Mutation rate was selected as 0.01. The location of mutate bits were selected randomly and those bits were complemented.

6. *Stop criteria:* One nearest neighbor classification was done with selected features. The stop criterion was checked, which was the change in highest fitness value. If it does not change more than 0.1, the algorithm stops. The selected features of the individual with highest fitness value were taken to

represent our data. If the fitness value changed more than 0.1, go step 3 and continue until stop criteria was achieved.

The algorithm was run several times and it is observed that the same features yielded the best fitness value in each run. Eight features selected by GA are mean value and standard deviation of RR intervals, HF band peak frequency, relative powers of LF and HF bands, dispersion of points perpendicular to line of identity in Poincare plot (SD1), sample entropy, short term fluctuation slope of detrended fluctuation analysis (α1).

CHAPTER FIVE
METHODS

Paroxysmal atrial fibrillation patients are diagnosed when an AF episode is detected in ECG record. However, PAF is a rhythm disorder that starts and terminates spontaneously in a short time. Then, it is difficult to get an ECG record taken during a PAF episode in routine physical examinations in a healthcare facility. Using a Holter or event recorder 24 h or more might be needed for diagnosis which is uncomfortable for the patient and an economic burden on the system (Cowan et al., 2014). The method proposed in this thesis uses ECG records taken at least 45 min away from any PAF episode to diagnose PAF patients. The records were preprocessed as explained in Chapter 3 and 31 HRV features were obtained. Atrial and ventricular ectopic beat numbers were also used as separate features and the dimension of the feature space was 33. Feature selection was done with genetic algorithm as explained in Chapter 4 and eight features were selected. Then, classification of PAF and non-PAF subjects was done using 4 different classifiers by utilizing both whole feature set and selected feature set as given in detail in the following sections.

The work is implemented in the software environment of MATLAB version 2008b. All the study was conducted using an Intel Core Duo-2.2 GHz computer with 2 GB DDR2 memory.

5.1 Data Acqusition

PhysioNet is a freely available web-based resource providing a wide range of physiologic signals and related open-source software to the researchers (Goldberger et al., 2000). The PAF Prediction Challenge Database (afpdb) from Physionet was used in this study. This database includes two-channel ECG recordings which have been created for use in the Computers in Cardiology Challenge 2001, where the goal of the challenge was developing automated methods for predicting PAF (Moody et al., 2001). ECG signals were sampled at 128 Hz and digitized with 12 bit resolution.

In addition, the database also includes automatically generated QRS occurance times of each record. However, these timings are not confirmed by the experts and may include ectopic beats or wrong detections (Moody et al., 2001). The RR interval series derived from these QRS occurance times were used in this study. The advantage of using these unaudited timings is reducing the sensitivity of the system to wrong QRS detections in real applications.

The database has 100 ECG record sets obtained from 98 different subjects. Each record set contains two 30 min records from the same subject. All records were extracted from longer ECG records. The distribution of the dataset is shown in Figure 5.1. Among these 100 record sets, 53 come from subjects who have previously been diagnosed with PAF. For these record sets, one record of the set is obtained just before a PAF attack (from here on will be called as prior-to-PAF records); the other record is obtained from ECG periods which are at least 45 min away from any PAF episode (from here on will be called as distant-from-PAF records). The remaining 47 record sets come from subjects who do not have any PAF history (from here on will be called as non-PAF records). These non-PAF subjects include healthy controls, patients referred for long term ambulatory ECG monitoring and patients in intensive care units without any PAF activity. In this study, only distant-from-PAF and non-PAF records were used; the prior-to-PAF records were not used. The reason for this approach is to make the performance evaluation of our systems more realistic. If the prior-to-PAF records had also been included, this would have increased the performance of the classifiers. However, this would be an erroneous bias because, in real life, it is almost impossible to get an ECG record just prior to a PAF episode from a subject. Since short-term HRV analysis would be done, 288 five minute segments from PAF patients and 510 five minute segments from non-PAF subjects were obtained with nonoverlapping windowing of the series.

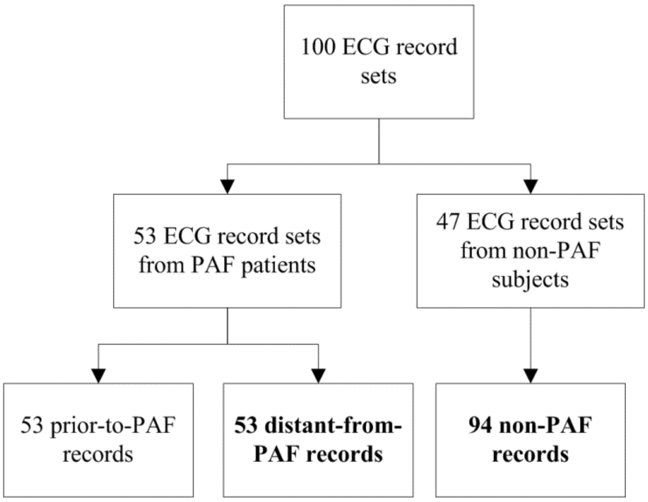


Figure 5.1 The distribution of the afpdb obtained from Physionet. The records shown with bold characters were used in the study.

5.2 Classification Stage

Pattern classification can be defined as assigning a label to a new pattern based on some a priori knowledge (labels, statistics, etc.) of similar data with a learning rule. A learning rule is a procedure that modifies the parameters of a network, which can also be referred to as a training phase. The learning rule is applied to train the network to perform some particular task. There are two learning methods: supervised and unsupervised.

In supervised learning, the learning rule is given with a set of examples (the training set) and their corresponding outputs (targets) such as $\{(p_1,t_1), (p_2, t_2), \ldots, (p_N, t_N)\}$ where p_i is an input to the network, and t_i is the corresponding correct (target) output. As the inputs are applied to the classifier network, the network outputs are compared to the targets and the learning rule adjust the parameters of the network in order to force the network outputs closer to the targets.

In unsupervised learning, on the other hand, the parameters of the network are modified in response to inputs only. There are no target outputs available. In this case, it establishes the classes based on the statistical properties of the patterns (Watanabe, 1985). Most of these algorithms perform clustering operations. They categorize the input patterns into a finite number of classes. This is especially useful in such applications as vector quantization.

Rapidly growing power and speed of computers enabled processing huge and complicated datasets. Thus, many recognition applications are not satisfied with a single classification approach. Therefore, multiple methods and approaches are needed to be used. Thus, combining several methods and classifiers is now commonly used in pattern recognition (Jain, Duin, & Mao, 2000).

In this thesis, supervised learning methods were preferred due to their simplicity and understandability. The Bayes', k-nearest neighbors, artificial neural network and support vector machine classifiers were used to find out the best classification of PAF and non-PAF subjects.

5.2.1 Bayes' Classification

In Bayesian decision theory, tradeoffs between probabilities and decision costs constitute the basis of a classification decision (Duda, Hart, & Stork, 2001). Bayes's rule provides a method of classifying an object optimally into one of c mutually exclusive classes, using

- the prior (a priori) probabilities $P(C_k)$
- the conditional densities $p(x|C_k)$

where $P(.)$ denotes probability mass function and $p(.)$ denotes probability density function. The class conditional probability density function ($p(x|Ck)$) determines how the feature values are distributed for each class. Feature vector x, where $x = \{x_1, \ldots, x_d\}$, will have a distribution described by the probability density function $p(x)$. The probability $P(x)$ of a feature vector x lying in a region R is found by the integration of the probability density function over R

$$P(x \in R) = \int_{x=R} p(x)dx \tag{5.1}$$

The class densities are the separate probability density functions formed for each class: $p(x|C_k)$. The unconditional probability density function is the sum of the class conditional probability density functions weighted by prior probabilities and its independent of class (Duda et al., 2001)

$$p(x) = \sum_{k=1}^{c} \big(p(x|C_k)P(C_k)\big) \tag{5.2}$$

The posterior probabilites may be calculated from

$$P(C_k|x) = \frac{p(x|C_k) \times P(C_k)}{p(x)} \tag{5.3}$$

where $P(C_k|x)$ is the posterior probability for class C_k, $p(x|C_k)$ is the class conditional probability density function for class C_k, $P(C_k)$ is the prior probability of class C_k, $p(x)$ is the unconditional probability density function from Equation (5.2).

To classify x, the class C_k with the highest posterior probability is chosen (or a random choice in the case of equal posterior probabilities) using

$$P(C_k|x) \geq P\big(C_j\big|x\big)\,, \forall j \neq k \tag{5.4}$$

With Bayes's rule, a classifier is a means of partitioning the input space into c decision regions R_k so that feature values in each decision region are associated with a particular class. The probability density of a correct classification at point x, for a class C_k, in a region is $p(x|C_k)P(C_k)$. Therefore, the overall probability of an accurate decision is obtained by integration of the probability density over all regions (Duda et al., 2001)

$$P_{correct} = \sum_{k=1}^{c} \int_{R_k} p(x|C_k)\, P(C_k) dx \tag{5.5}$$

$P_{correct}$ is maximized by the choice of regions to maximize each of c integrands, i.e. by assigning a pattern to class C_k that maximizes $p(x|C_k)P(C_k)$. Since it may be seen that the maximum $P_{correct}$ value is achieved by maximizing $P(C_k|x)$, the posterior probability.

If $p(x|C_k)$ for different classes overlap, e.g. for the cases of two curves with overlapping joint probability density functions, the best classification choice is to minimize the overlap area to maximize Equation (5.5); however, in this case, a perfect classifier cannot be realized.

Naïve Bayes classifier is based on applying Bayes' theorem with strong (naïve) independence assumption (Duda et al., 2001). A conditional probability is defined as the probability of an event given that another event has occurred. The conditional probability of x being in each class C_k is

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)} \tag{5.6}$$

where $\mathrm{P}(C_\mathrm{k}|\mathrm{x})$ is the probability of 'x' being in class $\mathrm{C_k}$

$\mathrm{P}(\mathrm{x}|C_k)$ is the probability of generating instance 'x' given class $\mathrm{C_k}$

$\mathrm{P}(\mathrm{C_k})$ is the probability of occurrence of class $\mathrm{C_k}$

$\mathrm{P}(\mathrm{x})$ is the probability of instance 'x' occurring.

However, 'x' might be a feature vector and it is difficult to calculate $\mathrm{P}(\mathrm{x}|\mathrm{C}_k)$ as in Equation (5.6).

$$\mathrm{P}(\mathrm{C_k}|\mathrm{x_1},\mathrm{x_2},\ldots,\mathrm{x_n}) = \frac{\mathrm{P}(x_1,\mathrm{x_2},\ldots,\mathrm{x_n}|\mathrm{C}_k)P(C_k)}{\mathrm{P}(\mathrm{x_1},\mathrm{x_2},\ldots,\mathrm{x_n})} \tag{5.7}$$

In order to overcome this difficulty, Naïve Bayes classifier assumes that each feature x_1, x_2, …, x_n is independent from each other and Equation (5.7) simplifies to

$$\mathrm{P}(C_k|\mathrm{x}_1,\mathrm{x}_2,\ldots,\mathrm{x}_\mathrm{n})=\frac{\mathrm{P}(x_1|\mathrm{C}_k)\mathrm{P}(\mathrm{x}_2|\mathrm{C}_\mathrm{k})\ldots\mathrm{P}(\mathrm{x}_\mathrm{n}|\mathrm{C}_k)\mathrm{P}(\mathrm{C}_\mathrm{k})}{P(\mathrm{x}_1)P(\mathrm{x}_2)\ldots\mathrm{P}(x_\mathrm{n})} \tag{5.8}$$

Then, performance of the Naïve Bayes classifier is strongly affected from feature dependencies. Kernel smoothing density estimate for probability distributions was used and the kernel smoother was selected normal. In this thesis, the bandwidth of the kernel smoothing window was selected automatically for each combination of features and class, using a value that is optimal for a Gaussian distribution.

5.2.2 K-Nearest Neighbor Classification

K-nearest neighbor (kNN) algorithm is a simple non-parametric method widely used for classification (Altman, 1992). kNN is one of the instance-based classifiers whose system parameters are simply the samples that are presented to the system. This algorithm assumes that all instances correspond to points in the *d*-dimensional space R^d (Mitchell, 1997). An object is classified according to majority of its neighbors. The number of neighbors (k) is a positive integer and generally small. If k = 1, then the object is simply assigned the class of its nearest neighbor. In binary classifications, it is reasonable to choose k as an odd number in order to avoid confusion in case of equal number of neighbors from each class (Nixon & Aguado, 2008).

In kNN calssification, the training set consists of objects whose correct classes are known. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance

$$d_{euclidean} = \sqrt{\sum_i (x_i - y_i)^2} \tag{5.9}$$

where x_i represents the test set features and y_i represents the train set features.

Although using Euclidean distance is very common, other distance measures, such as the Mahalanobis, city-block or cosine (1 minus the cosine of the included angle between observations) distances could be used instead.

There is no specific operation in training phase of the algorithm such as optimizing parameters etc. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the actual classification phase, the test sample (whose class is not known) is represented as a vector in the feature space. Distances from the new vector to all stored vectors are computed and k closest samples are selected. Then, the test sample is just assigned to the class of the majority of its k nearest neighbors (see Figure 5.2).

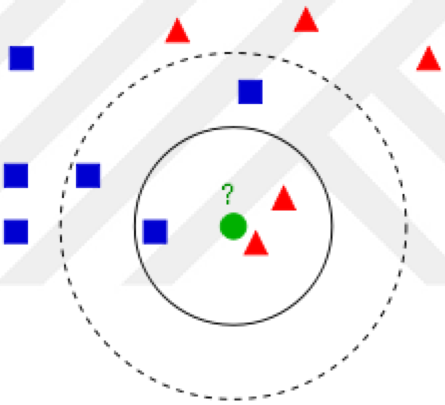


Figure 5.2 Classification of a sample with kNN algorithm. The sample to be classified (test sample) is shown as a circle. The train samples from two different classes are represented with squares and triangles. If k was selected 3, the test sample would be assigned to class ‘triangle’. If k was selected 5, the test sample would be assigned to class ‘square’.

kNN classification method does not depend on the data following any particular distribution unlike many other classifiers which assume a multivariate Gaussian distribution of the feature values. Thus, a kNN classifier allows a great deal of generality in the classification (Duda et al., 2001). But, there are some disadvantages of the algorithm. Firstly, for large data sets, this algorithm is very time-consuming because each sample in the training set is processed while classifying a new data and this requires a longer classification time. Secondly, the accuracy of the kNN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

5.2.3 Artificial Neural Network Classification

An artificial neural network (ANN) is an interconnected group of simple units with adaptive weights (Bishop, 1995). ANNs are inspired by biologic neural networks and are composed of neuronlike 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

$$y_i = f\left(\sum_{i=1}^{N} x_i w_i\right) \tag{5.10}$$

where x_i is the input, w_i are the weights, $f(.)$ is the activation function and y_i is the output of the i^{th} unit. Different functions can be utilized as the activation function but most often a sigmoid (or hyperbolic tangent) function is used.

Multi-layer perceptrons (MLPs) are the mostly used ANN structures. 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 5.3).

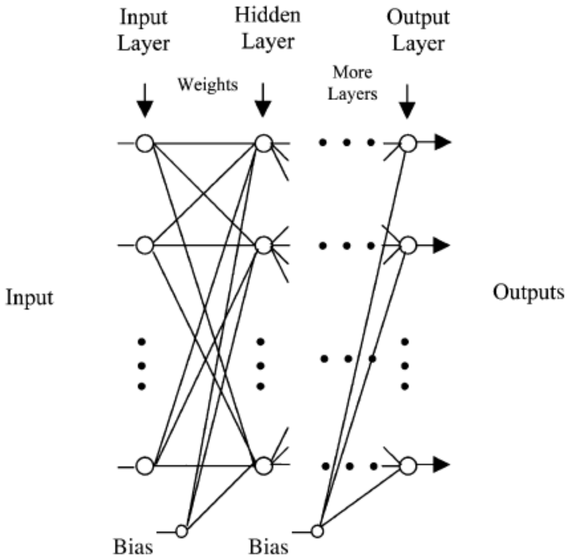


Figure 5.3 The architecture of an MLP network.

A training phase, where the values of the connection weights are adjusted, is needed for ANN algorithm to accomplish the desired task. Then 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). However, this should be noted that the ultimate aim of the training of 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 avoids overtraining and increases the generalization performance of the network (Hagiwara & Kuno, 2000). 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 5.4, the training is finished when the validation error reaches its minimum.

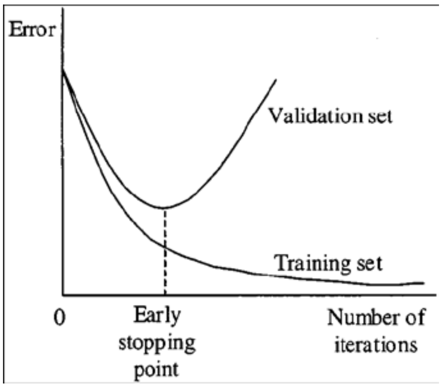


Figure 5.4 Training and validation errors versus iteration number. Training stops when the the error at validation set starts to increase

Commonly used training algorithms are summarized below:

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 et al., 2001; Yu, Efe, & Kaynak, 2002) Generally the error function used is the mean square error (MSE):

$$E_{MSE}(y) = \frac{1}{2}\sum_{j=1}^{N}\left(t_j - y_j\right)^2 \tag{5.11}$$

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

$$\Delta w_{ji}^{l} = \eta \delta_j^{(l)} y_i^{(l-1)} \tag{5.12}$$

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

$$\delta(n) = \frac{\partial E}{\partial w} \tag{5.13}$$

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

$$e_r(n) = \frac{E(n) - E(n-1)}{E(n)} \tag{5.14}$$

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

$$\begin{aligned} for\ \ e_r < 0, \eta(n+1) = \eta(n)(l + u.e^{-e_r(n)}) \\ for\ \ e_r > 0, \eta(n+1) = \eta(n)(l - u.e^{-e_r(n)}) \end{aligned} \tag{5.15}$$

where $\eta(n+1)$ is the updated learning rate, $\eta(n)$ 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) (Chen, Han, Au, & Tham, 2003; Hagan & Menhaj, 1994). When training with the LM method, the update of the weights are obtained as

$$\Delta w = [J^T J + \lambda I]^{-1} J^T e \tag{5.16}$$

where J is the Jacobian matrix, λ is the learning parameter, and e is the sum of error squares. In this study, LM algorithm was used for the training of ANN classifier.

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

$$\tilde{E} = E + v\Omega \quad where \quad \Omega = \frac{1}{n}\sum_{i=1}^{n} w_i^2 \tag{5.17}$$

Here *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.

5.2.4 Support Vector Machine Classification

Support vector machine (SVM) is a powerful technique proposed for solving pattern recognition problems (Blanz et al., 1996; Cortes & Vapnik, 1995; Osuna, Freund, & Girosi, 1997). According to the theory of SVMs (Vapnik, 1982, 1995), they minimize the structural risk while traditional techniques for pattern recognition are based on the minimization on the training set. This new induction principle, which is equivalent to minimize an upper bound on the generalization error, relies on the theory of uniform convergence in probability (Vapnik, 2006).

In the linearly separable case, the key idea of an SVM can be explained more easily. Suppose a training set *S* is given. The set contains points of either of two classes, a SVM separates the classes through a hyper-plane determined by certain points of *S*, termed “support vectors” (see Figure 5.5). In the separable case, this hyper-plane maximizes the margin, or twice the minimum distance of either class from the hyper-plane, and all the support vectors lie at the same minimum distance from the hyper-plane. In real cases, the two classes may not be linearly separable and the data vectors are mapped to a higher dimensional space by a kernel function. Then SVM finds a linear separating hyperplane with the maximal margins in this higher dimensional space. The solution is a trade-off between the largest margin and the lowest number of errors; trade-off is controlled by a regularization parameter.

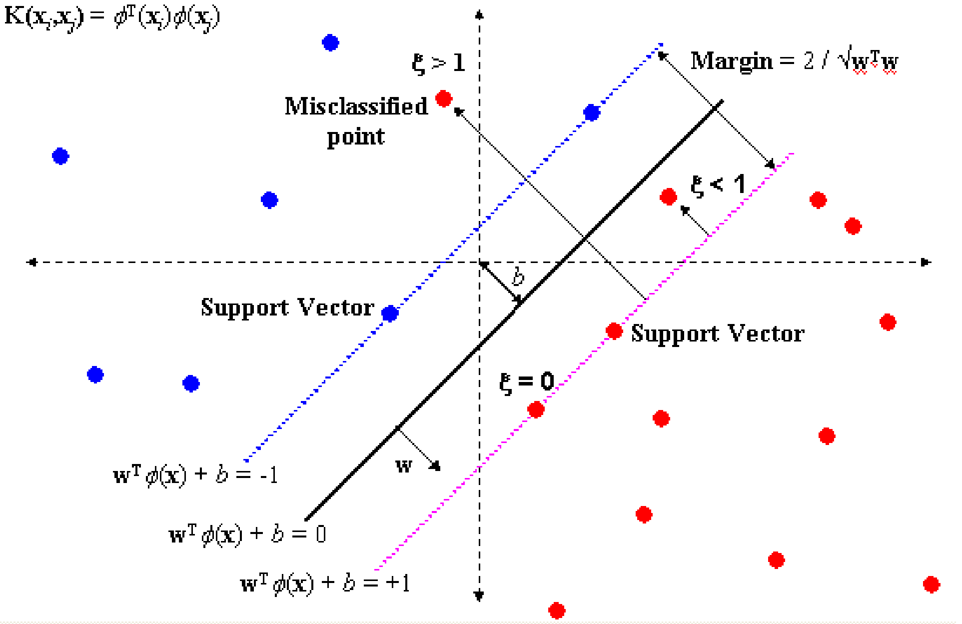


Figure 5.5 A representation of support vector machine algorithm. Red and blue dots represent two different classes. SVM finds the optimum hyper plane that maximizes the margin between support vectors while trying to minimize wrong classifications.

Given a training set of (x_i, y_i), i=1, 2, …,l where $x_i \in R^n$ and $y_i \in \{-1, 1\}$, the traditional SVM algorithm is reduced to the optimization problem

$$\min_{w,b,\xi} \left\{ \frac{1}{2} w^T w + C \left(\sum_{i=1}^{l} \xi_i \right) \right\} \tag{5.18}$$
$$\text{subject to}: \; y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \;\; \xi_i > 0 \; \forall i$$

where $\phi(x)$ is a nonlinear function that maps x into a high dimensional space (Wang, Liu, & Wan, 2005). w, b and ξ_i are the weight vector, bias and slack variable, respectively. C is a constant and determined a priori. Searching for the optimal hyperplane in Equation (5.18) is a quadratic programming problem, which can be solved by constructing a Lagrangian and transforming it into a dual maximization problem of the function *Q(a)*, defined as

$$\begin{gathered}\max Q(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j\mathrm{y}_i y_j K\left(\mathrm{x}_i,\mathrm{x}_j\right)\\ \text{subject to}:\ \sum_{i=1}^{l}\alpha_i\mathrm{y}_i=0\,,\ \ 0\leq\alpha_i\leq C\text{ for }i=1,2,\ldots,l\end{gathered}\tag{5.19}$$

where $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$ is the kernel function and, $\alpha = (\alpha_1, \alpha_2, . . ., \alpha_l)$ is the vector of nonnegative Lagrange multipliers. Assuming that the optimum values of the Lagrange multipliers are denoted as $\alpha_{o,i}(i=1, 2,. . ., l)$, it is then possible to determine the corresponding optimum value of the linear weight vector $\mathrm{w_o}$ and the optimal hyperplane as in Equations (5.20) and (5.21), respectively

$$\mathrm{w_o}=\sum_{i=1}^{l}\alpha_{o,i}\mathrm{y}_i\phi(\mathrm{x}_i)\tag{5.20}$$

$$\sum_{i=1}^{l}\alpha_{o,i}\mathrm{y}_i K(\mathrm{x},\mathrm{x}_i)+b=0\tag{5.21}$$

The decision function can be written as

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{l}\alpha_{o,i}\mathrm{y}_i\mathrm{K}(\mathrm{x},\mathrm{x}_i)+b\right)\tag{5.22}$$

In this work, the Gaussian radial basis function (RBF) given in Equation (5.23) was used as the kernel function.

$$K\left(x_i,x_j\right)=exp\left(-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}\right)\tag{5.23}$$

The parameters, which are kernel width 'σ' and regularization constant '*C*', were experimentally defined to achieve the best classification result. Grid search was performed to find σ and C values that give the minimum error rate in classification. In grid search, several (σ, C) pairs were used for classification and the pair with minimum error rate (maximum accuracy) was selected. Cross validation was made to take the classification success of whole data into consideration. Figure 5.6 and Figure 5.7 shows error rate for many (σ , C) values. The values of both variables were searched from 0 to 15 with 0.5 steps. The error rate was minimum for σ=10.5 and C=1 when all 33 features were used. The σ and C values found to be 4 and 1.5 respectively when selected 8 features were used.

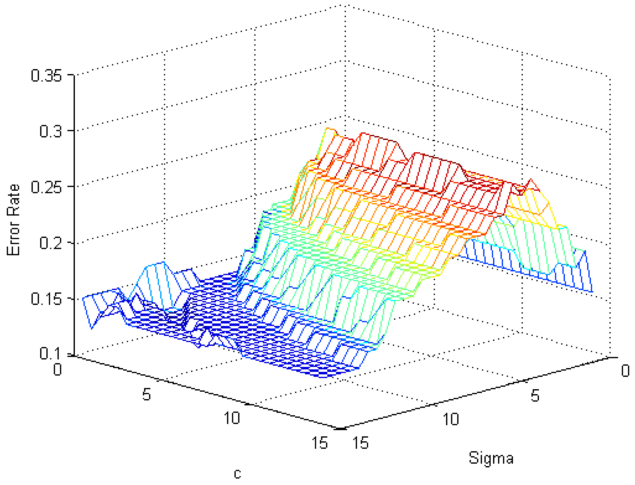


Figure 5.6 Error rate plot when all 33 features were used for classification. Minimum error rate is 0.13 when σ=10.5 and c=1.

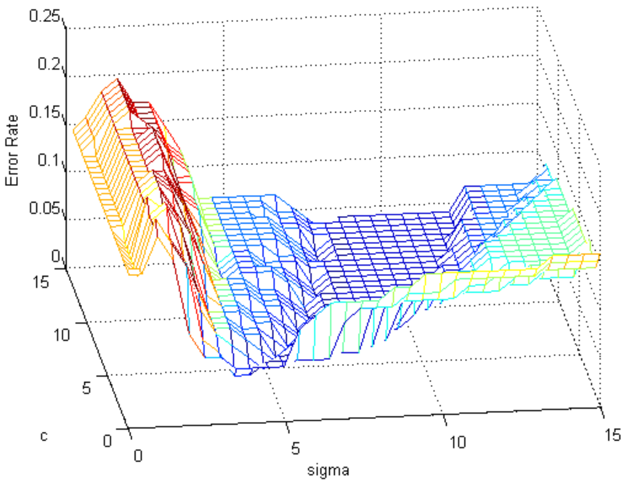


Figure 5.7 Error rate plot when selected 8 features were used for classification. Minimum error rate is 0.05 when σ=4 and c=1.5.

5.3 Model Evaluation

In this section, a few standard measures are explained to discuss the applications and accomplishments of various classifiers used in this study. These measures have been recommended by physicians and health-care workers by various organizations, and therefore are good measures for evaluating the performance of any automated system that is designed to assist these health-care professionals (Valafar, 2000).

5.3.1 Cross Validation

The performance of the features to discriminate between PAF patients and non-PAF subjects were evaluated using 10-fold cross validation. With this validation, the whole dataset was divided into ten groups randomly and one group was used for testing while the remaining nine groups were used for training. The same procedure was repeated for all folds and the average of performance values were used for the final evaluation. In this study, each fold had approximately 80 five minute RR interval series and the numbers of PAF and non-PAF segments were controlled to ensure that each fold represented the whole data adequately (approximately 50 from non-PAF subjects and 30 from PAF patients).

Due to the splitting of 30 min RR series into 5 min segments, there were at least six 5-minutes RR segments from the same subject. If the whole dataset had been divided into test and train groups randomly, some of the 5 min records belonging to a specific subject would have been in the train set while the remaining ones of the same subject would be in the test set. Such a situation would probably affect the results positively but erroneously. To avoid such a bias, the test and train sets were constructed on a subject basis. This way, all 5 min RR segments belonging to a specific subject was put in either the test or training set.

5.3.2 Performance Assessment

The output of a binary classifier is generally designed as 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 5.1. Throughout this thesis, PAF patients were regarded as ‘Positive’ and non-PAF subjects were regarded as ‘Negative’ during classifications.

Table 5.1 A two-by-two confusion matrix.

		System Output	
		Positive	**Negative**
True Condition	**Positive**	True positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Performance of a recognition system is measured by several parameters using the decision matrix (Eberhart, 2014). Sensitivity (SEN), selectivity (SEL), specificity (SPE), and overall accuracy (ACC) are the mostly used parameters.

Sensitivity is described as the ratio of the number of positives correctly classified by the recognition system to the total number of real positives. It shows the ratio of the correctly classified PAF patients to the total number of PAF patients:

$$Sen = \frac{TP}{TP + FN} \times 100 \tag{5.24}$$

Specificity is described as the ratio of the number of negatives correctly classified by the recognition system to the total number of real negatives. It shows the ratio of correctly classified non-PAF subjects to the total number of non-PAF subjects:

$$Spe = \frac{TN}{TN + FP} \times 100 \tag{5.25}$$

Selectivity is described as the ratio of the number of positives correctly classified by the recognition system to the total number of samples classified as positive. It shows the ratio of correctly classified PAF patients to the total number of subjects classified as PAF:

$$Sel = \frac{TP}{TP + FP} \times 100 \tag{5.26}$$

Overall Accuracy is the ratio of the total number of positives and negatives correctly classified by the recognition system to the all decisions.

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \times 100 \tag{5.27}$$

In addition to these performance measures, positive predictive value (PPV) and negative predictive value (NPV) are two other commonly used performance measures that describe the performance of a diagnostic test (Fletcher, Fletcher, & Wagner, 2005). Positive predictive value is the same with selectivity measure described by Equation (5.26). Negative predictive value is described as the ratio of the number of negatives correctly classified by the recognition system to the total number of samples classified as negative. It shows the ratio of correctly classified non-PAF subjects to the total number of subjects classified as non-PAF:

$$NPV = \frac{TN}{TN + FN} \times 100 \tag{5.28}$$

Another performance evaluation tool is Receiver Operating Characteristic (ROC) analysis. It is a plot of sensitivity (true positive rate) versus 1-specificity (false positive rate) values as shown in Figure 5.8. It is widely used in the medical applications to evaluate the performance of diagnostic tests. 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, Park, Hager, Brock, & Diener-West, 2004). 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 0.80-0.90 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 (Alberg et al., 2004; Fawcett, 2006; Van Erkel & Peter, 1998).

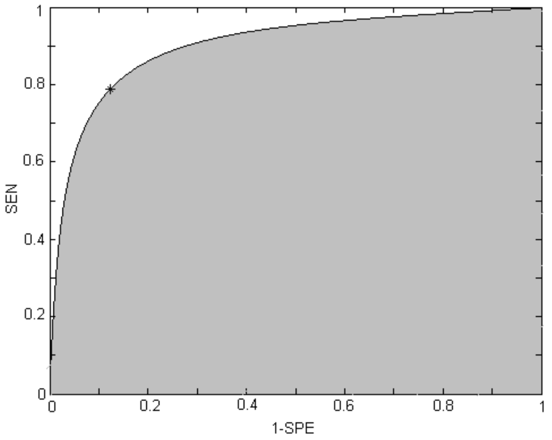


Figure 5.8 Area under ROC curve.

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.

Ideally, the predictive value in terms of ROC areas is based on a series of sensitivity-specificity pairs. However, in this study there is only one pair per each system and the ROC area reduces to a trapezoidal area (see Figure 5.9) which corresponds to the average of sensitivity and specificity values (Inan & Kuntalp, 2007). These areas were given in classification performance tables as AUC values.

$$AUC \approx \frac{SEN + SPE}{2} \tag{5.29}$$

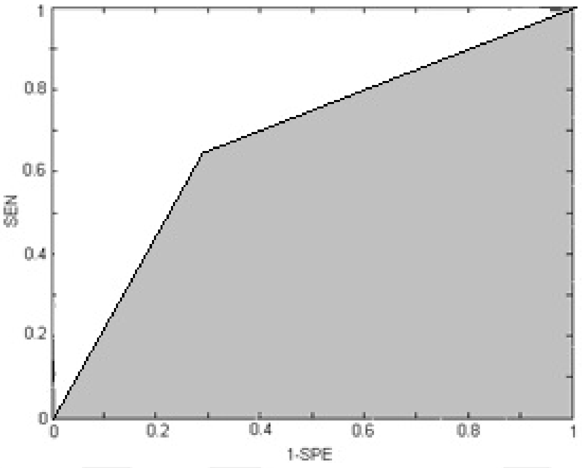


Figure 5.9 ROC curve in case of one sensitivity-specificity pair. The area under curve is the average of sensitivity and specificity values.

5.4 Representation of Data with 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, 2001). There are two usages of SOM 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. Hence, 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 were 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, 1993). In these models of SOM, very large numbers of neurons are used, generally over 1000.

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 in Figure 5.10. 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.

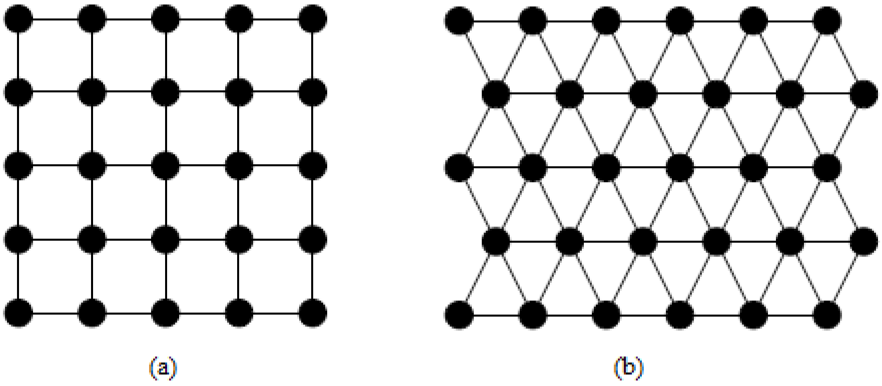


Figure 5.10 SOM network structures (a) Rectangular Grid (b) Hexagonal Grid.

An N-dimensional input is presented to each neuron of a SOM network as shown in Figure 5.11. Then, the winner unit (indicated by the index c), i.e. best match, is identified by the condition shown below for each sample

$$\|x_i(t) - w_c(t)\| = \min_t \|x_i(t) - w_i(t)\| \tag{5.30}$$

where x_i is the input vector with dimension N, w_i is the *i-th* weight, and c indicates the winning neuron.

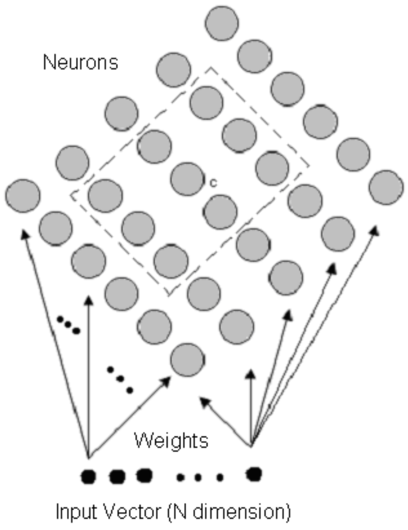


Figure 5.11 Self organizing map (SOM) structure.

The update of the weights in the SOM network is limited by neighborhood function ($\Omega_c(i)$). The neighborhood function plays a fundamental 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

$$\Delta w_i = \eta(x - w_i)\Omega_c(i), \qquad i \in NB_c \tag{5.31}$$

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.

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 SOM 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 (Iivarinen, Kohonen, Kangas, & Kaski, 1994; Oja et al., 2006; Ultsch, 1993). 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, 1993). 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 in Figure 5.12. This is one simple way of calculating of the U-matrix with dummy grids (Oja et al., 2006). 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 (Iivarinen et al., 1994).

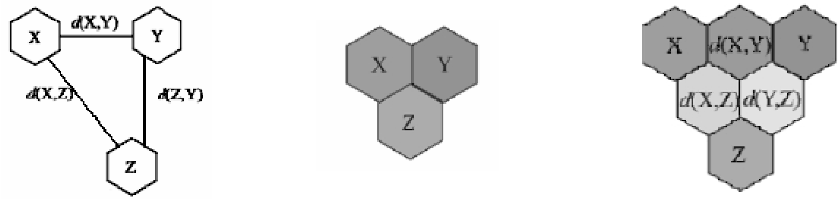


Figure 5.12 A simple way of calculating the U-matrix with dummy grids.

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 grayscale 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 5.13 shows a U-matrix representation of a SOM network with a 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.

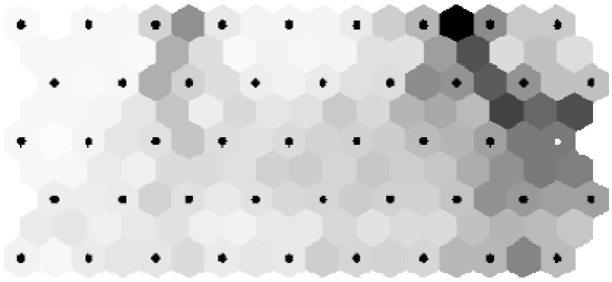


Figure 5.13 U-matrix representation of SOM network with gray-level image.

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. Thus, hills represent border of the clusters. 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.

CHAPTER SIX

RESULTS AND DISCUSSION

Diagnosing PAF patients from their normal sinus rhythm ECG is an important clinical challenge. This could lead to early and easy detection of the illness to take precautions. The whole study is visualized with a flowchart in Figure 6.1. In this study, a feature vector (see the first column in Table 6.4) was constructed from HRV measures and atrial & ventricular ectopic beat numbers. HRV measures were obtained from time domain, frequency domain, and nonlinear analysis of RR interval series. The RR intervals were obtained from Physionet Atrial Fibrillation Prediction Database (afpdb). Totally 288 five minute records from PAF patients and 510 five minute records from non-PAF subjects were obtained by windowing thirty minute signals. Short-term (5 min) HRV analysis was made. In the preprocessing stage, atrial and ventricular ectopic beats were eliminated from the series. The numbers of atrial and ventricular ectopic beats in each 5 min data were also kept as two features. For the frequency domain analysis, the series were interpolated using cubic spline method, resampled with 4 Hz for frequency domain analysis and then slowly varying trends were eliminated with smoothness priors detrending method with a threshold of α=500 which corresponds to 0.035 Hz cutoff frequency. Finally, feature vectors with length 33 (6 time-domain features, 12 frequency domain features, 13 nonlinear features, atrial and ventricular ectopic beat numbers) were constructed and they were fed to the inputs of different classifiers: k-nearest neighbor, Bayes', artificial neural network and support vector machine. Genetic algorithms was used for feature selection and eight features (mean of RR intervals, standard deviation of RR intervals, HF band peak frequency, relative power of LF band, relative power of HF band, dispersion of points perpendicular to line of identity in Poincare plot (SD1), sample entropy and short term fluctuation slope of detrended fluctuation analysis ($\alpha 1$) were found to be the best discriminating features. All classifications were made with two feature vectors: all features and selected features.

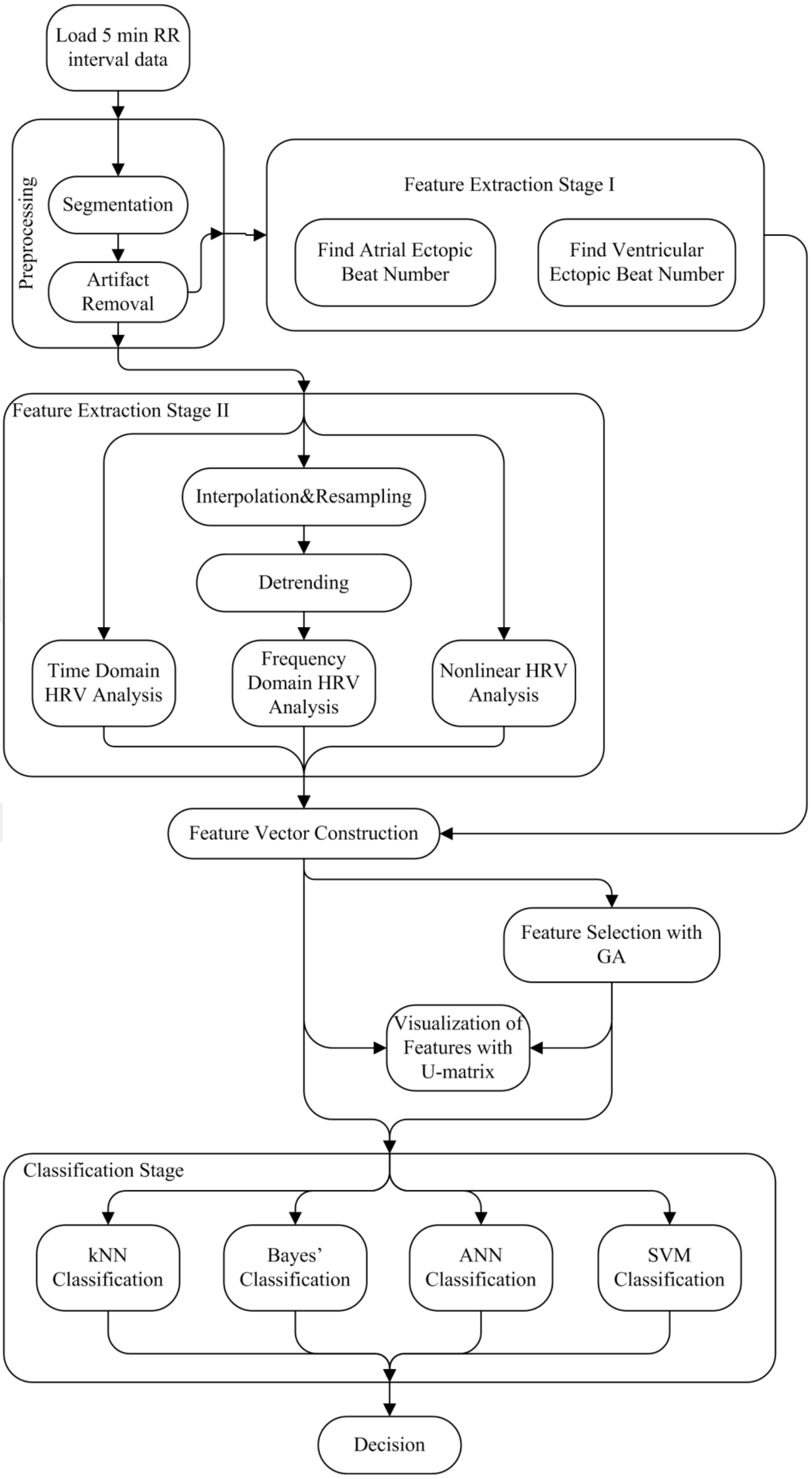


Figure 6.1 Flowchart of the whole study.

The first classification was performed with kNN classifier. The number of neighbors (k) and the distance metric were the two parameters that were examined for best performance. The number of nearest neighbors varied from 1 to 11 for all features and selected features. The distance metrics, which are ‘Euclidean’, ‘Cityblock’ and ‘Cosine’, were also examined with different runs of the classifier.

The results obtained by increasing the number of neighbors for all and selected features are given in Table 6.1-6.3. The results obtained with ‘Euclidean’ distance were given in Table 6.1, ‘Cityblock’ distance in Table 6.2 and ‘Cosine’ distance in Table 6.3. In all tables, SEN refers to sensitivity (%), SPE refers to specificity (%), PPV refers to positive predictive value (%), NPV refers to negative predictive value (%), ACC refers to overall accuracy (%), AUC refers to the area under ROC curve. In features (Feat.) column, “ALL” indicates that all features were used and “GA” indicates that selected features by GA were used for classification.

In all distance metrics, increasing the number of neighbors negatively affects the classification performance for both feature groups. This might be due to the unequal number of samples from PAF and non-PAF groups. In kNN classification, the more frequent examples tend to dominate the prediction of the new vector as they tend to come up in the k-nearest neighborhood due to their large numbers (Watanabe, 1985).

The most successful kNN classification was achieved with k=1 and ‘Euclidean’ distance for both all and selected features. For all features, the performance measures were 67% sensitivity, 80% specificity, 75% accuracy and the area under ROC curve was 0.74. The classification success increased when selected features were used. For selected features, the performance measures were 72% sensitivity, 87% specificity, 82% accuracy and the area under ROC curve was 0.8.

Table 6.1 Classification results where k is the parameter of the kNN classifier. The distance metric is ‘Euclidean’.

Euclidean Distance							
Feat.	**k**	**SEN**	**SPE**	**PPV**	**NPV**	**ACC**	**AUC**
ALL	1	67	80	65	81	75	0.74
	3	55	79	59	76	70	0.67
	5	52	80	61	75	70	0.66
	7	50	83	64	75	71	0.67
	9	49	85	66	75	72	0.67
	11	47	84	64	74	70	0.66
GA	1	72	87	77	85	82	0.80
	3	64	82	68	81	75	0.73
	5	59	81	68	77	73	0.70
	7	60	81	70	78	74	0.71
	9	58	82	71	77	74	0.70
	11	56	81	69	76	72	0.69

Table 6.2 Classification results where k is the parameter of the kNN classifier. The distance metric is ‘Cityblock’.

Cityblock Distance							
Feat.	**k**	**SEN**	**SPE**	**PPV**	**NPV**	**ACC**	**AUC**
ALL	1	49	76	54	72	66	0.63
	3	46	82	59	73	69	0.64
	5	45	84	63	73	69	0.65
	7	46	85	65	74	71	0.66
	9	43	84	61	72	69	0.64
	11	44	84	62	73	69	0.64
GA	1	66	82	68	82	76	0.74
	3	63	79	65	79	73	0.71
	5	58	81	67	77	72	0.70
	7	53	80	63	75	70	0.67
	9	53	81	66	75	71	0.67
	11	50	80	63	74	69	0.65

Table 6.3 Classification results where k is the parameter of the kNN classifier. The distance metric is ‘Cosine’.

Cosine Distance							
Feat.	**k**	**SEN**	**SPE**	**PPV**	**NPV**	**ACC**	**AUC**
ALL	1	65	79	63	80	74	0.72
	3	56	80	61	76	71	0.68
	5	51	81	61	75	70	0.66
	7	51	82	62	75	71	0.67
	9	49	83	63	75	71	0.66
	11	49	84	64	75	71	0.67
GA	1	64	81	66	80	75	0.73
	3	64	80	65	80	74	0.72
	5	61	81	67	79	74	0.71
	7	61	80	68	79	74	0.71
	9	59	79	65	77	72	0.69
	11	57	80	64	78	72	0.69

The comparison of the success of the distance metrics is given in Figure 6.2 and Figure 6.3 for all features and selected features, respectively. Euclidean distance has been shown to be superior to the other metrics for both all and selected features.

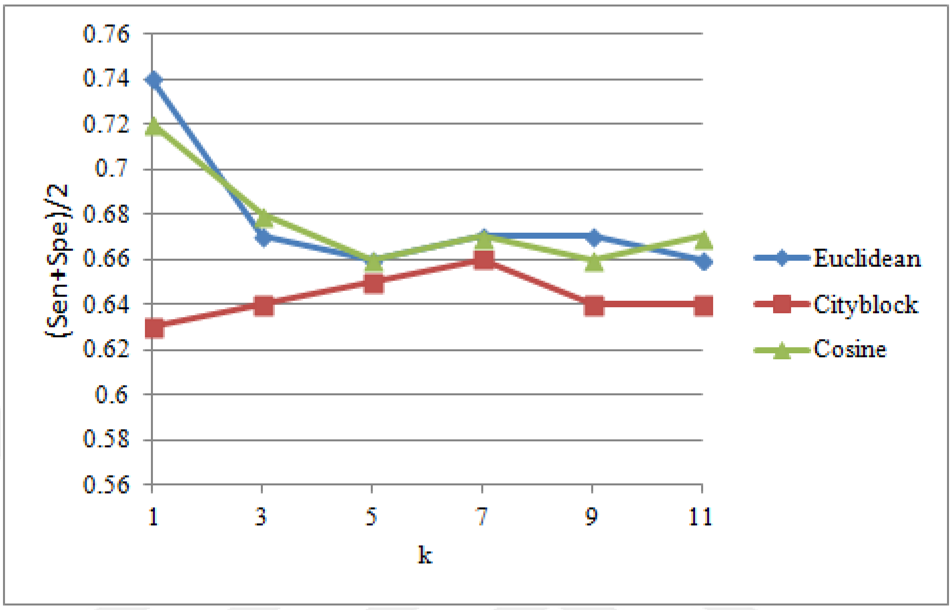


Figure 6.2 Number of neighbors versus the areas under ROC curve for three different distance metrics by using all features.

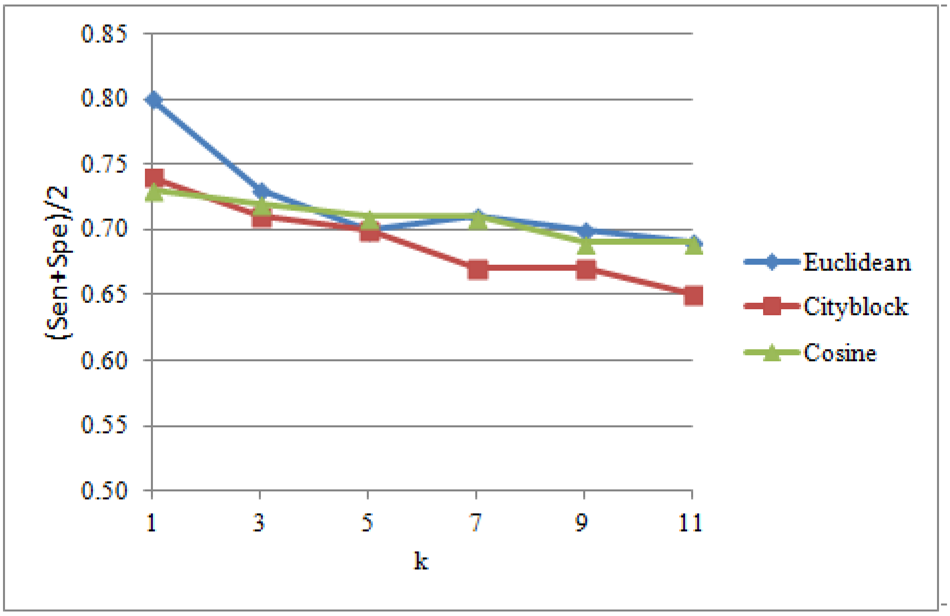


Figure 6.3 Number of neighbors versus the areas under ROC curve for three different distance metrics by using selected features.

After determining the best nearest neighbor number and distance metric, the classification performance of each feature was examined. The results are given in Table 6.4. Also the mean values and standard deviations of each feature for PAF and non-PAF groups are given together with p-values.

Table 6.4 Classification results of each feature with kNN classifier where k is 1 and distance metric is Euclidean. Features are presented as mean±standard deviation and their p-values are given.

Feature	PAF	Non-PAF	p	SEN	SPE	PPV	NPV	ACC	AUC
Mean RR	0.913±0.19	0.862±0.21	0	41	63	38	66	55	0.52
SDRR	0.023±0.01	0.029±0.02	0	38	61	35	63	52	0.50
SDHR	2.275±1.59	2.977±1.75	0	37	63	36	64	54	0.50
RMSSD	0.030±0.02	0.032±0.02	0.27	39	65	38	65	55	0.52
nn50	27.646±44.49	29.788±40.3	0.5	82	20	37	68	42	0.51
pnn50	8.815±13.69	9.575±13.81	0.45	56	51	39	67	53	0.54
VLF_peak	0.034±0.00	0.035±0.00	0.14	100	0	36	NaN	36	0.50
LF_peak	0.071±0.03	0.069±0.03	0.33	100	1	36	NaN	37	0.51
HF_peak	0.264±0.06	0.255±0.07	0.002	87	8	35	52	37	0.48
VLF_power	0.000±0.00	0.000±0.00	0.35	37	66	38	65	56	0.52
VLF_power_prc	10.459±10.08	11.391±9.28	0.19	40	66	39	66	57	0.53
LF_power	0.000±0.00	0.001±0.00	0	39	64	38	65	55	0.52
LF_power_prc	35.650±20.45	49.527±20.49	0	51	70	50	72	63	0.61
LF_power_norm	41.654±25.48	56.771±23.72	0	45	68	45	69	60	0.57
HF_power	0.000±0.00	0.000±0.00	0.02	39	66	39	66	56	0.53
HF_power_prc	53.891±26.53	39.083±23.03	0	45	69	46	70	61	0.57
HF_power_norm	58.346±25.48	43.229±23.72	0	43	68	44	68	59	0.56
LF/HF_power	1.418±2.02	2.627±3.09	0	41	67	42	67	58	0.54
SD1	0.021±0.01	0.023±0.02	0.038	40	35	39	66	56	0.38
SD2	0.045±0.03	0.060±0.04	0	40	69	42	67	58	0.55
ApEn	0.994±0.20	0.986±0.20	0.55	37	64	37	64	54	0.51
SampEn	1.456±0.51	1.358±0.50	0.218	36	64	35	64	54	0.50
DFA_α1	0.859±0.38	1.013±0.37	0	42	65	41	67	57	0.54
DFA_α2	0.970±0.31	0.987±0.28	0.44	39	66	40	60	57	0.53
CorDim_D2	1.068±1.20	1.439±1.27	0	38	61	35	63	52	0.50
RPA_L_{max}	187.417±111.84	226.396±140.84	0	59	51	40	69	54	0.55
RPA_L_{mean}	18.626±16.91	19.611±15.58	0.41	43	69	44	68	60	0.56
RPA_REC	0.388±0.17	0.410±0.14	0.06	39	62	37	64	53	0.51
RPA_DET	0.980±0.02	0.985±0.02	0	42	67	42	67	58	0.55
RPA_ShanEn	3.381±0.58	3.495±0.55	0.007	39	64	38	65	55	0.52
CCM	-0.355±0.38	-0.186±0.12	0	44	67	44	68	59	0.56
AEB #	1.740±5.54	0.941±3.69	0	99	1	36	NaN	36	0.50
VEB #	4.764±10.17	3.359±10.73	0.002	95	2	35	NaN	35	0.49

The statistical analysis of the dataset was made with unpaired t-test (see the Appendix). The p-values less than 0.05 were regarded as statistical evidence that there was significant difference between two groups. The eight features selected by GA were denoted with bold italic characters in Table 6.4. Other than SampEn, p-values of the selected features have been shown to be significantly different ($p<0.05$) for PAF and non-PAF groups.

Naive Bayes classification was the second method tried. Here, kernel smoothing density estimate for probability distributions was used. The kernel smoother was selected as normal. The bandwidth of the kernel smoothing window was selected automatically for each combination of feature and class, using a value that is optimal for a Gaussian distribution. The results were similar to kNN classification results as given in Table 6.5. The best performance rates are 69% sensitivity, 83% specificity, 76% accuracy for all features and 74% sensitivity, 88% specificity and 83% accuracy for selected features.

Table 6.5 Classification results with Bayes' classifier.

Alg.	**SEN**	**SPE**	**PPV**	**NPV**	**ACC**	**AUC**
ALL	69	83	73	81	76	0.71
GA	74	88	81	85	83	0.81

The third classification method was ANN classification. Levenberg-Marquardt algorithm was preferred for faster speed of convergence. The networks used with all features had 33 inputs, whereas the networks used with only the selected features had 8 inputs. The output of both neural network classifiers had two neurons and a single hidden layer. However, the number of neurons in the hidden layer was varied in order to find the classifier with the best performance as given in Table 6.6. The optimum number of hidden units was found to be 27 for the networks constructed for all features. This number turned out to be 8 for the networks which work with the selected features only. The transfer functions of the hidden layer units were all chosen as tangential sigmoid whereas they were linear for the units of the output layer. Mean squared error was chosen as the performance criterion.

The best performance for all features was achieved with 27 units in hidden layer with 78% sensitivity, 89% specificity, 83% accuracy, and 0.84 area under ROC curve. Classification with selected features was again more succesfull. With 8 neurons in hidden layer, 76% sensitivity, 95% specificity, 88% accuracy were obtained.

Table 6.6 Classification results where the number of hidden layer units is the parameter of the ANN classifier. Levenberg-Marquardt training algorithm was used.

Alg.	Hidden layer unit #	SEN	SPE	PPV	NPV	ACC	AUC
ALL	35	67	88	77	83	81	0.77
	34	76	81	71	85	79	0.78
	33	76	88	80	88	84	0.82
	32	75	87	80	86	83	0.81
	31	77	88	81	87	84	0.82
	30	74	89	81	86	84	0.82
	29	54	92	81	78	78	0.73
	28	62	90	78	80	79	0.76
	27	78	89	81	87	83	0.84
	26	73	83	74	84	80	0.78
	25	72	87	78	85	82	0.79
	24	73	85	74	84	80	0.79
	23	65	88	76	82	80	0.77
	22	68	91	83	84	84	0.80
GA	11	75	84	79	83	80	0.80
	10	77	94	91	87	87	0.86
	9	78	92	89	88	86	0.85
	8	76	95	92	87	88	0.86
	7	67	95	93	84	84	0.81
	6	70	91	83	83	82	0.81
	5	75	84	79	83	80	0.80
	4	65	87	73	81	78	0.76

The last employed classifier was the SVM classifier. Gaussian radial basis function (RBF) was used as the kernel function that maps the features to a higher dimensional feature space where the PAF and non-PAF subjects can be linearly separable. The kernel width σ and regularization constant C values were determined with a grid search for all features and selected features (see Section 5.2.4 for details). Different values of σ and C for all and selected features yield the best classification as given in Table 6.7. SVM classifier is the most successful classifier with 84% sensitivity, 91% specificity, 88% accuracy and area under ROC curve was 0.88 for all features. When selected features were used 93% sensitivity, 95% specificity, 95% accuracy were obtained and area under ROC curve was 0.94.

Table 6.7 Classification results with SVM.

Alg.	SEN	SPE	PPV	NPV	ACC	AUC
ALL (σ=10.5,C=1)	84	91	86	91	88	0.88
GA (σ=4,C=1.5)	93	95	93	96	95	0.94

Using selected features (summarized in Table 6.8) increased the performance for all classifiers. This is probably a result of the curse of dimensionality effect. The most successful classification results obtained from each classifier is summarized in Table 6.9. The most successful records were obtained by using a vector of the features given in Table 6.8 as the input to an SVM classifier which discriminated PAF patients and non-PAF subjects with 95% accuracy.

Table 6.8 Features selected by genetic algorithm given with their mean±std and p values.

Feature	Description	PAF	Non-PAF	p-value
Mean RR	Mean of RR intervals	0.913±0.19	0.862±0.21	0
SDRR	Standard deviation of RR intervals	0.023±0.01	0.029±0.02	0
HF peak	HF band peak frequency	0.264±0.06	0.255±0.07	0.002
LF power prc	Relative power of LF band (LF power/Total power)	35.650±20.45	49.527±20.49	0
HF power prc	Relative power of HF band (HF power/Total power)	53.891±26.53	39.083±23.03	0
SD1	Dispersion of points perpendicular to line of identity in Poincare plot	0.021±0.01	0.023±0.02	0.038
SampEn	Sample entropy	1.456±0.51	1.358±0.50	0.218
DFA α1	Short term fluctuation slope of detrended fluctuation analysis	0.859±0.38	1.013±0.37	0

Table 6.9 Classification results of all classifiers when all and selected features were used.

Alg.	**Classifier**	**SEN**	**SPE**	**PPV**	**NPV**	**ACC**	**AUC**
ALL	SVM (Gaussian RBF kernel, σ=10.5,c=1)	84	91	86	91	88	88
	ANN (27 units in hidden layer)	78	88	80	87	84	83
	Bayes' (Kernel smoothing density)	69	83	73	81	76	76
	kNN (1 neighbor, Euclidean distance)	68	77	64	80	74	73
GA	SVM (Gaussian RBF kernel, σ=4,c=1.5)	93	95	93	96	95	94
	ANN (8 units in hidden layer)	76	95	92	87	88	86
	Bayes' (Kernel smoothing density)	74	88	81	85	83	81
	kNN (1 neighbor, Euclidean distance)	72	89	82	85	83	81

The distributions of the normalized features belonging to PAF and non-PAF groups are also given in Figure 6.4. The need for complex combination of these features to successfully discriminate PAF and non-PAF subjects blurs the clinical meaning of each single feature in the model. However, it seems that there is an increase in relative power of HF band (vagal tone) and a decrease in relative power of LF band (sympathetic and vagal) in PAF patients. Also a decrease is observed in short term fluctuation slope of DFA (complexity) in PAF patients. Mean RR values of PAF patients were also found to be higher which means there is a decrease in their heart rate which supports the work of Hoshino et al. (Hoshino et al., 2013).

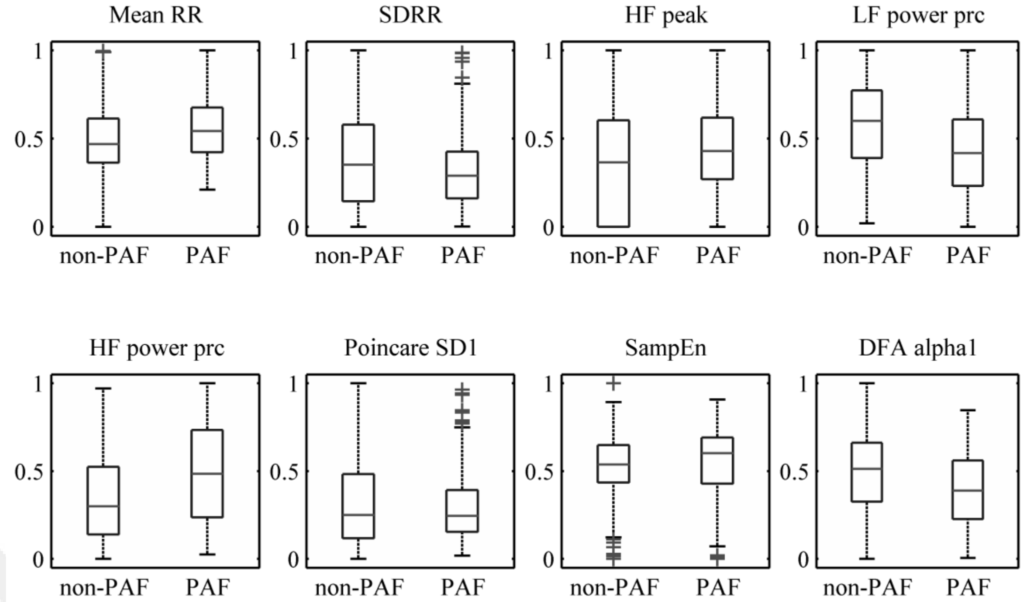
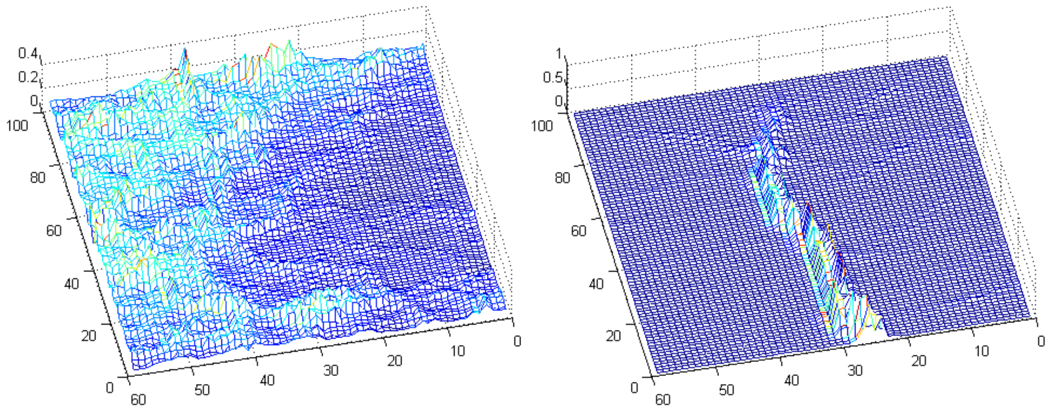


Figure 6.4 Distribution of non-PAF and PAF groups for selected features by the genetic algorithm. The features other than sample entropy (p-value 0.2183) found to be statistically different between PAF and non-PAF groups ($p<0.05$).

In order to visualize the high dimensional feature vector, Matlab SOM Toolbox was used to get the U-matrix representation of our data. A 30x50 grid was selected. The hill-valley representation for all the features is given in Figure 6.5a and for the selected features is given in Figure 6.5b. There is no obvious cluster in the data for all features whereas there are two valleys which are separated by a hill for selected features. However, there is a region where two clusters merge which makes the data not easily separable. The clusters can be seen in Figure 6.6 in 2-dimensions. The red region represents PAF patients whereas the blue region represents non-PAF subjects.



(a) (b)

Figure 6.5 Hill-valley representation of (a) all features (b) selected features. There is no obvious cluster in (a) whereas two clusters separated by a hill can be seen in (b).

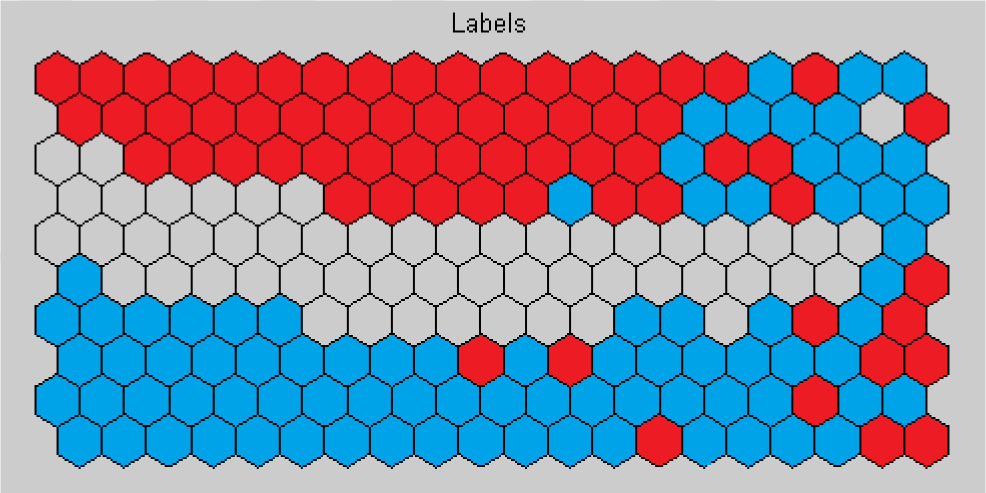


Figure 6.6 Clusters in U-matrix. Red cells represent PAF patients whereas blue cells represent non-PAF subjects.

Comparison of proposed system with other systems reported in the literature is really difficult because of the varieties in the classification techniques and datasets. But still, results show that the proposed novel system based on a combination of time domain, frequency domain, and nonlinear HRV measures with an SVM classifier performs very well in terms of sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy, and ROC area of 93%, 95%, 93%, 96%, 95%, and 0.88, respectively.

Martinez et al. utilized ECG records of 46 PAF patients and 53 healthy subjects. They got 2-hour ECG records preceding a PAF episode. They divided it into two groups: first is the one hour just prior to PAF onset and called ECG segment close to PAF episode and the second set is one hour away from the PAF episode called ECG segment away from PAF episode. Then they calculated the variability of many morphological features of P-wave. They found that using a decision tree with P-wave area and p-wave arc length achieved 95.42% accuracy to discriminate ECG segments of healthy subjects and patients suffering from PAF (includes both close to and away from PAF records) (Martínez et al., 2014).

In another work of Martinez et al., they obtained 2-hour ECG records of 24 PAF patients just preceding a PAF episode and 1-hour ECG records of 28 healthy subjects. They extracted morphological features from P-wave and obtained 91% accuracy to discriminate PAF patients and healthy subjects (Martínez et al., 2012).

Ros et al. used only the train set (50 non-PAF records and 50 records including both prior-to-PAF and distant-from-PAF records from PAF patients) of Physionet PAF Prediction Challenge Database (afpdb) (Goldberger et al., 2000). They obtained 92% correct classification rate using 22 parameters obtained from P-wave analysis (Ros et al., 2004).

Thuraisingham used the same dataset with Ros et al. and got a failure rate of 22.4% to detect PAF patients and false alarm rate of 20.4% in healthy records with wavelet decomposition of ECG signals (Thuraisingham, 2007).

The rest of the work summarized from here on used Physionet afpdb (Goldberger et al., 2000) similar to our work.

Schreier et al. used a correlation based assessment of the P-wave morphology of both regular and premature heartbeats from supraventricular origin and obtained 82% accuracy (Schreier et al., 2001).

Zong et al. developed an algorithm based upon the number and timing of the atrial premature complexes in the ECG and got 78% accuracy (Zong et al., 2001).

Maier et al. used different features obtained from HRV analysis describing the magnitude as well as the regularity of heart rate fluctuations and the number of supraventricular and ventricular premature beats. They discriminated two groups with 72% success (Maier et al., 2001).

Lynn and Chiang created feature vectors from return and difference maps of 30 minute ECG signals. They divided the maps into lattices and found the number of samples in each lattice. Then, they created vectors from those numbers and fed into k-nearest neighbor classifier which gave an accuracy of 68% (Lynn & Chiang, 2001).

Yang and Yin coded successive RR intervals as equal, accelerated, or decelerated and mapped to a single integer. Then, the histogram of those numbers was constructed for PAF patients and non-PAF subjects. The best cutoff frequency was found with receiver operator characteristics (ROC) analysis and resulted in 66% success rate (Yang & Yin, 2001).

Chazal and Heneghan examined features from the interval based power spectral density of RR intervals, time domain measures, P-wave amplitude features, and frequency representation of the P-wave. The effect of the length of the signal was also controlled by using 30-minute, 10-minute, and 5-minute windows of the ECG signals. Their best performance was obtained with power spectral density with 64% accuracy (de Chazal & Heneghan, 2001).

When compared to these previously reported works which have used the same dataset, the system developed in this study seems to be superior for the diagnosis of PAF from normal sinus rhythm ECG. As can be seen from these numbers, the best performing classifier designed in this study (SVM classifier with selected input features outperforms the others in the literature which are using the same dataset.

Moreover, the method presented in this study yields comparable performance when compared to the results of other works which have used different (their own) databases.

CHAPTER SEVEN
CONCLUSION

A new and efficient system for paroxysmal atrial fibrillation (PAF) diagnosis from normal sinus rhythm ECG records has been proposed in this thesis. Pattern recognition techniques, heart rate variability analysis, performance measures, different classification models, statistical analysis, and visualization of the dataset have been introduced.

Atrial fibrillation (AF) is an arrhythmia in which disordered electrical impulses cause atria to contract in an erratic way. As a result, the atria cannot empty the blood to ventricles completely and clot formation occurs which may lead to stroke. Paroxysmal AF is a class of AF in which the arrhythmia initiates and terminates spontaneously. PAF has no specific symptoms and the termination of the arrhythmia is so quick that it is difficult for the patient to go to a healthcare facility during a PAF episode. However, the diagnosis of PAF is currently based on the detection of arrhythmia on ECG which is very difficult to obtain. Generally, Holter monitors or event recorders are needed for diagnosis.

In this thesis, the goal is to diagnose PAF patients from their AF free ECG records. In this regard, nearly all HRV features obtained from the ECG signals of both PAF patients and non-PAF subjects were used. A total of 31 features were obtained from HRV analysis (time domain, frequency domain and nonlinear) of 5-minute RR interval time series. The numbers of atrial and ventricular ectopic beats in each 5-minute RR interval were also used as additional two features.

Four types of classifiers were designed based on k-nearest neighbor, Bayes, artificial neural network, and support vector machine networks. Parameters of the classifiers were examined to find out the ones which achieve the best discrimination. The results showed that none of these features provided satisfactory results when used alone. But using all 33 features as input to the classifiers, better results were obtained. Nevertheless, these results were still not comparable to results presented in the literature.

In order to increase the performance of the classifier systems, best discriminating features were selected with genetic algorithms. Eight features, which are mean of RR intervals, standard deviation of RR intervals, HF band peak frequency, relative power of LF band, relative power of HF band, dispersion of points perpendicular to line of identity in Poincare plot (SD1), sample entropy, and short term fluctuation slope ($\alpha1$) of detrended fluctuation analysis were selected by the genetic algorithm as the best discriminating feature subset. By using only those selected features (instead of using all features), the performances of all classifier systems were significantly increased. This is probably a result of the curse of dimensionality effect. The sensitivity, specificity and accuracy values obtained with the best performing classifier, which is the one based on SVM, were found to be 93%, 95%, 95%, sensitivity, specificity and accuracy values, respectively.

Due to the high discrimination rate and additional advantages, the system presented in this study can easily be used in health clinics, hospitals and even mobile phones for the quick and easy detection of PAF from normal sinus rhythm ECG recordings. Moreover, since the method uses only RR interval data, it could be easily modified to work with RR interval data obtained from a simple pulse meter instead of a regular ECG device.

In addition to producing high performance values, the developed system has some other advantages due to using only HRV derived features. First, the method presented in this study is relatively simple; it just uses 5-minutes RR interval data to obtain time-domain, frequency-domain and nonlinear HRV features. Second, studies with noticeable success rates have taken into account the morphological features of the ECG signal which are very sensitive to noise. In contrast, the RR-interval series data is more robust to noise. Third, using only the RR-interval series data effectively reduces the processing time when compared to the ECG based methods.

In addition, the AR-based frequency domain measures were not included in the study because of its limitation in selecting an adequate model order. Although there have been many papers on selecting an adequate model order, it is still an open question in studies related to the HRV analysis. After including these AR-based frequency-domain measures, the study may expose a higher classification power.

Although the classification results achieved in the thesis seem satisfactory, larger databases are needed to confirm the achieved results. In addition, there is a lack of demographic information such as drug use, physical activity, and emotional states, which should be considered during performing the HRV analysis. Because this information was not available in the database used, the studies covered in the thesis neglected these considerations. Also, the length of the records in the database is 30 min which allows only short-term HRV analysis. Using long-term HRV measures in addition to the short-time ones could also increase the diagnostic ability of the classifiers. Performances of the classification systems are expected to be improved further if these additional parameters are taken into account.

REFERENCES

Abarbanel, H. D. I., & Gollub, J. P. (1996). Analysis of observed chaotic data. *Physics Today*, *49* (11), 86.

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and Biological Engineering and Computing*, *44* (12), 1031–1051.

Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring. *Sports Medicine*, *33* (7), 517–538.

Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., & Luo, S. (1999). ECG beat detection using filter banks. *Biomedical Engineering, IEEE Transactions on*, *46* (2), 192–202.

Alberg, A. J., Park, J. W., Hager, B. W., Brock, M. V, & Diener-West, M. (2004). The use of “overall accuracy” to evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine*, *19* (5p1), 460–465.

Albrecht, P., & Cohen, R. J. (1988). Estimation of heart rate power spectrum bands from real-world data: dealing with ectopic beats and noisy data. In *Proceedings of Computers in Cardiology, 1988,* 311–314.

Alcaraz, R., Abásolo, D., Hornero, R., & Rieta, J. J. (2010). Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. *Computer Methods and Programs in Biomedicine*, *99* (1), 124–32.

Altman, N. S. (1992). An Introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46* (3), 175–185.

Ashcroft, S. J. H., & Pereira, C. (2003). *Practical statistics for the biological* *sciences: simple pathways to statistical analyses*. New York: Palgrave Macmillan.

Barbieri, R., Triedman, J. K., & Saul, J. P. (2002). Heart rate control and mechanical cardiopulmonary coupling to assess central volume: a systems analysis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *283* (5), 1210–1220.

Barrett, C. J., Navakatikyan, M. A., & Malpas, S. C. (2001). Long-term control of renal blood flow: what is the role of the renal nerves?. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *280* (5), 1534–1545.

Baselli, G., Cerutti, S., Civardi, S., Lombardi, F., Malliani, A., Merri, M., et al. (1987). Heart rate variability signal processing: a quantitative approach as an aid to diagnosis in cardiovascular pathologies. *International Journal of Bio-Medical Computing*, *20* (1), 51–70.

Benitez, D., Gaydecki, P. A., Zaidi, A., & Fitzpatrick, A. P. (2001). The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine*, *31* (5), 399–406.

Berne, R. M., & Levy, M. N. (1997). *Cardiovascular physiology* (7th edition). St. Louis: Mosby.

Berntson, G. G. (1997). Heart rate variability: origins, methods and interpretive caveats. *Psychophysiology*, *34*, 623–648.

Bigger, J. T., Albrecht, P., Steinman, R. C., Rolnitzky, L. M., Fleiss, J. L., & Cohen, R. J. (1989). Comparison of time-and frequency domain-based measures of cardiac parasympathetic activity in Holter recordings after myocardial infarction. *The American Journal of Cardiology*, *64* (8), 536–538.

Bigger, J. T., Fleiss, J. L., Rolnitzky, L. M., Steinman, R. C., & Schneider, W. J. (1991). Time course of recovery of heart period variability after myocardial infarction. *Journal of the American College of Cardiology*, *18* (7), 1643–1649.

Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation*, *85 (1)*, 164–171.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press, Inc.

Blanz, V., Schölkopf, B., Bülthoff, H., Burges, C., Vapnik, V., & Vetter, T. (1996). Comparison of view-based object recognition algorithms using realistic 3D models. In *Proceedins of 6th International Conference on Artificial Neural Networks,* 251–256.

Bootsma, M., Swenne, C. A., Van Bolhuis, H. H., Chang, P. C., Cats, V. M., & Bruschke, A. V. (1994). Heart rate and heart rate variability as indexes of sympathovagal balance. *American Journal of Physiology-Heart and Circulatory Physiology*, *266* (4), 1565–1571.

Braga, A. N. G., Lemos, M. D. S., Da Silva, J. R., Fontes, W. R. P., & Dos Santos, R. A. S. (2002). Effects of angiotensins on day-night fluctuations and stress-induced changes in blood pressure. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *282* (6), 1663–1671.

Bravi, A., Green, G., Herry, C., Wright, H. E., Longtin, A., Kenny, G. P., & Seely, A. J. E. (2013). Do physiological and pathological stresses produce different changes in heart rate variability?. *Frontiers in Physiology*, *4*, 197.

Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability?. *IEEE*

Transactions on Bio-Medical Engineering, *48* (11), 1342–1347.

Camm, A. J., Kirchhof, P., Lip, G. Y. H., Schotten, U., Savelieva, I., Ernst, S., et al. (2010). Guidelines for the management of atrial fibrillation. *European Heart Journal*, ehq278.

Camm, A. J., Malik, M., Bigger, J., & Breithardt, G. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17* (3), 354–381.

Carrasco, S., Gaitan, M. J., Gonzalez, R., Yanez, O. (2001). Correlation among Poincaré plot indexes and time and frequency domain measures of heart rate variability. *Journal of Medical Engineering & Technology*, *25* (6), 240–248.

Chan, Z. S. H., Ngan, H. W., Rad, A. B., & Ho, T. K. (2002). Alleviating'overfitting' via genetically-regularised neural network. *Electronics Letters*, *38* (15), 809–810.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40* (1), 16–28.

Chen, T., Han, D., Au, F. T. K., & Tham, L. G. (2003). Acceleration of Levenberg-Marquardt training of neural networks with variable decay rate. In *Proceedings of the International Joint Conference on Neural Networks, 2003,* 3, 1873–1878.

Circulatory system, (n.d.). Retrieved May 10, 2016, from http://humanbodywq.weebly.com/circulatory-system.html.

Clifford, G. D. (2002). *Signal processing methods for heart rate variability*. Phd Thesis, University of Oxford, Oxford.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20* (3), 273–297.

Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, *89* (6), 68102.

Costa, M., Goldberger, A. L., & Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Physical Review E*, *71* (2), 21906.

Cowan, C., Campbell, J., V-Lin, C., Chung, G., Fay, M., Fitzmaurice, D., & Lip, G. (2014). *Atrial Fibrillation: the management of atrial fibrillation. NICE clinical guideline 180*.

Dabiré, H., Mestivier, D., Jarnet, J., Safar, M. E., & Chau, N. P. (1998). Quantification of sympathetic and parasympathetic tones by nonlinear indexes in normotensive rats. *American Journal of Physiology - Heart and Circulatory Physiology*, *275*(4), 1290–1297.

de Chazal, P., & Heneghan, C. (2001). Automated assessment of atrial fibrillation. In *Proceedings of Computers in Cardiology 2001, 28,* 117–120.

de Jong, K. A. (1975). *Analysis of the behaviour of a class of genetic adaptive systems*. Phd thesis, University of Michigan, Michigan.

de Vos, C. B., Pisters, R., Nieuwlaat, R., Prins, M. H., Tieleman, R. G., Coelen, R.-J. S., et al. (2010). Progression from paroxysmal to persistent atrial fibrillation: clinical correlates and prognosis. *The Journal of the American College of Cardiology*, *55* (8), 725–731.

Deboer, R. W., Karemaker, J. M., & Strackee, J. (1984). Comparing spectra of a series of point events particularly for heart rate variability data. *Biomedical Engineering, IEEE Transactions on, 4,* 384–387.

Dokur, Z., Olmez, T., Yazgan, E., & Ersoy, O. K. (1997). Detection of ECG waveforms by neural networks. *Medical Engineering & Physics*, *19* (8), 738–741.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: John Wiley & Son.

Duprez, D. A., De Sutter, J. H., De Buyzere, M. L., Rietzschel, E. R., Rimbaut, S., Kaufman, J. M., et al. (1995). Renin-angiotensin-aldosterone system, RR interval, and blood pressure variability during postural changes in borderline arterial hypertension. *American Journal of Hypertension*, *8* (7), 683–688.

Eberhart, R. C. (2014). *Neural network PC tools: a practical guide*. San Diego: Academic Press.

Eckberg, D. L. (1997). Sympathovagal balance a critical appraisal. *Circulation*, *96* (9), 3224–3232.

Evans, W. (1959). The management of paroxysmal atrial fibrillation. *Progress in Cardiovascular Diseases*, *60* (2), 480–484.

Ewing, D. J., Martyn, C. N., Young, R. J., & Clarke, B. F. (1985). The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetes Care*, *8* (5), 491–498.

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

Fleisher, L. A., Frank, S. M., Sessler, D. I., Cheng, C., Matsukawa, T., & Vannier, C. A. (1996). Thermoregulation and heart rate variability. *Clinical Science*, *90* (2), 97–103.

Fletcher, R. H., Fletcher, S. W., & Wagner, E. H. (2005). *Clinical epidemiology: the essentials*. Baltimore: Lippincott Williams & Wilkins.

Francis, D. P., Davies, L. C., Willson, K., Ponikowski, P., Coats, A. J. S., & Piepoli,

M. (2000). Very-low-frequency oscillations in heart rate and blood pressure in periodic breathing: role of the cardiovascular limb of the hypoxic chemoreflex. *Clinical Science*, *99* (2), 125–132.

Friberg, L., Hammar, N., & Rosenqvist, M. (2010). Stroke in paroxysmal atrial fibrillation: report from the Stockholm Cohort of Atrial Fibrillation. *European Heart Journal*, *31* (8), 967–975.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. San Diego: Academic Press.

Furberg, C. D., Psaty, B. M., Manolio, T. A., Gardin, J. M., Smith, V. E., Rautaharju, P. M., et al. (1994). Prevalence of atrial fibrillation in elderly subjects (the Cardiovascular Health Study). *The American Journal of Cardiology*, *74* (3), 236–241.

Fuster, V., Rydén, L. E., Cannom, D. S., Crijns, H. J., Curtis, A. B., Ellenbogen, K. A., et al. (2006). ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice . *The Journal of the American College of Cardiology*, *48* (4), 149–246.

Gallagher, M. M., & Camm, J. (1998). Classification of atrial fibrillation. *The American Journal of Cardiology*, *82* (7), 18–28.

Gertsch, M. (2003). *The ECG: a two-step approach to diagnosis*. Berlin: Springer-Verlag.

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.

Gladstone, D. J., Spring, M., Dorian, P., Panzov, V., Thorpe, K. E., Hall, J., et al. (2014). Atrial fibrillation in patients with cryptogenic stroke. *New England Journal of Medicine*, *370* (26), 2467–2477.

Go, A., Hylek, E., Phillips, K., & Chang, Y. (2001). Prevalence of diagnosed atrial fibrillation in adults - National implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Journal of the American Medical Association*, *285* (18), 2370 – 2375.

Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning*. Menlo Park: Addison-Wesley Reading.

Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, *1*, 69–93.

Goldberger, A., Amaral, L., Glass, L., & Hausdorff, J. (2000). PhysioBank, PhysioToolkit, and PhysioNet - Components of a new research resource for complex physiologic signals. *Circulation*, *101* (23), 215–220.

Goldstein, B., Fiser, D. H., Kelly, M. M., Mickelsen, D., Ruttimann, U., & Pollack, M. M. (1998). Decomplexification in critical illness and injury: relationship between heart rate variability, severity of illness, and outcome. *Critical Care Medicine*, *26* (2), 352–357.

Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters*, *50*, 346–349.

Greenwood, J. P., Batin, P. D., & Nolan, J. (1997). Assessment of cardiac autonomic function. *British Journal of Cardiology*, *4*, 154–158.

Greenwood, J. P., Durham, N. P., & Nolan, J. (1998). Autonomic assessment of cardiovascular disease. *Hospital Medicine, 59* (9), 714–718.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*, 1157–1182.

Guzik, P., Piskorski, J., Krauze, T., Schneider, R., Wesseling, K. H., Wykretowicz, A., & Wysocki, H. (2007). Correlations between the Poincare plot and conventional heart rate variability parameters assessed during paced breathing. *The Journal of Physiological Sciences*, *57* (1), 63–71.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *Neural Networks, IEEE Transactions on*, *5* (6), 989–993.

Hagiwara, K., & Kuno, K. (2000). Regularization learning and early stopping in linear networks. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks,* 4, 511–516.

Hart, R. G., Pearce, L. A., Rothbart, R. M., McAnulty, J. H., Asinger, R. W., & Halperin, J. L. (2000). Stroke with intermittent atrial fibrillation: incidence and predictors during aspirin therapy. *Journal of the American College of Cardiology*, *35* (1), 183–187.

Hassanien, A. E., Kim, T.-H., Kacprzyk, J., & Awad, A. I. (2014). *Bio-inspiring cyber security and cloud services: trends and innovations*. Berlin: Springer-Verlag.

Haykin, S. (1999). *Neural networks*. India: Prentice Hall.

Heist, E. K., Mansour, M., & Ruskin, J. N. (2011). Rate control in atrial fibrillation targets, methods, resynchronization considerations. *Circulation*, *124* (24), 2746–2755.

Henry, B., Lovell, N., & Camacho, F. (2012). Nonlinear dynamics time series analysis. *Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling, 2*, 1–39.

Hinojosa-Laborde, C., Rickards, C. A., Ryan, K. L., & Convertino, V. A. (2011). Heart rate variability during simulated hemorrhage with lower body negative pressure in high and low tolerant. *Frontiers in Physiology*, *2,* 85.

Hohnloser, S. H., Pajitnev, D., Pogue, J., Healey, J. S., Pfeffer, M. A., Yusuf, S., & Connolly, S. J. (2007). Incidence of stroke in paroxysmal versus sustained atrial fibrillation in patients taking oral anticoagulation or combined antiplatelet therapy: an ACTIVE W Substudy. *Journal of the American College of Cardiology*, *50* (22), 2156–2161.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* Cambridge: MIT Press.

Hon, E. H., & Lee, S. T. (1963). Electronic evaluation of the fetal heart rate patterns preceding fetal death, further observations. *American Journal of Obstetrics and Gynecology*, *87*, 814–826.

Hoshino, T., Ishizuka, K., Nagao, T., Shimizu, S., & Uchiyama, S. (2013). Slow sinus heart rate as a potential predictive factor of paroxysmal atrial fibrillation in stroke patients. *Cerebrovascular Diseases*, *36* (2), 120–125.

Hu, K., Ivanov, P. C., Chen, Z., Carpena, P., & Stanley, H. E. (2001). Effect of trends on detrended fluctuation analysis. *Physical Review E*, *64* (1), 11114.

Huikuri, H. V, & Stein, P. K. (2012). Clinical application of heart rate variability after acute myocardial infarction. *Frontiers in Physiology*, *3,* 41.

Huikuri, H. V., Mäkikallio, T. H., & Perkiömäki, J. (2003). Measurement of heart rate variability by methods based on nonlinear dynamics. *Journal of Electrocardiology*, *36*, 95–99.

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

Inan, Z. H., & Kuntalp, M. (2007). A study on fuzzy C-means clustering-based systems in automatic spike detection. *Computers in Biology and Medicine*, *37* (8), 1160–6.

Işler, 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*, *37* (10), 1502–1510.

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

Jimenez, L. O., & Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *28* (1), 39–54.

Kadambe, S., Murray, R., & Boudreaux-Bartels, G. F. (1999). Wavelet transform-based QRS complex detector. *Biomedical Engineering, IEEE Transactions on*, *46* (7), 838–848.

Kamath, M. V, & Fallen, E. L. (1995). Correction of the heart rate variability signal for ectopics and missing beats. *Heart Rate Variability*, 75–85.

Kamen, P. W., Krum, H., & Tonkin, A. M. (1996). Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. *Clinical Science*, *91* (2), 201–208.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., & Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and Its Applications*, *295* (3), 441–454.

Karmakar, C. K., Khandoker, A. H., Gubbi, J., & Palaniswami, M. (2009). Complex correlation measure: a novel descriptor for Poincaré plot. *Biomedical Engineering Online*, *8* (1), 17.

Kato, T., Yamashita, T., Sagara, K., Iinuma, H., & Fu, L. (2004). Progressive nature of paroxysmal atrial fibrillation. *Circulation Journal*, *68* (6), 568–572.

Kerr, C. R., Humphries, K. H., Talajic, M., Klein, G. J., Connolly, S. J., Green, M., et al. (2005). Progression to chronic atrial fibrillation after the initial diagnosis of paroxysmal atrial fibrillation: results from the Canadian Registry of Atrial Fibrillation. *American Heart Journal*, *149* (3), 489–496.

Keselbrener, L., & Akselrod, S. (1998). Autonomic responses to blockades and provocations. In M. Malik, (Ed.). *Clinical guide to cardiac autonomic tests* (101-148). London: Kluwer Academic Publishers.

Keselbrener, L., Keselbrener, M., & Akselrod, S. (1997). Nonlinear high pass filter for R-wave detection in ECG signal. *Medical Engineering & Physics*, *19* (5), 481–484.

Klabunde, R. (2011). *Cardiovascular physiology concepts*. New York: Lippincott Williams & Wilkins.

Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, *59* (4), 256–262.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43* (1), 59–69.

Kohonen, T. (2001). *Self-Organizing Maps*. Berlin: Springer.

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science*, *1* (2), 111–117.

Laitio, T., Jalonen, J., Kuusela, T., & Scheinin, H. (2007). The role of heart rate variability in risk stratification for adverse postoperative cardiac events. *Anesthesia & Analgesia*, *105* (6), 1548–1560.

Langley, P., di Bernardo, D., Allen, J., Bowers, E., Smith, F. E., Vecchietti, S., & Murray, A. (2001). Can paroxysmal atrial fibrillation be predicted? In *Proceedings of Computers in Cardiology 2001, 28,* 121–124.

Lévy, S., Camm, A. J., Saksena, S., Aliot, E., Breithardt, G., Crijns, H., et al. (2003). International consensus on nomenclature and classification of atrial fibrillation. *Europace*, *5* (2), 119–122.

Li, C., Zheng, C., & Tai, C. (1995). Detection of ECG characteristic points using wavelet transforms. *Biomedical Engineering, IEEE Transactions on*, *42* (1), 21–28.

Lip, G. Y. H., & Li Saw Hee, F. (2001). Paroxysmal atrial fibrillation. *QJM: an international journal of medicine*, *94* (12), 665–678.

Lip, G. Y. H., & Tse, H. (2007). Management of atrial fibrillation. *Lancet*, *370* (9587), 604–18.

Lip, G. Y. H. (1999). How would i manage a 60-year-old woman presenting with atrial fibrillation? *Proceedings of Royal College of Physicians of Edinburgh*, *29*, 301–306.

Lippman, N., Stein, K. M., & Lerman, B. B. (1994). Comparison of methods for removal of ectopy in measurement of heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology*, *267* (1), 411–418.

Lynn, K., & Chiang, H. (2001). A two stage solution algorithm for paroxymal atrial fibrillation prediction. In *Proceedings of Computers in Cardiology 2001, 28,* 405–407.

Maier, C., Bauch, M., & Dickhaus, H. (2001). Screening and prediction of paroxymal atrial fibrillation by analysis of heart rate variability parameters. In *Proceedings of Computers in Cardiology 2001, 28,* 129–132.

Majeed, A., Moser, K., & Carroll, K. (2001). Trends in the prevalence and management of atrial fibrillation in general practice in England and Wales, 1994-1998: analysis of data from the general practice research database. *Heart*, *86* (3), 284–288.

Malik, M., & Camm, A. J. (1993). Components of heart rate variability: what they really mean and what we really measure. *The American Journal of Cardiology*, *72* (11), 821–822.

Malik, M., & Camm, A. J. (1995). *Heart rate variability*. New York: Armonk, Futura Pub. Co. Inc.

Malik, M., Farrell, T., Cripps, T., & Camm, A. J. (1989). Heart rate variability in relation to prognosis after myocardial infarction: selection of optimal processing techniques. *European Heart Journal*, *10* (12), 1060–1074.

Malpas, S. C. (2002). Neural influences on cardiovascular variability: possibilities and pitfalls. *American Journal of Physiology-Heart and Circulatory Physiology*, *282* (1), 6–20.

Martínez, A., Alcaraz, R., & Rieta, J. J. (2012). Study on the P-wave feature time course as early predictors of paroxysmal atrial fibrillation. *Physiological Measurement*, *33* (12), 1959–74.

Martínez, A., Alcaraz, R., & Rieta, J. J. (2014). Morphological variability of the P-wave for premature envision of paroxysmal atrial fibrillation events. *Physiological Measurement*, *35* (1), 1–14.

Mateo, J., & Laguna, P. (2000). Improved heart rate variability signal analysis from the beat occurrence times according to the IPFM model. *Biomedical Engineering, IEEE Transactions on*, *47* (8), 985–996.

Mitchell, T. M. (1997). *Machine learning*. Boston: McGraw-Hill Companies Press.

Montano, N., Ruscone, T. G., Porta, A., Lombardi, F., Pagani, M., & Malliani, A. (1994). Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt. *Circulation*, *90* (4), 1826–1831.

Moody, G., Goldberger, A., McClennen, S., & Swiryn, S. (2001). Predicting the onset of paroxysmal atrial fibrillation: the Computers in Cardiology Challenge 2001. In *Proceedings of Computers in Cardiology 2001, 28,* 113–116.

Naccarelli, G. V, Varker, H., Lin, J., & Schulman, K. L. (2009). Increasing prevalence of atrial fibrillation and flutter in the United States. *The American Journal of Cardiology*, *104* (11), 1534–1539.

Nattel, S., & Harada, M. (2014). Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *Journal of the American College of Cardiology*, *63* (22), 2335–45.

Nixon, M. S., & Aguado, A. S. (2008). *Feature extraction & image processing* (2nd edition). San Diego: Academic Press.

Oja, M., Nikkila, J., Törönen, P., Wong, G., Castrén, E., & Kaski, S. (2006). Exploratory clustering of gene expression profiles of mutated yeast strains. *Computational and Statistical Approaches to Genomics*, 61–74.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: an application to face detection. In *Proceedings of IEEE Computer Society Conference on Computer vision and Pattern Recognition,* 130–136.

Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., et al. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, *59* (2), 178–193.

Pagani, M., Montano, N., Porta, A., Malliani, A., Abboud, F. M., Birkett, C., & Somers, V. K. (1997). Relationship between spectral components of cardiovascular variabilities and direct measures of muscle sympathetic nerve activity in humans. *Circulation*, *95* (6), 1441–1448.

Page, R. L., Wilkinson, W. E., Clair, W. K., McCarthy, E. a, & Pritchett, E. L. (1994). Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. *Circulation*, *89* (1), 224–227.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *Biomedical Engineering, IEEE Transactions on*, (3), 230–236.

Papaioannou, V., Pneumatikos, I., & Maglaveras, N. (2013). Association of heart rate variability and inflammatory response in patients with cardiovascular diseases: current strengths and limitations. *Frontiers in Physiology*, *4:174*.

Peltola, M. A. (2012). Role of editing of R-R intervals in the analysis of heart rate variability. *Frontiers in Physiology*, 1–10.

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, *5* (1), 82–87.

Perkiömäki, J. S. (2011). Heart rate variability and non-linear dynamics in risk stratification. *Frontiers in Physiology*, *2,* 81.

Pernkopf, F. (2007). Tracking of multiple targets using on-line learning for appearance model adaptation. *Image Analysis and Recognition*, 602–614.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, *88* (6), 2297–2301.

Pincus, S. M., & Goldberger, A. L. (1994). Physiological time-series analysis: what does regularity quantify?. *American Journal of Physiology-Heart and Circulatory Physiology*, *266* (4), 1643–1656.

Ponikowski, P., Chua, T. P., Piepoli, M., Amadi, A. A., Harrington, D., Webb-Peploe, K., et al. (1997). Chemoreceptor dependence of very low frequency rhythms in advanced chronic heart failure. *American Journal of Physiology-Heart and Circulatory Physiology*, *272* (1), 438–447.

Porter, G. A., & Rivkees, S. A. (2001). Ontogeny of humoral heart rate regulation in the embryonic mouse. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *281* (2), 401–407.

Pumprla, J., Howorka, K., Groves, D., Chester, M., & Nolan, J. (2002). Functional assessment of heart rate variability: physiological basis and practical applications. *International Journal of Cardiology*, *84* (1), 1–14.

Purves, D., Augustine, G. J., Fitzpatric, D., Hall, W. C., LaMantia, A.-S., & White, L. E. (2012). Autonomic control of cardiovascular function. In R. D. Mooney & M. L. Platt (Eds.), *Neuroscience* (5th ed.). Sunderland: Sinauer Associates, Inc.

Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, *4* (2), 164–171.

Rentero, N., Cividjian, A., Trevaks, D., Pequignot, J. M., Quintin, L., & McAllen, R. M. (2002). Activity patterns of cardiac vagal motoneurons in rat nucleus ambiguus. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *283* (6), 1327–1334.

Richman, J., & Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278* (6), 2039–2049.

Rompelman, O. (1993). Rhythms and analysis techniques. *The Physics of Heart and Circulation*, 101–120.

Ros, E., Mota, S., Fernández, F. J., Toro, F. J., & Bernier, J. L. (2004). ECG characterization of paroxysmal atrial fibrillation: parameter extraction and automatic diagnosis algorithm. *Computers in Biology and Medicine*, *34* (8), 679–96.

Salo, M. A., Huikuri, H. V, & Seppanen, T. (2001). Ectopic beats in heart rate variability analysis: effects of editing on time and frequency domain measures. *Annals of Noninvasive Electrocardiology*, *6* (1), 5–17.

Sanna, T., Diener, H.-C., Passman, R. S., Di Lazzaro, V., Bernstein, R. A., Morillo, C. A., et al. (2014). Cryptogenic stroke and underlying atrial fibrillation. *New England Journal of Medicine*, *370* (26), 2478–2486.

Savelieva, I., & Camm, A. J. (2000). Clinical relevance of silent atrial fibrillation: prevalence, prognosis, quality of life, and management. *Journal of Interventional Cardiac Electrophysiology*, *4* (2), 369–382.

Schreier, G., Kastner, P., & Marko, W. (2001). An automatic ecg processing algorithm to identify patients prone to paroxymal atrial fibrillation. In *Proceedings of Computers in Cardiology 2001, 28,* 133–135.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press IL.

Sherwood, L. (2015). *Human physiology: from cells to systems*. Cengage learning.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, *10* (5), 335–347.

Sleight, P., La Rovere, M. T., Mortara, A., Pinna, G., Maestri, R., Leuzzi, S., et al. (1995). Physiology and pathophysiology of heart rate and blood pressure variability in humans: is power spectral analysis largely an index of baroreflex gain?. *Clinical Science*, *88* (1), 103–109.

Spruyt, V. (2014). *The curse of dimensionality in classification.* Retrieved 20 April, 2016, from http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/#top.

Suárez, K. V, Silva, J. C., Berthoumieu, Y., Gomis, P., & Najim, M. (2007). ECG beat detection using a geometrical matching approach. *Biomedical Engineering, IEEE Transactions on*, *54* (4), 641–650.

Tarkiainen, T. H., Kuusela, T. A., Tahvanainen, K. U. O., Hartikainen, J. E. K., Tiittanen, P., Timonen, K. L., & Vanninen, E. J. (2007). Comparison of methods for editing of ectopic beats in measurements of short-term non-linear heart rate dynamics. *Clinical Physiology and Functional Imaging*, *27* (2), 126–133.

Tarvainen, M. P., & Niskanen, J.-P. (2008). *Kubios HRV (version 2.0), user's guide*.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A, Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV-heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113* (1), 210–220.

Tarvainen, M. P., Ranta-Aho, P. O., & Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. *IEEE Transactions on Bio-Medical Engineering*, *49* (2), 172–175.

Taylor, J. A., Carr, D. L., Myers, C. W., & Eckberg, D. L. (1998). Mechanisms underlying very-low-frequency RR-interval oscillations in humans. *Circulation*, *98* (6), 547–555.

Thuraisingham, R. (2007). An electrocardiogram marker to detect paroxysmal atrial fibrillation. *Journal of Electrocardiology*, *40* (4), 344–347.

Thuraisingham, R. A. (2006). Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. *Computer Methods and Programs in Biomedicine*, *83* (1), 78–82.

Tobaldini, E., Nobili, L., Strada, S., Casali, K. R., Braghiroli, A., & Montano, N. (2015). Heart rate variability in normal and pathological sleep. *Heart Rate Variability: Clinical Applications and Interaction Between HRV and Heart Rate*, *16*, 93.

Todd, W. (2013). *Conduction Anatomy.* Retrieved April 20, 2016, from http://wesleytodd.blogspot.com.tr/2013/06/triangle-koch.html.

Ultsch, A. (1993). *Self-organizing neural networks for visualisation and classification*. Berlin: Springer.

Ultsch, A. (2003). Maps for the visualization of high-dimensional data spaces. In *Proceedings of Workshop on Self organizing Maps*, 225–230.

Valafar, F. (2000). Applications of Neural Networks in Medicine and Biological Sciences. In A. Zilouchian & M. Jamshidi (Eds.), *Intelligent Control Systems Using Soft Computing Methodologies*. Boca Raton: CRC Press, Inc.

Van Erkel, A. R., & Peter, M. (1998). Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *European Journal of Radiology*, *27* (2), 88–94.

Van Gelder, I. C., & Hemels, M. E. W. (2006). The progressive nature of atrial fibrillation: a rationale for early restoration and maintenance of sinus rhythm. *Europace : European Pacing, Arrhythmias, and Cardiac Electrophysiology : Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology*, *8* (11), 943–949.

van Ravenswaaij-Arts, C. M. A., Kollee, L. A. A., Hopman, J. C. W., Stoelinga, G. B. A., & van Geijn, H. P. (1993). Heart rate variability. *Annals of Internal Medicine*, *118* (6), 436–447.

Vapnik, V. (2006). *Estimation of dependences based on empirical data*. Berlin: Springer-Verlag.

Verleysen, M., & François, D. (2005). The curse of dimensionality in data mining and time series prediction. *Computational Intelligence and Bioinspired Systems*, 758–770.

Vijaya, G., Kumar, V., & Verma, H. K. (1998). ANN-based QRS-complex analysis of ECG. *Journal of Medical Engineering & Technology*, *22* (4), 160–167.

Wang, L., Liu, B., & Wan, C. (2005). Classification using support vector machines with graded resolution. In *Granular Computing, 2005 IEEE International Conference on*, 2, 666–670.

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

Webber, C., & Zbilut, J. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, *76* (2), 965–973.

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

Webster, J. G. (1995). *Design of cardiac pacemakers*. New York: IEEE Press.

Williams, T. D., Chambers, J. B., Henderson, R. P., Rashotte, M. E., & Overton, J. M. (2002). Cardiovascular responses to caloric restriction and thermoneutrality in C57BL/6J mice. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *282* (5), 1459–1467.

Wolf, M. M., Varigos, G. A., Hunt, D., & Sloman, J. G. (1978). Sinus arrhythmia in acute myocardial infarction. *The Medical Journal of Australia*, *2* (2), 52–53.

Xu, X., & Liu, Y. (2004). ECG QRS complex detection using slope vector waveform (SVW) algorithm. In Proceedings of *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 2, 3597–3600.

Yang, A., & Yin, H. (2001). Prediction of paroxysmal atrial fibrillation by footprint analysis. In *Proceedings of Computers in Cardiology 2001, 28,* 401–404.

Yanowitz, F. G. (2012). *Introduction to ECG interpretation V8. 0.* Salt Lake City: Intermountain Healthcare.

Yeh, Y.-C., & Wang, W.-J. (2008). QRS complexes detection for ECG signal: The Difference Operation Method. *Computer Methods and Programs in Biomedicine*, *91* (3), 245–254.

Yeragani, V. K., Sobolewski, E., Jampala, V. C., Jerald, K. A. Y., Yeragani, S., & Gina, I. (1998). Fractal dimension and approximate entropy of heart period and heart rate: awake versus sleep differences and methodological issues. *Clinical Science*, *95* (3), 295–301.

Yu, C.-C., & Liu, B.-D. (2002). A backpropagation algorithm with adaptive learning rate and momentum coefficient. In *Proceedings of the 2002 International Joint Conference on Neural Networks,* 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.

Zong, W., Mukkamala, R., & Mark, R. G. (2001). A methodology for predicting paroxymal atrial fibrillation based on ECG arryhythmia feature analysis. In *Proceedings of Computers in Cardiology 2001, 28,* 125–128.

Zuern, C. S., Barthel, P., & Bauer, A. (2011). Heart rate turbulence as risk-predictor after myocardial infarction. *Frontiers in Physiology*, *2*, 99.

APPENDIX
STATISTICAL ANALYSIS

Hypothesis Testing

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. There are four general steps to test (Ashcroft & Pereira, 2003):

1. Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_1 (commonly, that the observations show a real effect combined with a component of chance variation).

2. Identify a test statistic that can be used to assess the truth of the null hypothesis.

3. Compute the *p*-value, which is the probability that a statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller *p*-value, the stronger the evidence against the null hypothesis.

4. Compare the *p*-value to an acceptable significance value, α. If $p \leq \alpha$, that the observed effect is statistically significant, and the alternative hypothesis is valid. The most commonly used significance level is $\alpha = 0.05$ (for a two-sided test, α/2).

Hypotheses are generally defined as:

$$H_0: \mu_1 = \mu_2 \tag{A.1}$$

and

$$H_1: \mu_1 \neq \mu_2 \tag{A.2}$$

where $\mu_1$and $\mu_2$are the mean values of two groups.

General t-Testing

A t-test is a statistical hypothesis test in which the test statistic has a Student's t distribution if the null hypothesis is true. It is applied when the population is assumed to be normally distributed but the sample sizes are small enough that the statistic on which inference is based is not normally distributed because it relies on an uncertain estimate of standard deviation rather than on a precisely known value. The general formula for t-test is as follows:

$$t = \frac{(Statistics) - (Hypothesized\ \ value\ of\ parameter)}{Estimated\ standard\ error\ of\ statistics} \tag{A.3}$$

Among the t tests, the unpaired t-Test with Unequal Variance method, which is given in the following subsection, is used in the thesis.

Unpaired t-Test with Unequal Variance

The difference from the previous test is that the variance in two samples is extremely different, meaning samples are very different in size. The two sample t test for unpaired data is defined as:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{A.4}$$

where n_1 and n_2 are the sample sizes, $\overline{x_1}$ and $\overline{x_2}$ are the sample means, s_1^2 and s_2^2 are the sample variances. It assumes that the degree of freedom (*d.f.*) is calculated by:

$$d.f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \qquad \text{(A.5)}$$

where *d.f.* is rounded to an integer value after this calculation is completed.