ISLAMIC UNIVERSITY OF TECHNOLOGY, (IUT)

NEURAL NETWORK BASED ECG ARRHYTHMIA PERFORMANCE OPTIMIZATION

By

Jamshed Karim (092453)

Md. Shajjadul Abedin (092445)

Yasir Hoque (092423)

A Dissertation

Submitted in Partial Fulfillment of the Requirement for the

Bachelor of Science in Electrical and Electronic Engineering

Academic Year: 2012-2013

Department of Electrical and Electronic Engineering

Islamic University of Technology (IUT)

A Subsidiary Organ of OIC

Dhaka, Bangladesh.

A Dissertation on,

NEURAL NETWORK BASED ECG ARRHYTHMIA PERFORMANCE OPTIMIZATION

Submitted By

Jamshed Karim (092453) Md. Shajjadul Abedin (092445) Yasir Hoque (092423)

Approved By

Dr. Md. Shahid Ullah

Professor & Head of the Department of EEE, IUT.

Dr. Mohammad Rakibul Islam

Professor

Thesis Supervisor

Department of EEE, IUT

ACKNOWLEDGEMENT

Every honor on earth is due to the Great Almighty, descended from Him and must be ascribed to Him. He has given me the capability to do this work with good health. This thesis is a result of one year and this is by far the most significant scientific accomplishment in my life. It would be impossible without support and appreciation of those who mattered most.

I would like to thank my supervisor, Dr. Mohammad Rakibul Islam for his continuous guidance, inspiration and enthusiasm during the progress of the work. He has always been generous with his time, listening carefully and criticizing fairly.

I am also grateful to the Head of the Department Prof. Dr. Md. Shahid Ullah for his inspirations throughout my career and encouragement to complete the work.

Last but not the least I am thankful to my respected colleagues, family, friends and relatives for their non-scientific support over the whole time of my work.

Abstract

Any disturbance in the activity of heart that can cause irregular heart rhythm is known as cardiac arrhythmia. Electrocardiogram (ECG) is one of the most promising tools for detecting different types of arrhythmia, which is necessary until it goes fatal and causes loss of life. For detecting ECG arrhythmia Different techniques have been used till now. But one problem that most of the detection methods have faced is the trade-off between speed and accuracy. So in this thesis work we have tried to create a bridge among these two. We have tried to maintain a good processing speed to detect the ECG abnormality by maintaining a good accuracy rate. In this thesis work we have mainly chosen Neural Network based algorithm to detect arrhythmia but before choosing that we have also worked with two other techniques which are Wavelet packet transform and Non-linear based model. In the detection of the abnormality of ECG beat Leverberg-Marquardt algorithm in the Neural Network model has been used as it is the mostly appraised algorithm for its high processing speed. From this thesis work, the final result creates a balance between the speed and accuracy in the detection of ECG arrhythmia which can be improved in future through more research. So this Neural Network based thesis work comes with an option of choosing a method which balances the detection speed and accuracy as well as comes with a great scope of improvement.

Contents

Chapter1 Introduction

Chapter 2 Wavelet based packet ECG arrhythmia classification

Chapter 3 Non-linear analysis of ECG arrhythmia detection and classification

Chapter 4 Neural Network

Chapter 5 Conclusion

List of figures

List of Abbreviations

ECG Electrocardiogram

LFD local fractal dimension

EMD Empirical mode Decomposition

IMF Intrinsic Mode function

WPD Wavelet Packet Decomposition

VFE Variance Based Fractal Dimension Estimation

PSDFE Power Spectrum Density Based Fractal Dimension Estimator

ROC Receiver Operating Characteristics

GSI Geometric Separability Index

BD Bhattacharya distance

Chapter 1

Introduction

The term arrhythmia suggests abnormalities in the heart rate, although it is commonly used to refer to changes in the shape of the heart wave. The electrocardiogram (ECG) is the most important bio-signal used by cardiologists for diagnostic purposes. The ECG signal, if properly analyzed, provides key information about the electrical activity of the heart. Life threatening situations, signaled by the occurrence of premature ventricular contractions (PVC), ventricular fibrillation (VF) or other serious arrhythmias, can often be reversed if detected in time for proper treatment. However, ECG being a non-stationary signal, the irregularities may not be periodic and may not show up all the time, but would manifest at certain irregular intervals during the day. So continuous ECG monitoring permits observation of cardiac variations over an extended period of time, either at the bedside or when patients are ambulatory, providing more information to physicians. Thus, continuous monitoring increases the understanding of patients' circumstances and allows more reliable diagnosis of cardiac abnormalities. Detection of abnormal ECG signals is a critical step in administering aid to patients. Often, patients are hooked up to cardiac monitors in hospital continuously. This requires continuous monitoring by the physicians. Since clinical observation of ECG can take long hours and can be very tedious, the cardiac intensive care unit has been incorporated into hospitals in order to improve the chances of survival of and to treat patients suffering traumatic episodes of heart diseases. Modern era of medical science is facilitated by computer aided feature extraction and disease diagnostics in which various signal processing techniques have been utilized in extracting different kinds of features from the ECG signals. The objective of computer aided digital signal processing of ECG signal is to reduce the time taken by the cardiologists in interpreting the results. Due to the large number of patients in intensive care units, the need for continuous observation of them and the high possibility of the analyst missing the vital information by visual

analysis, over the years researchers have developed a variety of relatively effective signal processing techniques in time or frequency or time-frequency domain in order to classify ECG arrhythmia accurately. However, the problem of classifying different types of ECG arrhythmia still poses a challenge to the area of signal processing. In a particular arrhythmia condition, morphological changes occur in different sections of a normal ECG beat and such changes vary from beat to beat under the same arrhythmia. Thus extracting the very detail characteristics of each arrhythmia through signal processing techniques into a feature vector capable of correctly classifying among different types of ECG arrhythmia is difficult task. Complexity and ease of implementation of the cardiac arrhythmia classification methods is also of concern in real life applications. The overall goal of arrhythmia classification technique is to find a simple yet effective method capable of performing the classification task with greater sensitivity, specificity and accuracy.

1.1 Electrocardiogram:

Electrocardiogram (ECG) is diagnosis tool that reported the electrical activity of heart recorded by skin electrode. The morphology and heart rate reflects the cardiac health of human heart beat. It is a noninvasive technique that means this signal is measured on the surface of human body, which is used in identification of the heart diseases. Any disorder of heart rate or rhythm, or change in the morphological pattern, is an indication of cardiac arrhythmia, which could be detected by analysis of the recorded ECG waveform. The amplitude and duration of the P-QRS-T wave contains useful information about the nature of disease afflicting the heart. The electrical wave is due to depolarization and repolarization of Na ions and K ions in the blood. The ECG signal provides the following information of human heart:

a. Heart position and its relative chamber size
b. Impulse origin and propagation
c. Heart rhythm and conduction disturbances
d. Extent and location of myocardial concentrations
e. Drug-effects on the heart.

ECG does not afford data on cardiac contraction or pumping function.

1.2 The Heart Anatomy:

The heart contains four chambers that is right atrium, left atrium, right ventricle, left ventricle and several atrioventricular and Sino atrial node as shown in the Fig. 1.1. The two upper chambers are called the left and right atria, while the lower two chambers are called the left and right ventricles. The atria are attached to the ventricles by fibrous, non-conductive tissue that keeps the ventricles electrically isolated from the atria. The heart conduction system is shown in Fig.1.2. The right atrium and the right ventricle together form a pump to the circulate blood to the lungs. Oxygen poor blood is received through large veins called the superior and inferior vena cava and flows into the right atrium. The right ventricle then pumps the blood to the lungs where the blood is oxygenated. Similarly, the left atrium and the left ventricle together form a pump to circulate oxygen-enriched blood received from the lungs to the rest of the body.

In heart Sino-atrial (S-A) node spontaneously generates regular electrical impulses, which then spread through the conduction system of the heart and initiate contraction of the myocardium. Propagation of an electrical impulse through excitable tissue is achieved through a process called depolarization. Depolarization of the heart muscles collectively generates a strong ionic current. This current flows through the resistive body tissue generating a voltage drop. The magnitude of the voltage drop is sufficiently large to be detected by electrodes attached to the skin. ECGs are thus recordings of voltage drops across the skin caused by ionic current flow generated from myocardial depolarization. Atrial depolarization results in the spreading of the electrical impulse throughout the ventricular myocardium.

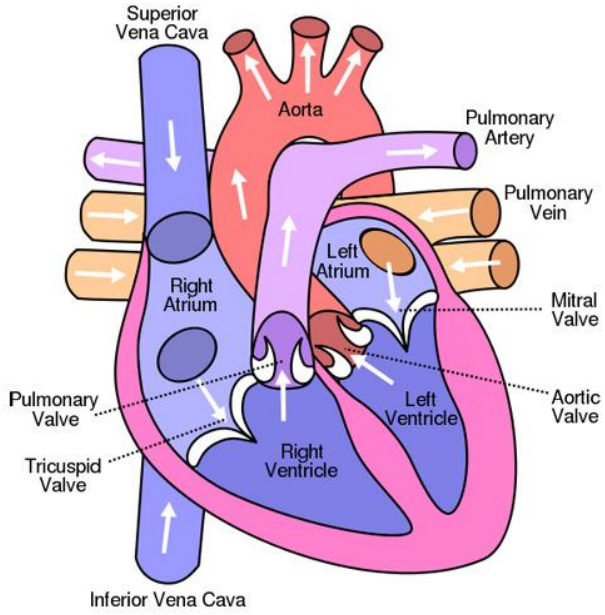

Fig. 1.1: A diagrammatic structure of human heart

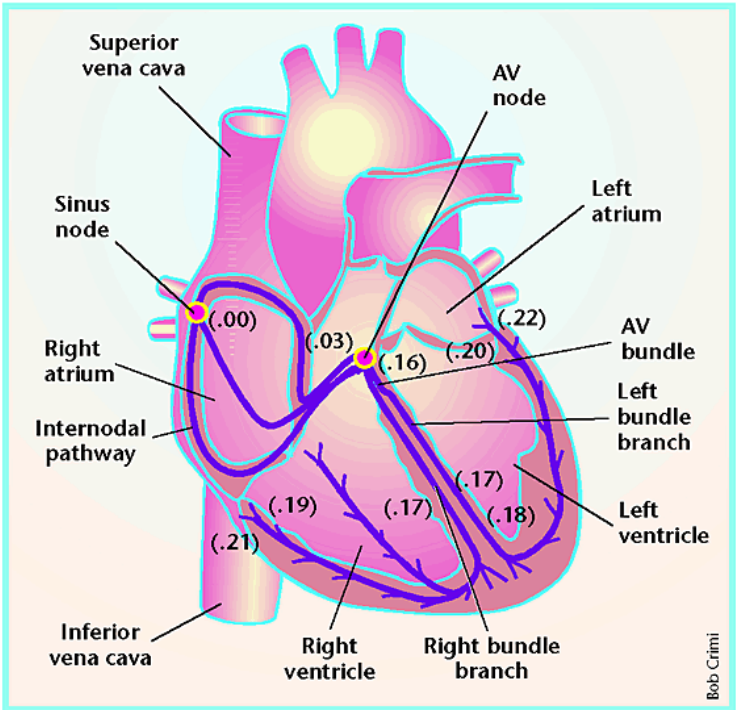

Fig. 1.2: Heart conduction system

1.3 Leads in ECG:

The standard ECG has 12 leads which include 3 bipolar leads, 3 augmented unipolar leads and 6 chest leads. In Fig.1.3 all leads' position are shown in human body. A lead is a pair of electrodes

(+ve&-ve) placed on the body in designated anatomical locations & connected to an ECG record. Bipolar leads record the potential difference between two points (+ve & -ve poles), whereas unipolar leads record the electrical potential at a particular point by means of a single exploring electrode.

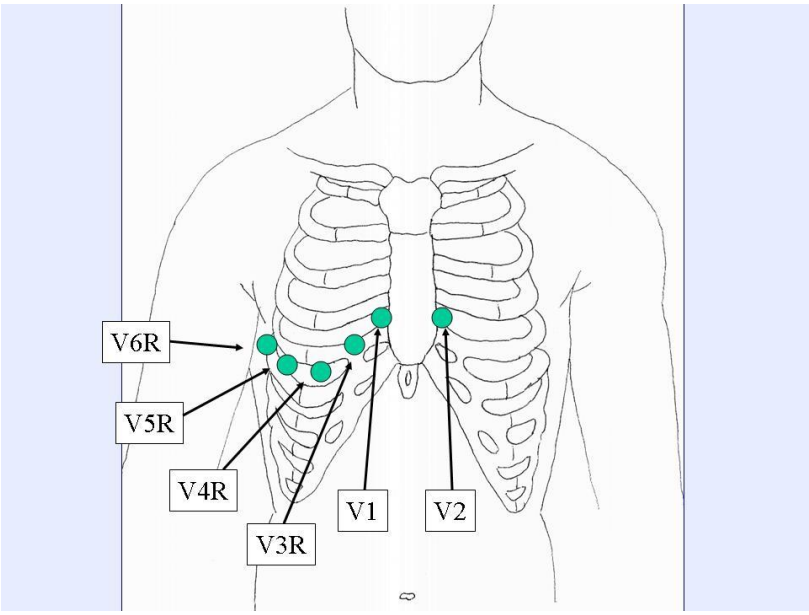

Fig. 1.3: All leads' position on the body

1.3.1 Einthoven Leads:

a. Lead 1: records potentials between the left and right arm,
b. Lead 2: between the right arm and left leg, and
c. Lead 3: those between the left arm and left leg.

1.3.2 Unipolar Limb Leads:

a. aVR: when the +ve terminal is on the right arm,
b. aVL: when the +ve terminal is on the left arm and
c. aVF: when the +ve terminal is on the left leg.

One lead connected to +ve terminal acts as the different electrode, while the other two limbs are connected to the –ve terminal serve as the indifferent electrode. Wilson leads (V1-V6) are unipolar chest leads positioned on the left side of the thorax in a nearly horizontal plane. The indifferent electrode is obtained by connecting the 3 standard limb leads. When used in combination with the unipolar limb leads in the frontal plane, they provide a three dimensional view of the integral vector.

1.3.3 Chest (precordial) Leads:

a. V1: 4th intercostal space, right sternal edge.
b. V2: 4th intercostal space, left sternal edge.
c. V3: between the 2nd and 4th electrodes.
d. V4: 5th intercostal space in the midclavicular line.
e. V5: on 5th rib, anterior axillary line.
f. V6: in the midzxillary line.

To make recordings with the chest leads, the three limb leads are connected to form an indifferent electrode with high resistances. The chest leads mainly detect potential vectors directed towards the back. These vectors are hardly detectable in the frontal plane. Since the mean QRS vector is usually directed downwards and towards the left back region, the QRS vectors recorded by leads V1-V3 are usually negative, while those detected by V5 and V6 are positive.

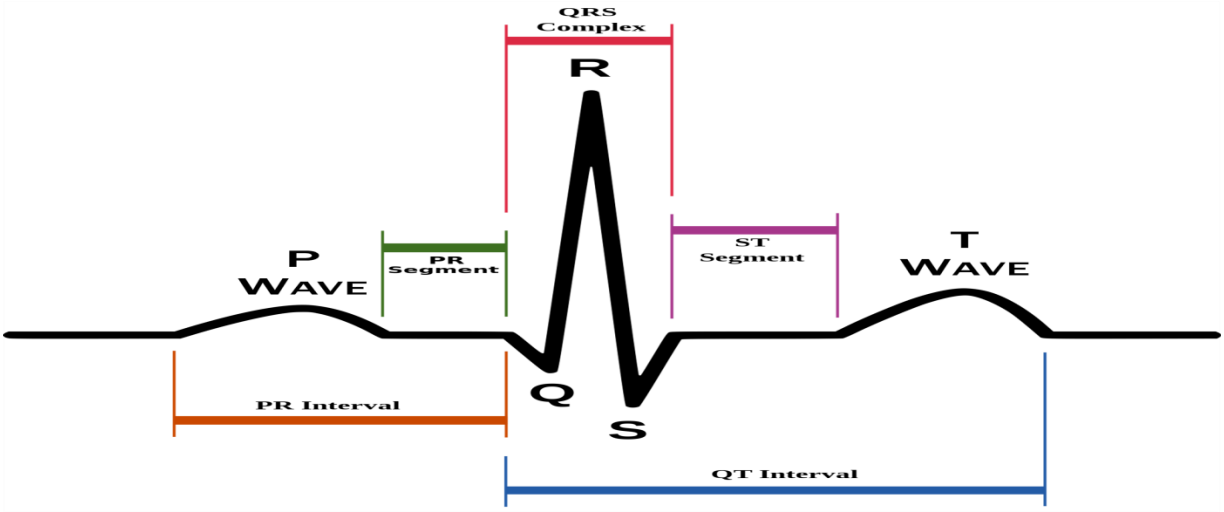

Fig. 1.4: Schematic representation of normal ECG waveform.

In leads V1 and V2, QRS = -ve because, the chest electrode in these leads is nearer to the base of the heart, which is the direction of electronegativity during most of the ventricular depolarization process. In leads V4, V5, V6, QRS = +ve because the chest electrode in these leads is nearer the heart apex, which is the direction of electro positivity during most of depolarization.

1.3.4 P Wave

P wave represents the sequential activation of the right and left atria, and it is common to see notched or biphasic P waves of right and left atrial activation. It is very difficult to analyze P waves with a high signal-to-noise ratio in ECG signal. A clear P wave before the QRS complex represents sinus rhythm. Absence of P waves may suggest atrial fibrillation, sinus node arrest, junctional rhythm or ventricular rhythm.

1.3.5 PE Segment:

The PR segment connects the P wave and the QRS complex. The impulse vector is from the AV node to the bundle of His to the bundle branches and then to the Purkinje Fibers. This electrical activity does not produce a contraction directly and is merely traveling down towards the ventricles and this shows up flat on the ECG. The PR interval is more clinically relevant than PR segment to any abnormality.

1.3.6 PR Interval:

The PR interval is measured from the beginning of the P wave to the first part of the QRS complex. It includes time for atrial depolarization, conduction through the AV node, and conduction through the His-Purkinje system. Disruption at any point can prolong the PR interval. The length of the PR interval changes with heart rate. Long PR interval suggests 1^{st} degree atrioventricular block whereas varying PR interval implies Mobitz type 1 atrioventricular block or multifocal atrial tachycardia.

1.3.7 QRS complex:

The QRS complex is the largest voltage deflection of approximately 10-20 mV but may vary in size depending on age, and gender. The QRS represents the simultaneous activation of the right and left ventricles, although most of the QRS waveform is derived from the larger left ventricular musculature. Duration of the QRS complex indicates the time for the ventricles to depolarize and may give information about conduction problems. Wide QRS complex indicates right or left bundle branch block, ventricular flutter or fibrillation etc.

1.3.8 T wave:

T wave represents ventricular repolarization. The interval from the beginning of the QRS complex to the apex of the T wave is referred to as the absolute refractory period. Tall T wave may suggest left bundle branch block, ventricular hyper trophy, hyperkalemia, stroke whereas small, flatted or inverted T wave implies myocardial ischemia, myocarditis, anxiety, certain drugs, right bundle branch block, hypokalemia etc.

1.3.9 ST Segment:

The ST segment occurs after ventricular depolarization has begun. The ST segment is usually isoelectric and has a slight upward concavity. It is always measured from the J point, where the QRS and ST segments meet. Elevation of ST segment implies myocardial ischemia, acute MI, LBBB, left ventricular hypertrophy, hyperkalemia, hypothermia whereas depression of ST segment may suggest, myocardial ischemia, acute posterior MI, LBBB, RBBB etc.

1.3.10 QT Interval:

The QT interval consists of the QRS complex along with the ST segment and T wave, which constitutes the majority of the duration. The QT interval is used primary as a measure of membrane repolarization. Since the QT interval varies with heart rate, the QT interval is corrected to make comparisons between ECG beats. A prolonged QT interval is a risk factor for ventricular tachyarrhythmia and sudden death. Short QT interval may suggest hyperkalemia, hypomagnesaemia, Graves' disease etc.

1.3.11 U Wave:

The U wave is hypothesized to be caused by the repolarization of the interventricular septum. Their amplitude is normally one-third of the following T wave and even more often completely absent. Ti may be seen following the T wave and can make interpretation of the QT interval especially difficult. This is associated with metabolic disturbances, typically hypokalemia, hypomagnesaemia and ischemia.

1.4 Arrhythmias in ECG Signal:

The normal rhythm of the heart where there is no disease or disorder in the morphology of ECG signal is called Normal sinus rhythm. The heart rate of NSR is generally characterized by 60 to 100 beats per minute. The regularity of the R-R interval varies slightly with the breathing cycle. The source of the rhythm is the Sino-atrial node, which is the normal pacemaker of the heart.

When the heart rate increases above 100 beats per minute, the rhythm is known as sinus tachycardia. This is not arrhythmia but a normal response of the heart which demand for higher blood circulation. If the heart rate is too slow then this is known as bradycardia and this can adversely affect vital organs. When the heart rate is too fast, the ventricles are not completely filled before contraction for which pumping efficiency drops, adversely affecting perfusion. Rhythms that deviate from NSR are called arrhythmias since they are abnormal and dysfunctional. Arrhythmias may be easily understood by categorizing them in the following manner:

a. Sinus Node Arrhythmias
b. Atrial Arrhythmias
c. Junctional Arrhythmias
d. Ventricular Arrhythmias
e. Atrioventricular Blocks
f. Bundle Branch and Fascicular Blocks

1.4.1 Sinus Node Arrhythmia:

This type of arrhythmia arises from the S-A node of heart. As the electrical impulse is generated from the normal pacemaker, the characteristic feature of these arrhythmias is that P-wave morphology of the ECG is normal. These arrhythmias are the following types: Sinus arrhythmia, Sinus bradycardia, and Sinus arrest etc.

Sinus Arrhythmia:

This is not a disorder or a true arrhythmia, but a normal, physiologic variation in the sinus rate with the phases of respiration. The slowest instantaneous heart beat may be less than 60 beats per minute, while the highest may exceed 100 beats per minute.

Sinus Bradycardia:

In sinus bradycardia, the rhythm originates from the S-A node but at a rate of less than 60 beats per minute. The ECG appears normal except for the slow heart rate. Mild sinus bradycardia is usually asymptomatic, while marked sinus bradycardia may lead to hypotension and result in insufficient perfusion of the brain and other vital organs. Treatment is indicated if the bradycardia is symptomatic.

Sinus Arrest:

In sinus arrest the S-A node intermittently fails to fire. There is no P-wave and therefore no accompanying QRS complex and no T-wave. Bradycardia may result if the occurrence of sinus arrests is frequent. Sinus arrest results from a marked depression of the automaticity of the S-A node. Since the automaticity of the S-A node is abnormal, the longest P-P interval will not be a multiple of the shortest P-P interval; unlike Sino-atrial exit block.

1.4.2 Atrial Arrhythmias:

Atrial arrhythmias originate outside the S-A node but within the atria in the form of electrical impulses. These arrhythmias types are given bellow

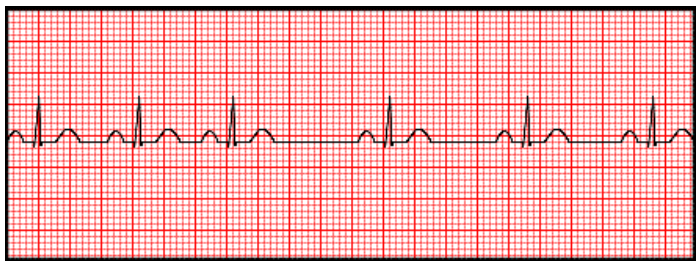

Fig. 1.5 Sinus arrhythmias

Premature Atrial Contractions (PAC):

This arrhythmia results an abnormal P-wave morphology followed by a normal QRS complex and a T-wave. This happens because of an ectopic pacemaker firing before the S-A node. PACs may occur as a couplet where two PACs are generated consecutively. When three or more consecutive PACs occur, the rhythm is considered to be atrial tachycardia.

Atrial Tachycardia:

The heart rate atrial tachycardia is fast and ranges from 160-240 beats per minute in atrial tachycardia. Frequently atrial tachycardia is accompanied by feelings of palpitations, nervousness, or anxiety.

Atrial Flutter:

In atrial flutter, the atrial rate is very fast, ranging from 240 to 360 per minute. The abnormal P-waves occur regularly and so quickly that they take morphology of saw-tooth waveform which is called flutter waves.

Atrial Fibrillation:

The atrial rate exceeds 350 beats per minute in this type of arrhythmias. This arrhythmia occurs because of uncoordinated activation and contraction of different parts of the atria. The higher atria rate and uncoordinated contraction leads to ineffective pumping of blood into the ventricles. Atrial fibrillation may be intermittent, occurring in paroxysms or chronic.

1.4.3 Junctional Arrhythmias:

Junctional arrhythmias are originated within the A-V junction in the form of the impulse comprising the A-V node and its Bundle. The abnormal in P wave morphology occurs because of these arrhythmias. The polarity of the abnormal P0wave would be opposite to that of the normal sinus P0wave since depolarization is propagated in the opposite direction from the A-V node towards the atria.

Premature Junctional Contractions (PJC):

It is a ventricular contraction initiated by an ectopic pacemaker in the atrioventricular (A-V) node. In premature junctional escape contraction, a normal looking QRS complex prematurely appears, but without a preceding P0wave, but the morphology of T-wave is normal.

Paroxysmal supraventricular Tachycardia:

In paroxysmal supraventricular tachycardia, the heart rate ranges from 160 to 240 beats per minute. PSVT may occur as a result of a reentry circuit in the A-V junction. It may also occur as a result of a re-entry circuit involving an accessory pathway between the atria and ventricles. The onset and termination of PSVT is abrupt, and may occur in repeated episodes that last for seconds, hours or days, In AVNRT, P-waves are usually buried in the QRS complex and hence not visible; whereas for AVRT the P waves may be visible.

1.4.4 Ventricular arrhythmias:

In this type of arrhythmia, the impulses originate from the ventricles and move outwards to the rest of the heart. In Ventricular arrhythmias, the QRS complex is wide and bizarre in shape.

Premature Ventricular Contractions (PVC):

In PVC the abnormality is originated from ventricles. PVCs usually do not depolarize the atria or the S-A node and hence the morphology of P waves maintains their underlying rhythm ad occurs at the expected time. PVCs may occur anywhere in the heart beat cycle. PVCs are described as isolated if they occur singly, and as couplets if two consecutive PVCs occur.

Ventricular Tachycardia (VT):

The heart rate of ventricular tachycardia is 110 to 250 beats per minute. In VT the QRS complex is abnormally wide, out of the ordinary in shape, and of a different direction from the normal QRS complex. VT is considered life threatening as the rapid rate may prevent effective ventricular filling and result in a drop in cardiac output.

Ventricular Fibrillation:

Ventricular fibrillation occurs when numerous ectopic pacemakers in the ventricles cause different parts of the myocardium to contract at different times in non- synchronized fashion. Ventricular flutter exhibits a very rapid ventricular rate with a saw tooth like ECG waveform.

1.4.5 Atrioventricular Blocks:

It is the normal propagation of the electrical impulse along the conduction pathways to the ventricles, but the block may delay or completely prevent propagation of the impulse to the rest of the conduction system.

First-Degree AV Block:

A first degree AV block is occurred when all the P0waves are conducted to the ventricles, but the PR-interval is prolonged.

Second-Degree AV Block:

Second degree AV blocks are occurred when some of the P-waves fail to conduct to the ventricles. Here, conduction intermittently fails to go down the bundle of His and bundle branches. There is usually a pre-existing complete block in one bundle branch, such that when an intermittent block occurs in the remaining bundle branch, conduction down to the ventricles fails. As a result of the pre-existing bundle branch conduction down to the ventricles fails. As a result of the pre-existing bundle branch block, the QRS complex is commonly wide. Commonly the conduction ratio is 4:3 or 3:2 P-waves to QRS complexes. This type of AV block is dangerous as complete heart block can occur unpredictably, and is an indication for implantation of an artificial pacemaker. Third-Degree AV block, the rhythm of the P0waves is completely dissociated from the rhythm of the QRS complexes. The atria and the ventricles beat independently, each beat at their own rate.

Paced Beat (Implant):

Hearts implanted with an artificial pacemaker will generally beat around 50 to 70 beats per minute depending on the setting of the artificial pacemaker. The pacing electrode is commonly attached to the apex of the right ventricular cavity or the right atrium, or both. The artificial pacemaker produces a narrow, often biphasic spike. A pacemaker lead positioned in an atrium produces a pacemaker spike followed by a P-wave. A pacemaker lead positioned in a ventricle produces a pacemaker spike followed by a wide, bizarre QRS complex. Sometimes it is also called pacemaker rhythm.

1.4.6 Bundle Branch Blocks:

Bundle branch block, cease in the conduction of the impulse from the AV node to the whole conduction system. Due to this block there may occur myocardial infarction or cardiac surgery.

Right Bundle Branch Block (RBBB):

When the right bundle branch is blocked, the electrical impulse from the AV node is not able propagate to the Purkinje network to depolarize the right ventricular myocardium. Instead, the impulse propagates in a convoluted manner through the left ventricular myocardium to reach the right ventricular myocardium. Since propagation through the myocardium is much slower than through the specialized conduction tissue, the QRS complex becomes widened. The morphology of the QRS complex also changes and takes on a bizarre appearance because of the different direction of depolarization.

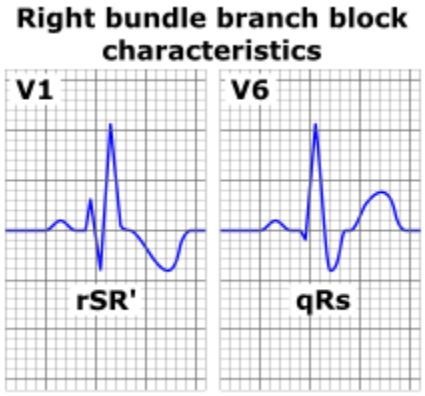

Fig. 1.6: Right Bundle Block (RBBB)

Left Bundle Branch Block (LBBB):

Similar to right bundle branch block, a block in the left bundle branch will prevent the electrical impulses from the A-V node from depolarizing the left ventricular myocardium in the normal way. For LBBB, the right ventricle is depolarized primarily, generating an electrical wave front that eventually spreads to the left ventricular myocardium causing the myocardium to depolarize.

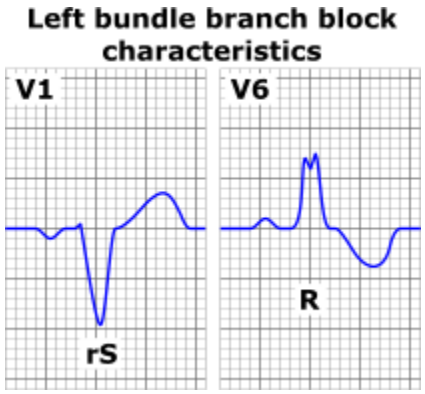

Fig. 1.7: Left Bundle Branch Blocks

Chapter 2

Wavelet based packet ECG arrhythmia classification

2.1 Introduction:

In the EMD based LFD estimation method, for better approximation of LFD at each sample point, at least three IMFs are to be determined, which is dependent on the length of the segment of an ECG signal used for EMD operation[1]. For some types of ECG arrhythmia, it is difficult to optimize the segment length in order to ensure the extraction of at least three IMFs. This problem can be overcome by analyzing the ECG signal in time-frequency domain via Wavelet Packet Decomposition (WPD). Here time frequency analysis of ECG signal in WPD domain is performed for calculating the HE as well as deriving a set of more effective LFD features. Features extracted from these WPD coefficients can efficiently represent the wavelet space is also decomposed along with the scale space decomposition thus making the higher frequency band decomposition possible [2]. We have used state of art methods PSDFE and VFE for comparison of ECG arrhythmia classification performance with the proposed method based on LFD estimation in WPD domain in terms of different performance evaluation criteria.

2.2 Wavelet Transform:

Discrete wavelet Transform (DWT) is multi-resolution transform with a very fast implementation. DWT is a lost less linear transformation of a signal or data into coefficients on a basis of mother wavelet functions. A family of mother wavelet is available having the energy spectrum concentrated around the low frequencies like the ECG signal as well as better resembling the QRS complex of the ECG signal. Therefore, for the analysis of an ECG signal at different scales [3], wavelet transform (DWT) is used in practice. In discrete wavelet transform (DWT), for analyzing both the low and high frequency components, it is passed through a series of low pass and high pass filters with different cut off frequencies. This process results in a set of approximate and detail DWT coefficients, respectively. The filtering operations in DWT result in a change in the signal resolution. Thus, DWT decomposes the signal into approximate and detail

information thereby helping in analyzing it at different frequency bands with different resolutions. In wavelet analysis, only scale space is decomposed, but wavelet space is not decomposed. This results in a logarithmic frequency resolution [4], which does not work well for all the signals. By the restriction of Heisenberg's uncertainty principle, the spatial resolution and spectral resolution of high frequency band become poor thus limiting the application of wavelet transform.

2.2.1 Wavelet Packet Transform selection:

ECG signal sequence of cardiac cycles or 'beats' is one point for wavelet packet transform. ECG is not strictly a periodic signal differences in period and amplitude level of beats. Each region has different frequency components-QRS has high frequency oscillations region has lower frequencies and U regions have very low frequencies. Signal contains noise components due to various sources that are suppressed during processing of ECG signal. Fourier Transform - provides only frequency information, time information is lost. Short Term Fourier Transform (STFT) - provides both time and frequency information, but resolves all frequencies equally. Wavelet transform - provides good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. Useful approach when signal at hand has high frequency components for short duration and low frequency components for long duration as in ECG.

2.2.2 Discrete Wavelet Transform:

Time-scale representation of signal obtained using digital filtering techniques. Resolution of the signal is changed by filtering operations. Scale is changed by up sampling and down sampling (subsampling) operations [5]. Subsampling is reducing sampling rate or removing some of the samples of the signal. Up sampling will be increasing sampling rate by adding new samples to the signal. The equation used in discrete wavelet transform is stated below

$$w_{m,n} = \langle x(t), \psi_{m,n} \rangle = a_0^{m/2} \int f(t) \psi(a_0^m(t) - nb_0) dt$$

2.2.3 Discrete Wavelet Transform Analysis:

In many practical applications and especially in the application described in this report the signal of interest is sampled. In order to use the results we have achieved so far with a discrete signal we have to make our wavelet transform discrete too. Remember that our discrete wavelets are not time-discrete, only the translation- and the scale step are discrete. Simply implementing the wavelet filter bank as a digital filter bank intuitively seems to do the job. But intuitively is not good enough, we have to be sure DWT of original signal is obtained by concatenating all coefficients starting from the last level of decomposition. DWT will have same number of coefficients as original signal. Frequencies most prominent (appear as high amplitudes) are retained and others are discarded without loss of information.

$$\varphi(2^{j}t) = \sum_{k} h_{j+1}(k)\varphi(2^{j+1}t - k)$$

The two-scale relation states that the scaling function at a certain scale can be expressed in terms of translated scaling functions at the next smaller scale [5]. Do not get confused here: smaller scale means more detail. The first scaling function replaced a set of wavelets and therefore we can also express the wavelets in this set in terms of translated scaling functions at the next scale. More specifically we can write for the wavelet at level j:

$$\Psi(2^{j}t) = \sum_{k} g_{j+1}(k)\varphi(2^{j+1}t - k)$$

This is the two-scale relation between the scaling function. Since our signal f (t) could be expressed in terms of dilated and translated wavelets up to a scale j-1, this leads to the result that f(t) can also be expressed in terms of dilated and translated scaling functions at a scale j.

2.3 The Fast Wavelet Transform:

Start from the definition, if the form of scaling and wavelet function is known, its coefficients is defined. We can find another way to find the coefficients without knowing the scaling and

dilation version of scaling and wavelet function. The computation time can be reduced. The final equation will be

$$W_{\Psi}[j,k] = h_{\Psi}[-n] * W_{\Psi}[j+1,n]_{\,n=2k,k\geq 0}$$

For the commonly used discrete signal, say, a digital image, the original data can be viewed as approximation coefficients with order J. That is, f[n] = WΨ [J; n] next level of approximation and detail can be obtained. This algorithm is "fast" because one can find the coefficients level by level rather than directly using above equation to find the coefficients. This algorithm was proposed in that way.

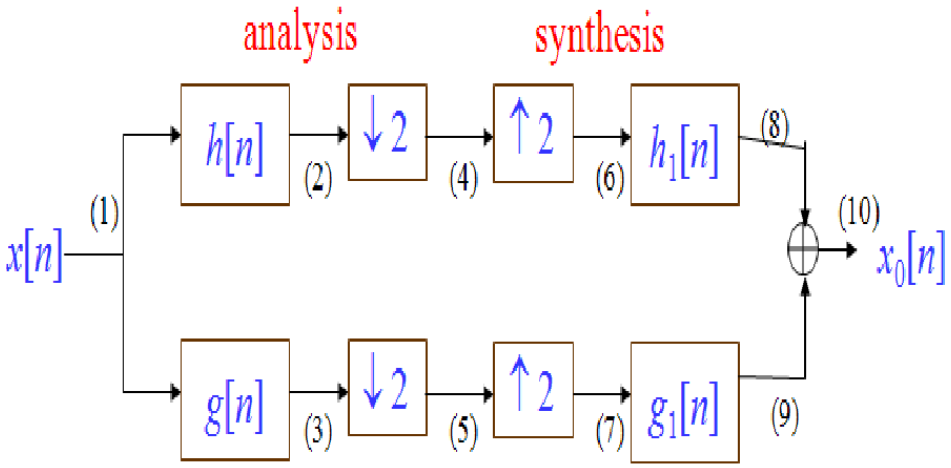

Fig. 2.1: The schematic diagram to realize discrete wavelet transforms. Here the filter names are changed

2.3.1 Sub band Coding:

The satisfying result in the above equation is much simpler to implement to

Convolve WΨ[j; n] with different filters, h Ψ [-n] and h [-n]. Down sample by the factor, we will find the next-level coefficients. Multiresolution analysis can be achieved by cascading the structure above. A schematic diagram of this algorithm is shown in Fig. 4.3. The synthesis process does the inverse operation of the analysis process. First up sample by a factor of 2 and

then convolve the signal with a known inverse filter. We expect the reconstructed signal, x0[n], to be exactly the same as the original signal, x[n].

2.3.2 Mother wavelets:

Mother wavelets or wavelet functions are used for characterization of different wavelet shapes. They cover entire domain of interests. The selection of Mother Wavelet is very much important in the WPD procedure. There are several wavelet families like – Harr, Daubechies, Biorthogonal, Coiflects, Symlets, Morlet, Mexican Hat, Meyer etc. Here Mexican Hat, Morlet and DB6 has been shown below

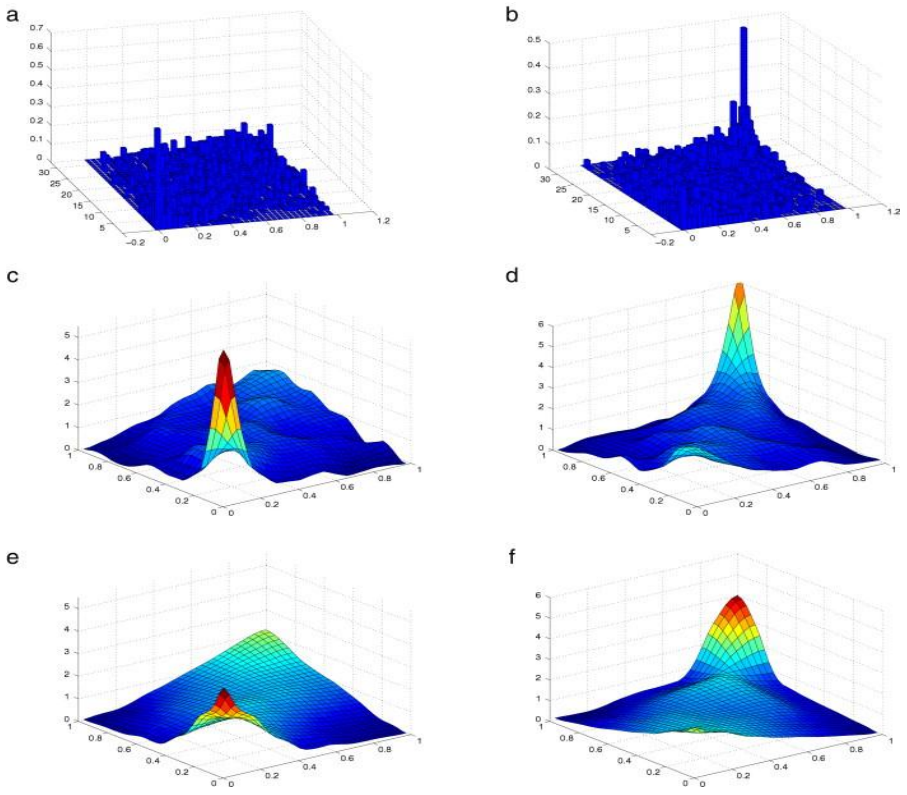

Fig. 2.2: different type of wavelets

2.4 Continuous Wavelet Transform:

In this chapter, we talk about continuous wavelet transform. This transform works when we use a continuous wavelet function to find the detailed coefficients of a continuous signal. Like the concept in 4.3, we have to establish a basis to do such analysis. First, we give the definition of continuous wavelet transform and do some comparison between that and the Fourier transform.

2.4.1 Definition:

We define a mother wavelet function Ψ (t) is a subset of $L^2(R)$, which is limited in time domain. That is, Ψ(t) has values in a certain range and zeros elsewhere. Another property of mother wavelet is zero-mean. The other property is that the mother wavelet is normalized. Mathematically, they are

$$\Psi_{s,u}(t) = \frac{1}{\sqrt{s}}\Psi\frac{t-u}{s}$$

u is the translating parameter, indicating which region we concern. s is the scaling parameter greater than zero because negative scaling is undefined. The multi resolution property ensures the obtained set of $\Psi_{u,s}(t)$.g is orthonormal. Conceptually, the continuous wavelet transform is the coincident of the basis $\Psi_{u,s}(t)$.

2.5: Different result to simplify WPT method:

- To increase Specificity, Sensitivity and Accuracy.
- KLT (Karhunen-Loke transform) method, several other compression methods with WPT technique.
- In this figure below, we simulated a signal from MIT-BIH database which was only done by using WPT and the resultant classification was Ventricular Arrhythmia
-

2.5.1 KLT and Wavelet packets:

The growing amount of data coming from digital recording of ECG signals needs data compression techniques [6]. Data compression can be understood as the process of detecting and removing redundancies in a signal**.** Three different redundancy sources can be distinguished in

the ECG signal: time correlation between the samples of a single beat (intra-beat correlation), beat- to-beat quasi-periodic behavior (inter-beat correlation) and the inter-lead correlation.
It would be very useful to compare the rate distortion performance of the best basis obtained by the wavelet packets methodology to the optimum KLT. We selected 10 minutes of record 100 from MIT Arrhythmia database. Moreover, the best basis chosen from the Wavelet packets obtain a very close performance to KLT. However, any data compression system must be evaluated in a rate-distortion sense, i.e., taking into account also the number of bits used to code the coefficients and the overhead information. The overhead information for the KLT is very large (mainly composed of the basic functions). In contrast, the overhead information related the best basis is very small due to the binary-tree structure of the wavelet packets expansions. We show that RD curves for KLT and WP, where it is shown the clear higher performance of the wavelet packets approach. In next sections we will only consider the wavelet packets based algorithm.

2.5.2 Pure wavelet packet transforms techniques:

ECG Analysis has enormous benefits on diagnosis the abnormality and the pathological conditions on those patients with cardiac diseases [7]. The ECG for diagnosis consists of P wave, QRS complex and T wave, which come out according to directions of cardiac rhythms to each parts of the heart. There are several obstacles to analyses ECG waves and detect parameters within each ECG wave since ECG has a non-stationary characteristic and there are several kinds of noises, which complicate the analyses. There are several ways to perform ECG analyses such as the calculation for the first derivative Adaptive threshold algorithm Hidden Markov modeling [8], neural network and so on. Each method of analyses has its own shortcomings such as the complexities, time for calculation and other shortcomings. One of interesting methods is a wavelet packet, which has several frequency bands, suitable for the signals with changing frequencies according to the change of the time along with the property from eliminating noises at different frequencies.

The procedure for detecting and analyzing T wave and P wave is similar to the QRS complex detection, with the only different is the application of wavelet packet transform at the 4th scales which more suitable to the low frequency signals, thus enhancing T wave and P wave as shown

in the fig 2.3(a) and fig 2.3(b). After eliminating QRS complex, the system would use the onset and offset of QRS complex to mark the onset of T wave and the offset of P wave before making further calculations.

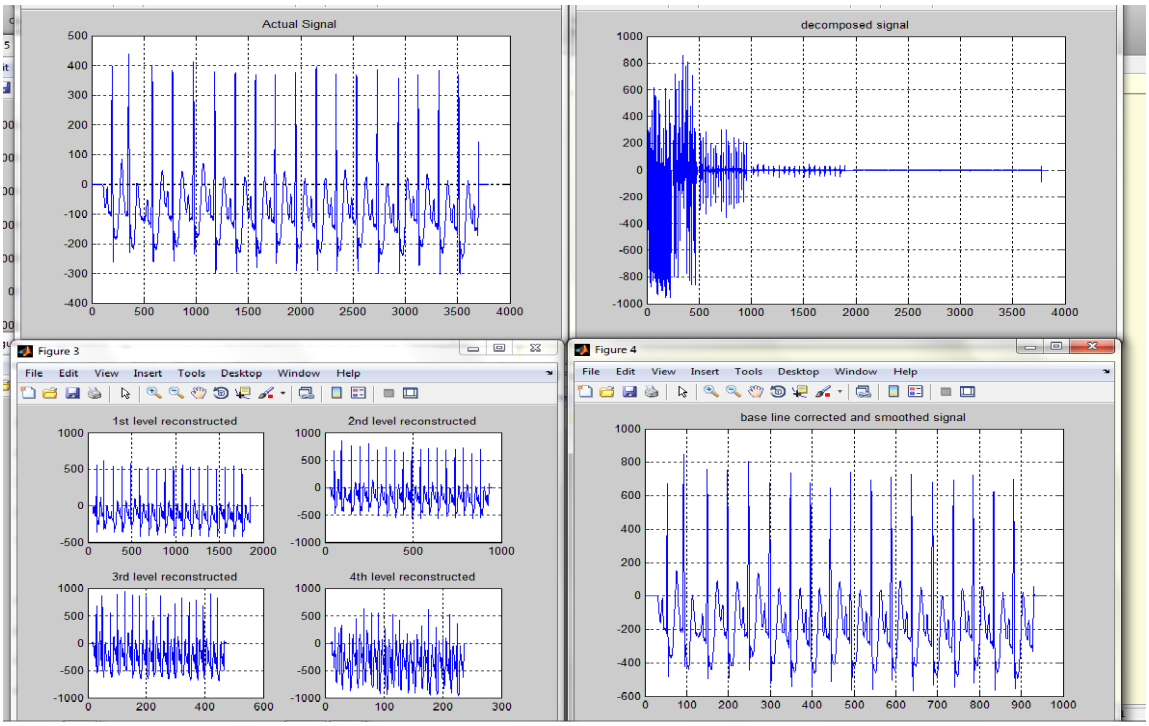

Fig. 2.3(a): Different Figures of the Simulation results of MIT-BIH Ventricular arrhythmia Data (203.m). Actual signal (fig. 1), Decomposed signal (fig.2), reconstructed signal (fig.3)

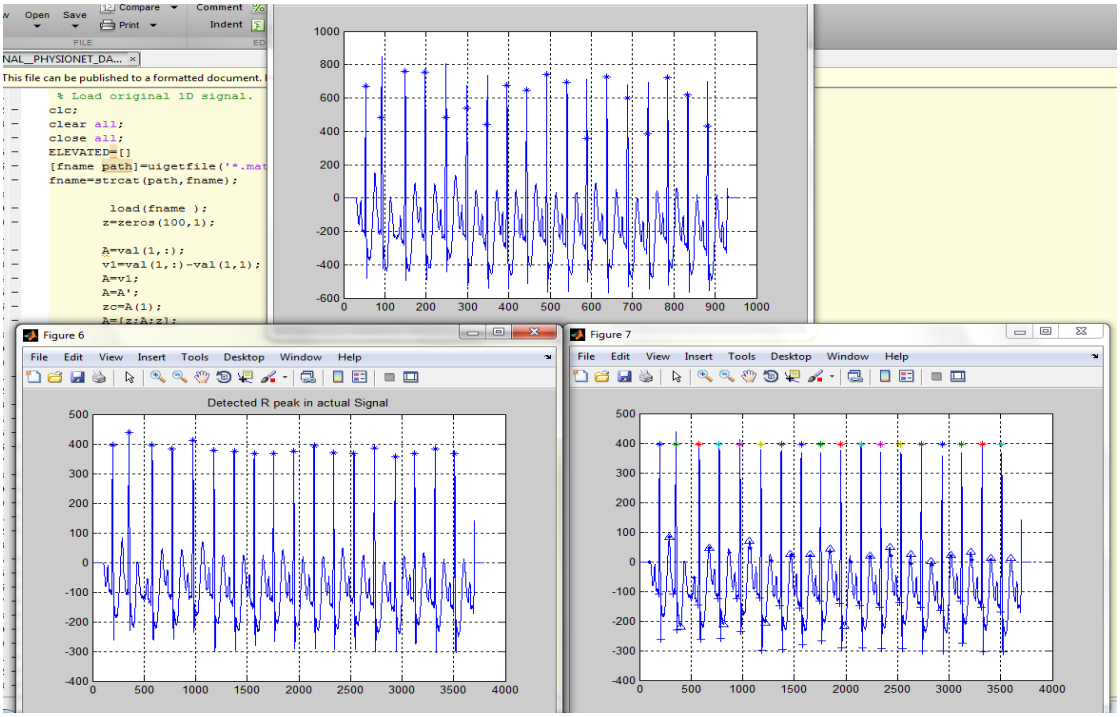

Fig. 2.3(b): Base line smoothed signal (fig.4), Detection of R peak in signal (fig.6)

2.6 Conclusion:

In the present work, the possibilities of the wavelet transform have been shown as an alternative atrial extraction method. The electrocardiography registers have an implicit atrial that it is underhanded with the synthesized one. This could justify the obtained cross-correlation values. Methodology 2 extracts a smoother signal with higher cross correlation coefficients, but it needs two consecutive transformations and it couldn't be convenient due to the increase of the computational load. The presented results show a good similarity between the expected AA and the obtained signal. The principal peak in the 5-8 Hz range and the energy concentration around the band peak are characteristic of this type of arrhythmias.

An extension of data compression methods based on transform coding to multi lead ECG recordings was shown. Two orthogonal transforms were analyzed: the Karhunen-Loeve transform and a fast approximation using wavelet packets. In spite of the optimality of the Karhunen- Loeve transform, it obtained a lower performance (in a rate-distortion sense) than wavelet packets due to the overhead information needed to code its basis functions

Chapter 3

Non-linear analysis of ECG arrhythmia detection and classification

ECG signals actually show non-linear characteristics. So, to do so a non-linear algorithm should be used to detect the problem with the ECG signal. There are many filtering and classifiers used in detecting the arrhythmias. Like KNN- classifier, Bayes classifier etc. There are some algorithms that are used to detect the ECG correlation dimension, largest lyapunov exponent and so on. We also work on like HRV (heart rate variability) analysis, RR interval analysis and also on QRS complex detection analysis. So, by dint of these processes and algorithms we came to know how AF (atrial fibrillation), VA (Ventricular arrhythmia) these fetal arrhythmias are detected and their optimization are confirmed. So, non-linear analysis to detect the ECG arrhythmias is a huge progress towards any heart diseases and failures.

3.1 Correlation dimension and Lyapunov exponent analysis:

We present a study of the nonlinear dynamics of electrocardiogram (ECG) signals for arrhythmia characterization [9]. The correlation dimension and largest Lyapunov exponent are used to model the chaotic nature of five different classes of ECG signals. The model parameters are evaluated for a large number of real ECG signals within each class and the results are reported. The presented algorithms allow automatic calculation of the features. The statistical analysis of the calculated features indicates that they differ significantly between normal heart rhythm and the different arrhythmia types and, hence, can be rather useful in ECG arrhythmia detection. On the other hand, the results indicate that the discrimination between different arrhythmia types is difficult using such features**.**

3.1.1 Correlation dimension estimation:

The mathematical description of a dynamical system consists of two parts: the state which is a snapshot of the process at a given instant in time and the dynamics which is the set of rules by which the states evolve over time [10]. variables is not known. The time window length (W) is defined by the time spanned by each embedding vector,

$W = (m - I) L$

After determining m using FFN, we need to select the optimal time window length (W). The selection procedure will become apparent later in this section. To study the dynamics of such system, we first need to reconstruct the state space trajectory [11]. The most common method to do this is using delay time embedding theorem to create a larger dimensional geometric object by embedding into a larger m-dimensional embedding space. The embedding dimension m [12] [13] must be large enough for delay time embedding to work. The correlation dimension D2 is defined as the slope of the linear region of the plot of log(C(r)) versus log(r) for small values of r. In practice, the determination of the linear scaling region is not an easy task because of the presence of noise, which makes it not practical to compute the slope for very small values of r. Moreover, this determination was found to be not repeatable using manual selection. C(r) is the average number of neighbors each point has within a given distance r, given as,

$$c(r) = \frac{1}{Np}\sum_{i,j} \theta[r - (\{(i) - x(j)\}]$$

In our implementation, we tried this approach combined with computerized regression and the results were not satisfactory. Then, we improved our implementation using a second-order regression for the whole curve. The linear regression was then obtained for the part of this curve that appeared linear by vision. More consistent values for
D2 were obtained.

$$D_2 = \lim_{r \to 0} \frac{\log(C(r))}{\log(r)}$$

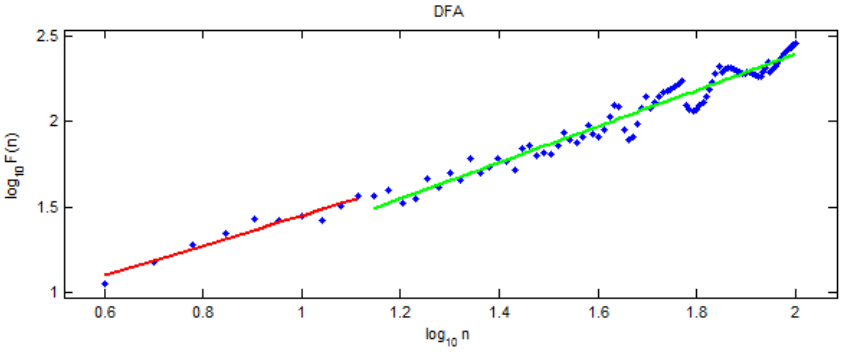

Fig. 3.1: log(C(r)) versus log(r) curve

3.1.2 Calculation of Lyapunov Exponents by WOLF'S algorithm:

Lyapunov exponents are defined as the long time average exponential rates of divergence of nearby states. If a system has at least one positive Lyapunov exponent, then the system is chaotic. The larger the positive exponent, the more chaotic the system becomes (i.e., the shorter the time scale of system predictability) [14]. Lyapunov exponents will be arranged such that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \cdots \ldots \geq \lambda_n$, where λ_1 and λ_n correspond to the most rapidly expanding and contracting principal axes, respectively. Therefore, λ_1 may be regarded as an estimator of the dominant chaotic behavior of a system. The equation is

$$\lambda_1 = \frac{1}{t_s - t_0} \sum_{k=1}^{s} \log_2 \left(\frac{d_1(t_k)}{d_{0(t_{k-1})}} \right)$$

We used a software implementation of Wolf's algorithm. This software is divided into two programs: BASGEN and FET. BASGEN stands for "dataBASe GENerator". It is considered as a preprocessing step for the main program FET. There are a lot of parameters that need to be defined for the two programs. We selected those parameters for BASGEN as follows: embedding dimension m=4, time delay L=60, and grid resolution ires= 20. It should be noted that for Lyapunov exponent calculations, the embedding dimension m was chosen as D2 rounded to the next highest integer. Also, the grid resolution refers to the fact that BASGEN places the reconstructed data into a grid of dimension m, with a resolution of iris cells per side. This grid will be used later by FET to efficiently find nearest neighbors to any point.

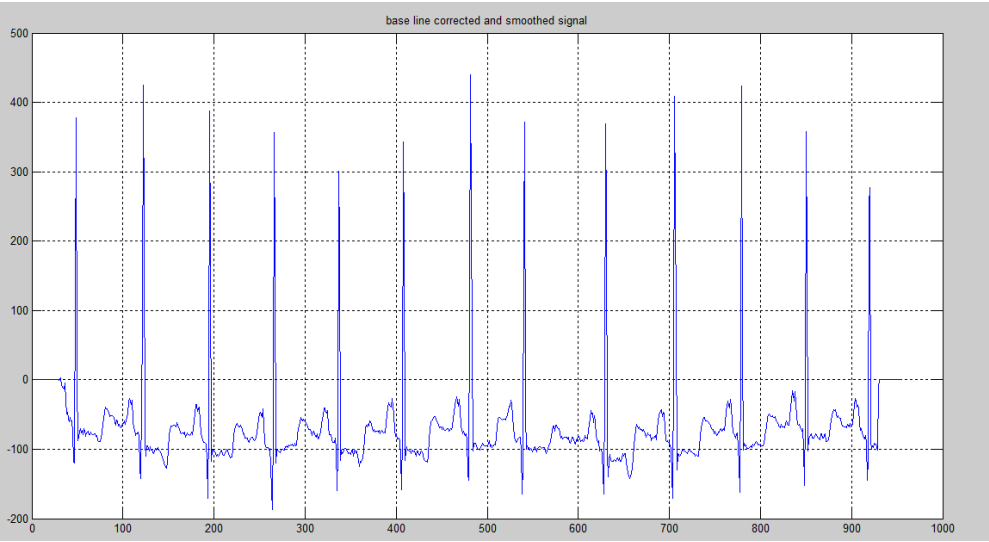

Fig. 3.2: base line corrected and smoothed signal

3.2 AF (Atrial fibrillation) detection and approximation:

Atrial Fibrillation (AF) is a supraventricular arrhythmia characterized by uncoordinated atrial activation [15]. AF occurs when the electrical impulses in the atria degenerate into a rapid chaotic pattern. On the ECG, AF is described by the replacement of P waves by fibrillatory waves (f waves) that vary in size, shape, and timing, associated with an irregular ventricular response. Consequently, when AF occurs, a notably disorganized atrial activity (AA) can be observed.

3.2.1 Sample entropy:

Sample Entropy (SampEn) examines time series for similar epochs and assigns a non-negative number to the sequence [16], with larger values corresponding to more complexity or irregularity in the data. SampEn(m, r) is the negative logarithm of the conditional probability that two sequences similar during m points remain similar at the next point, where self-matches are not included in calculating the probability. Although m and r are critical in determining the outcome of SampEn, no guidelines exist for optimizing their values. Nevertheless, them and r values suggested by Pincus are m = 1 or m = 2 and r between 0.1 and 0.25 times the standard deviation of the original time series.

3.3 Chaotic HRV analysis by non-linear means:

By computing Heart Rate Variability (HRV) signal form electrocardiogram (ECG) signals, 14 carefully selected time domain, frequency domain, nonlinear and chaotic features are extracted and used to train MLP neural networks [18]. HRV signal is known to be less sensitive to noise as compared to ECG and has higher chaotic and nonlinear characteristics. 7 arrhythmias (i.e. PVC, AF, CHB, LBBB, NSR, VF, and VT) have been classified with accuracies ranged from 95% to 100% on MIT-BIH dataset.

3.3.1 Methods of representing HRV:

x(t) is a direct measure of heart rate variability (HRV). This is transformed to a continuous HRV-signal y(t) **as** shown in Figure IC. The transformation of x(t) to y(t) is described by the transfer

function h(t) or its Fourier transform H(f) respectively. It can be shown [7] that the spectrum of the point process Sx(f) is related to the spectrum of the step-like interpolation algorithm Sy(f) by

$$S_y(f) = |H(f)|^2 S_x(f)$$

The transfer function was derived in**.** Denoting the square of the normalized standard deviation , the transfer function is given by

$$|H(f)|^2 = <I>^4 (1 + 4\gamma + 3\gamma^2)[\sin(\pi f <I>/\pi f <I>)]^{n(\gamma)+2}$$

Where, by the exponent n(γ) decreases from 4.2 to 2 when y is decreased from 0.1 to 0.01. In the present investigation, the normalized standard deviation was found in the range0.023 <σ/I< 0.129. This demonstrates that the step-like interpolation algorithm [19] introduces a low pass filter characteristic which is dependent on the forth power of <I> and on the quotient of breathing Sequence to mean heart rate. This induces a systematical error. Applying (l), the spectrum of the step-like Interpolation algorithm can be corrected for the transfer Function giving rise to

$$S_x(f) = S_y(f)/<I>^4 [\sin(\pi f <I>/\pi f <I>)]^4$$

The transfer function after linear interpolation must be calculated by computer simulation.

The transfer function contains respiratory rate (f) and mean heart rate (I). If both parameters were changed the cut off frequency of the low pass filter depends on the product PI of respiratory rate and mean heart rate. To assess the dependence on respiratory rate mean heart rate must be nearly constant.

3.3.2 Methods for assessing RSA:

We recorded one ECG lead of twelve healthy volunteers during metronome controlled respiration (8**,** 12 and 18 cycle shin = 0.13, 0.2 and 0.3 Hz) over a time period of 6 minutes during n o d sinus rhythm. R-R intervals were derived from the ECG in using an algorithm which automatically senses R-waves. Mean and standard deviation of R-R intervals were calculated far each record. Applying a step-like interpolation algorithm, the sequence of R-R intervals x(t) is transformed to a continuous process y(t).

3.3.3 Feature Extraction:

HRV signal demonstrates both linear .and nonlinear (specifically, chaotic) behaviors. In this paper, we use a combination of seven linear time domain features and two frequency domain features together with five chaotic features of HRV signal [20]. We choose a signal length of N=32 for our analysis. Seven linear time domain features used are as follows:

Mean: the average of RR intervals of HRV signal.

SD: standard deviation of RR interval of HRV signal.

Mean HR: mean of heart rates.

SD HR: standard deviation of heart rates.

RMS: root mean square of RR intervals.

PNN50: ratio of the Number of successive difference of intervals which differ by more than 50ms to the total number of all RR intervals.

$$\mathrm{PNN50\%} = \frac{\mathrm{NN50}}{\mathrm{N}-1} * 100$$

M_x: M_x is defined as the roots of ratio of first differential of time series variance to its own variance:

$$\mathrm{M_x} = \sqrt{\frac{\delta^2 \acute{\mathrm{x}}}{\delta^2 \mathrm{x}}} = \frac{\delta \acute{\mathrm{x}}}{\delta \mathrm{x}}$$

Although, time domain features are very informative, they are not able to distinguish between sympathetic and parasympathetic contents of HRV signal. Here we use LF (Low frequency power spectral density) and HF (high frequency power spectral density) as two important frequency domain features.

3.3.4 Feature Dimension reduction:

Generalized discriminate analysis (GDA) is a tool to decrease the distance of patterns belong to the same class and simultaneously increase the distance of patterns of different categories; thereby, GDA increase the classification performance. In addition, it reduces the number of feature of an N-class problem to N-1. The fact demonstrates the effect of application of GDA to features of HRV signal. In the left figure, patterns of different classes are overlapped, but after

application of GDA, patterns are mostly linear separable even in this two dimension feature space. Note that, as we want to classify seven arrhythmias, GDA will reduce the previously specified 14-dimensional feature space to a 6-dimension feature space.

3.3.5 Ignoring non-informative samples:

As mentioned before, to improve the learning and generalization capability of our neural network classifier, it is important to prevent misleading patterns to be fed to NN during its learning phase. To determine which samples may mislead the net, or equally, to determine which samples are more representative of their belonging categories, we profiteered the learning patterns. To do so, we learn 350 HRV samples to a self-Organizing Map (SOM) neural network. SOM is a self-organizing classifier, in the manner that it classifies similar input patterns to nearby output neurons without receiving the appropriate output for input patterns. After a SOM is learned, each region of output neurons represents a category of the learning patterns. By, reapplying the patterns belong to a specified class to learned SOM. LLE verses CD before application of GDA (a) and after GDA (b) may determine the winning neurons as representative of those categories. After that, neurons which responded to more than one category are representative of overlapping (i.e. misleading) patterns, while neurons responded to only one category are representative of that category. Applying this procedure to our 350 data set, we choose 20 most representative patterns for each arrhythmia classes. That is, our learning data set consist of 140 patterns.

3.4 Non-Linear Features of the RR Interval Signal:

We explore the RR interval signal to detect arrhythmic segments in electrocardiograms (ECG) using non-linear analysis. Initially, the RR interval signal is extracted and it is segmented into small segments Time-frequency analysis is used for the calculation of the total energy. These characteristics are fed into a neural network to classify each segment as normal or arrhythmic.

The proposed approach is validated using the MITBIH database for various segment sizes (32, 64, 128, 256 and 512 RR intervals). The method results in high sensitivity and specificity (85% sensitivity and 92% specificity) for arrhythmic segment detection.

3.4.1 Method:

In the first stage, QRS detection is performed on the ECG recordings and the RR interval signal is constructed. The signal is segmented into small segments. In the second stage 4 features are extracted from each segment: standard deviation, total energy, ApEn and NormEn. The calculation of ApEn is based on a method proposed in parameters m and r needed for the calculation of ApEn are set to m = 2 and r = 20% of the standard deviation of the segment, as proposed in. NormEn was calculated as:

$$\text{NormEn}=\frac{\text{Entopy}}{\text{power}}$$

Where entropy is the Shanon entropy defined as

$$\text{Entropy}= -\sum_{i=1}^{M}\frac{N_i}{N}\log_2(Pi)$$

Where N is the segment length (i.e. number of RR intervals in the segment), M the total possible values of an RR interval, Ni the number of appearances of value I (i.e. length of an RR interval) in the segment and Pi the probability of appearance of value i in all segments. Pi is calculated as:

$$P_i = \frac{\#\text{of appearances of value i in all segments}}{\text{length of all segments}}$$

The energy of the segment is calculated using Smoothed Pseudo Wigner-Ville distribution (SPWV),

$$\text{SPWV}_x = \int_{-\infty}^{+\infty} h(\tau) \int_{-\infty}^{+\infty} g(s-t)x(s+\tau/2)x*(s-\tau/2)ds e^{-i2\pi\omega\tau}d\tau$$

Where x is the signal, x* is the complex conjugate of the signal, g is a time smoothing window and h is a frequency smoothing window in the time domain. To avoid marginal problems in the segments, the calculation of the time-frequency plane was made before the segmentation of the RR interval signal. The energy of the segment is defined as the integration of SPWVx over the time frequency plane: The power of the segment, used for the calculation of the NormEn, is defined as the ratio of the energy and the segment duration (i.e. the sum of the duration of all RR intervals in the segment).

The features extracted from stage b are used to train a back propagation neural network (stage c). The chosen architecture of the neural network contains: 4 inputs, one hidden layer with 20 neurons and one output, being a real number in the interval [0, 1]. Different RR interval segment sizes are used, with 32, 64, 128, 256 and 512 RR intervals. 20% of the RR interval segments, for each segment size, are used for the training of the neural network and the rest 80% for testing. The training of the neural network ends if the sum of the square errors for all segments is less than 0.001 or the maximum number of training epochs is reached (2.000 epochs).

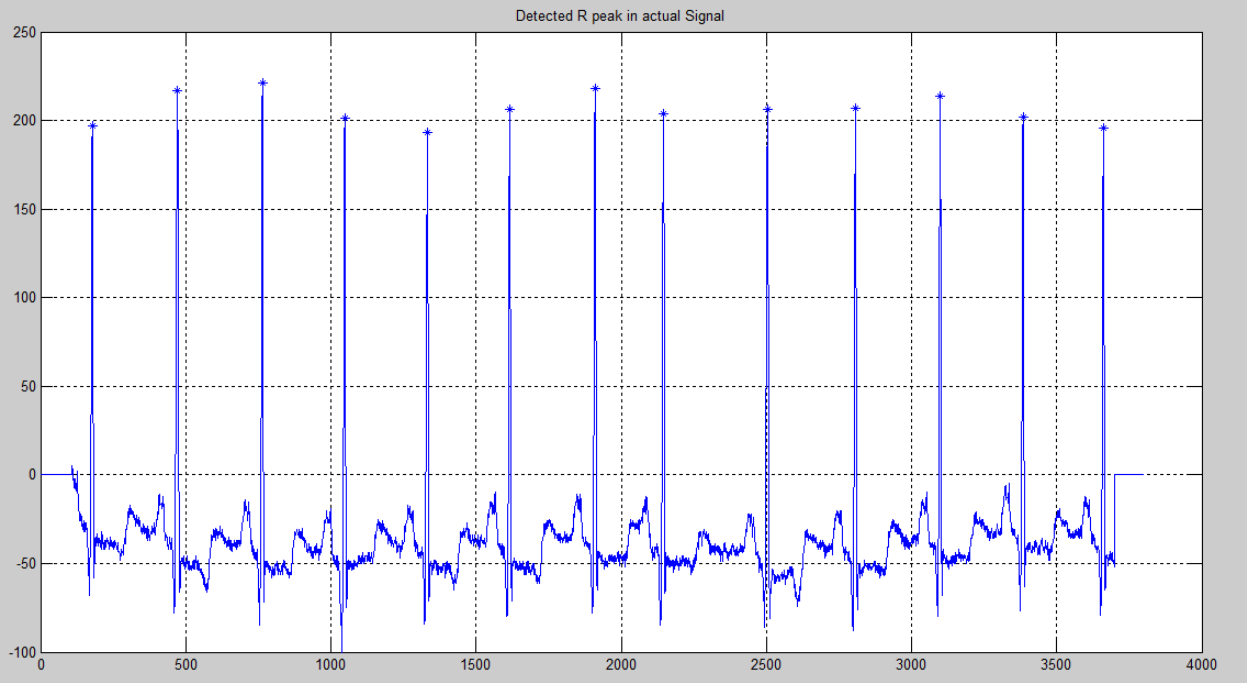

Fig. 3.3: R-R interval (peak to peak)

3.4.2 The dataset:

The MIT-BIH arrhythmia database is used for the training and testing of the proposed method. The database beat and rhythm annotation was used to determine if an RR interval segment is arrhythmic or normal. The annotation of the second beat of the RR interval (i.e. the beat containing the second R wave) was used to annotate the RR interval. An RR interval segment is annotated arrhythmic or normal according to the following rules: (i) if an RR interval is annotated as N, L, R, P, f, p or Q and the rhythm is annotated as (N, (IVR, (B, (T, (P or (PREX then it is considered normal, otherwise it is considered arrhythmic, and (ii) if there are less or equal than n arrhythmic RR intervals in an RR interval segment then it is considered normal,

otherwise it is considered arrhythmic. Several values were used for n: 0, 1, 2 and 3 RR intervals and 5%, 10% and 15% of the segment length (i.e. for 128 RR interval segment 6, 13 and 19, which are 5%, 10% and 15% of 128, also used as n values).

3.5 QRS Complex Detection by Non Linear Thresholding:

Electrocardiogram (ECG) signal is used to analyze the cardiovascular activity in the human body and has a primary role in the diagnosis of several heart diseases. The QRS complex is the most distinguishable component in the ECG. Therefore, the accuracy of the detection of QRS complex is crucial to the performance of subsequent machine learning algorithms for cardiac disease classification. The aim of the present work is to detect QRS wave from ECG signals. Wavelet transform filtering is applied to the signal in order to remove baseline drift, followed by QRS localization. By using the property of R peak, having highest and prominent amplitude, we have applied Thresholding technique based on the median absolute deviation(MAD) of modulus maxima's to detect the complex. In order to evaluate the algorithm, the analysis has been done on MIT-BIH Arrhythmia database. The results have been examined and approved by medical doctors.

3.5.1 BASELINE WANDERING REMOVAL BY WAVELET DECOMPOSITION:

In practical situations, most of the signal contains significant amount of noise elements which does not allow the accurate analysis of signal, namely power line interference, baseline drift (caused by patient breathing, bad electrodes or improper electrode site) and electromyogram. Reduction of baseline drift is desirable for implementing amplitude threshold strategy. Wavelet transform has been used as a tool for analyzing the bio medical signals, which have a tendency to change their statistical properties with time. As these baseline wandering corresponds to the low frequency element, therefore by using the approximation coefficients of the discrete wavelet transform, it is possible to reduce this noise element.

3.5.2 Discrete Wavelet Transform:

The N noisy data are transformed via the discrete wavelet transform, to obtain N noisy wavelet coefficients (y_j;k). The DWT of a signal x is calculated by passing it through a series of filters.

The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass)

$$a_{j+1}[p]=\sum_{N=-\infty}^{\infty} h[n-2p]a_{j[n]}$$

$$d_{j+1}[p]=\sum_{N=-\infty}^{\infty} g[n-2p]a_{j[n]}$$

After little iteration in multistate analysis of DWT, the approximation coefficients correspond to the lower frequency base line drift component of the ECG signal. The zeroing of the scaling coefficients results in removing the effect of baseline wandering. In order to obtain good results the important parameter to consider in this approach is the level of decomposition. The wrong selection of the decomposition level might results in either over fitting effect in baseline approximation or conversely with poor approximation due to high level. The result in this work is based on the empirical analysis of the signal.

3.5.3 QRS detection:

Many methods have been developed in the past for the detection of QRS complex. However, most of these methods suffer mainly from of problems like, QRS waveform varies from patient to patient, noise and QRS complexes pass band overlaps, noise removal technique results in modification of original signal which ultimately results in false detection. The effect of threshold which is used to detect QRS complexes is limited for the variability of QRS waveforms for different beats of the same subject. Therefore in order to overcome these problems, we are proposing to decompose the signal and by using the amplitude property of modulus maxima's of R peak, we can apply the median absolute deviation Thresholding to classify these maxima of R peaks from the rest. The localization problem of continuous wavelet transform has been solved by selecting a window within the cone of influence of respective maxima's

a) Continuous Wavelet Transform:

The Morlet-Grossmann definition of the continuous wavelet transform for a 1-dimensional signal f(x), the space of all square integral functions, is

$$WT(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(x)\varphi * \frac{(x-b)}{a}dx$$

Where: -WT (a, b) is the wavelet coefficient of the function f (x) - x is the analyzing wavelet - a (>0) is the scale parameter - b is the position parameter.

b) Wavelet Transform Modulus Maxima (WTMM):

The Modulus maximum is termed as, any point (u0; s0) such that We (u; s0) j is locally maximum at u = u0 and the maxima line to any connected curve S(u) in the scale-space plane (u; s) along which all points are modulus maxima. These modulus maxima's correspond to the significant point or peaks in the signal, as illustrated in the signal, R peak have distinctive maxima's even at coarser scales.

c) Threshold Selection:

The threshold criteria are based on MAD of the negative maxima (m) of the decomposed signal. The threshold is computed and applied on the 10th scale of decomposition

$$\text{Thresh} = \sigma\sqrt{2\log(n)}$$

$$\sigma = \frac{\text{MAD}(m)}{0.6745}$$

Where n is total number of samples which will converge to the expected R peaks present in the signal

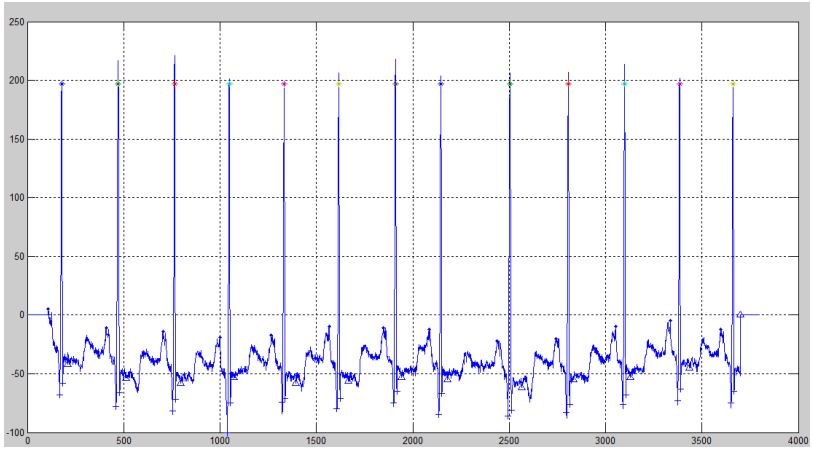

Fig. 3.4 Expected R peaks in the ECG signal

3.6 Conclusion:

We have proposed a new method to detect the QRS complex by wavelet based approaches. Wavelet transform has always been considered as a significant tool in the analysis of a non-stationary signal. By using Gaussian wavelet, we have identified the interest point by relating them to the local extreme in wavelet transform. The proposed threshold criteria based on median absolute deviation of maxima can reliably predict the position of R peaks. Continuous wavelet transform suffer from the problem of delocalization which has been taken care of by selecting a window around the maxima. At the initial stages the algorithm has been tested on part of MIT-BIH arrhythmia database. In future we are proposing an algorithm based on these modulus maxima lines to compute the duration of QRS complex. At the same time left bundle branch block and right bundle branch block can also be detected by tracking these maxima.

A novel and effective method is proposed to classify the seven most important heart arrhythmias. We based our method on HRV signal which is more robust to the noise and has better chaotic characteristics.

Chapter 4

Neural Network

In computer science and related fields, artificial neural networks are models inspired by animal central nervous systems (in particular the brain) that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network.

For example, in a neural network for handwriting recognition, a set of input neurons may be activated by the pixels of an input image representing a letter or digit. The activations of these neurons are then passed on, weighted and transformed by some function determined by the network's designer, to other neurons, etc., until finally an output neuron is activated that determines which character was read. The inspiration for neural networks came from examination of central nervous systems. In an artificial neural network, simple artificial nodes, called "neurons", "processing elements" or "units", are connected together to form a network which mimics a biological neural network. There is no single formal definition of what an artificial neural network is. Commonly, though, a class of statistical models will be called "neural" if they

1. Consist of sets of adaptive weights, i.e. numerical parameters that are tuned by a learning algorithm.
2. They are capable of approximating non-linear functions of their inputs.

The adaptive weights are conceptually connection strengths between neurons, which are activated during training and prediction. Neural networks are also similar to biological neural networks in performing functions collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned. The term "neural network" usually refers to models employed in statistics, cognitive psychology and artificial intelligence. Neural network models which emulate the central nervous system are part of theoretical neuroscience and computational neuroscience.

In modern software implementations of artificial neural networks, the approach inspired by biology has been largely abandoned for a more practical approach based on statistics and signal processing. In some of these systems [21] [22], neural networks or parts of neural networks (like artificial neurons) form components in larger systems that combine both adaptive and non-adaptive elements.

4.1 Neural network in ECG:

The electrocardiogram examination has been widely used in clinical diagnosis of cardiovascular diseases owing to its non-invasiveness and reliability. In many cases, such as Intensive Care Unit (ICU), or Holter System, long-time recording and analyzing of ECG are required and automatic ECG pattern recognition techniques helps to reduce physicians' work and improve diagnosis efficiency. A typical ECG waveform contains P wave, QRS complex and T wave in each heart beat (Fig.1). Recognizing an ECG pattern is essentially the process of extracting and classifying ECG feature parameters, which may be obtained either from the time domain or transform domain. The features being frequently used for ECG analysis in time domain include the wave shape, amplitude, duration, areas, and R-R intervals [25]. The basic problem of automatic ECG analysis occurs from the non-linearity in ECG signals and the large variation in ECG morphologies of different patients. And in most cases, ECG signals are contaminated by background noises, such as electrode motion artifact and electromyogram-induced noise, which also add to the difficulty of automatic ECG pattern recognition [26].

Work flow of Neural Network

- Collect data
- Target detection
- Create the network
- Configure the network
- Initialize the weights and biases
- Train the network
- Validate the network (post-training analysis)
- Use the network

4.1.1 Collecting the data:

We've used the MIT-BIH database for our thesis purpose. The MIT-BIH Arrhythmia Database [27] was the first generally available set of standard test material for evaluation of arrhythmia detectors, and it has been used for that purpose as well as for basic research into cardiac dynamics at about 500 sites worldwide since 1980. It has lived a far longer life than any of its creators ever expected. Together with the American Heart Association Database, it played an interesting role in stimulating manufacturers of arrhythmia analyzers to compete on the basis of objectively measurable performance, and much of the current appreciation of the value of common databases, both for basic research and for medical device development and evaluation, can be attributed to this experience.

4.1.2 Target Detection:

Scientists and engineers often need to know if a particular object or condition is present. For instance, geophysicists explore the earth for oil, physicians examine patients for disease, astronomers search the universe for extraterrestrial intelligence, etc. These problems usually involve comparing the acquired data against a threshold. If the threshold is exceeded, the target (the object or condition being sought) is deemed present. For example, suppose you invent a device for detecting cancer in humans. The apparatus is waved over a patient, and a number between 0 and 30 pops up on the video screen [28]. Low numbers correspond to healthy subjects, while high numbers indicate that cancerous tissue is present. You find that the device works quite well, but isn't perfect and occasionally makes an error. The question is: how do you use this system to the benefit of the patient being examined?

Figure 4.1 illustrates a systematic way of analyzing this situation. Suppose the device is tested on two groups: several hundred volunteers known to be healthy (non-target), and several hundred volunteers known to have cancer (target). Figures (a) & (b) show these test results displayed as histograms. The healthy subjects generally produce a lower number than those that have cancer (good), but there is some overlap between the two distributions (bad does not specify if the output is a real number, requiring a pdf, or an integer, requiring a pmf .

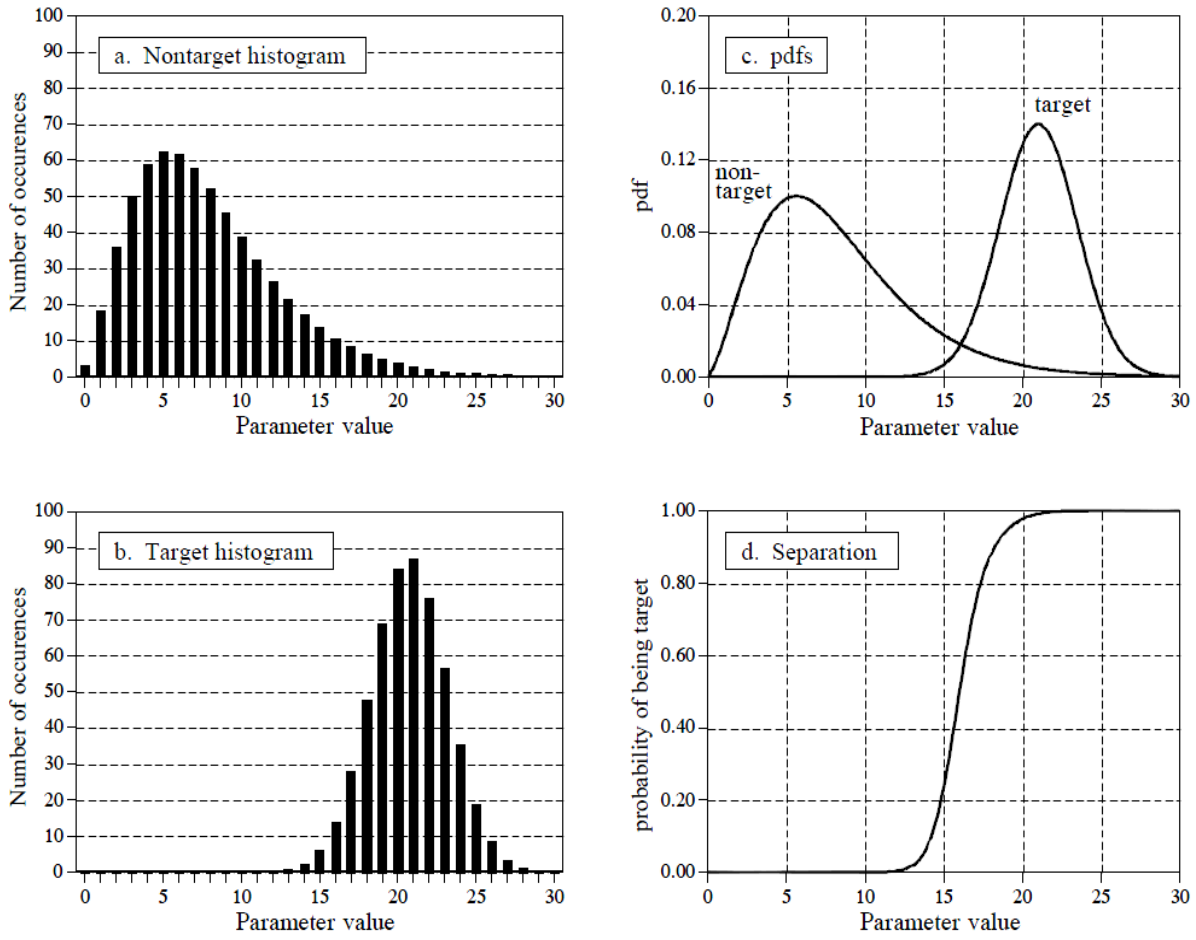

Fig. 4.1 Probability of target detection

For example, if a patient has cancer, and the test properly detects the condition, it is said to be a true-positive. Likewise, if a patient does not have cancer, and the test indicates that 454 The Scientist and Engineer's Guide to Digital Signal Processing cancer is not present, it is said to be a true-negative. A false-positive occurs when the patient does not have cancer, but the test erroneously indicates that they do. This results in needless worry, and the pain and expense of additional tests. An even worse scenario occurs with the false negative [29], where cancer is present, but the test indicates the patient is healthy. As we all know, untreated cancer can cause many health problems, including premature death.

The human suffering resulting from these two types of errors makes the threshold selection a delicate balancing act. How many false-positives can be tolerated to reduce the number of false-negatives? Figure 4.3 shows a graphical way of evaluating this problem, the ROC curve (short for Receiver Operating Characteristic). The ROC curve plots [30] the percent of target signals

reported as positive (higher is better), against the percent of non-target signals erroneously reported as positive (lower is better), for various values of the threshold. In other words, each point on the ROC curve represents one possible tradeoff of true-positive and false-positive performance. Figures (a) through (d) show four settings of the threshold in our cancer detection example. For instance, look at (b) where the threshold is set at 17 [31].

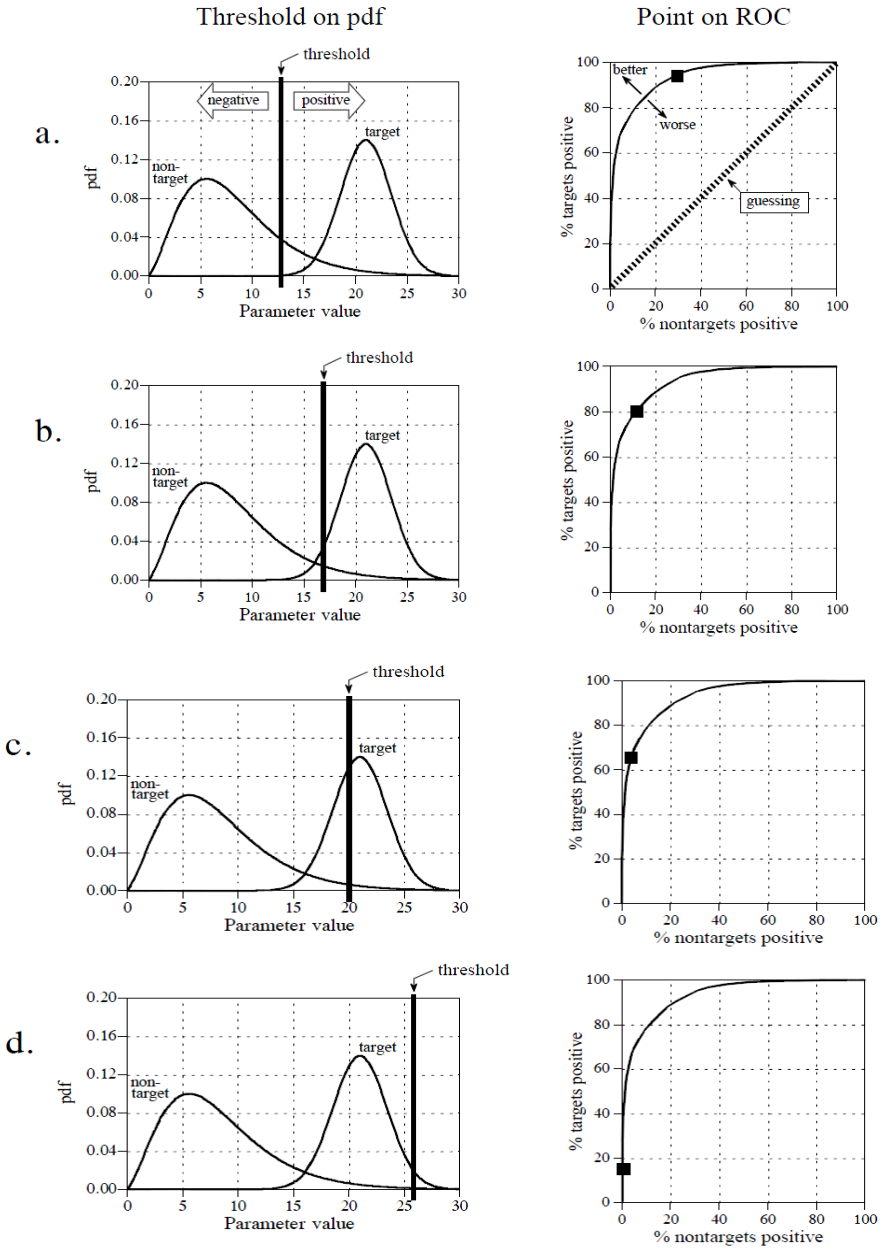

fig 4.2Relationship between ROC curves and pdfs.

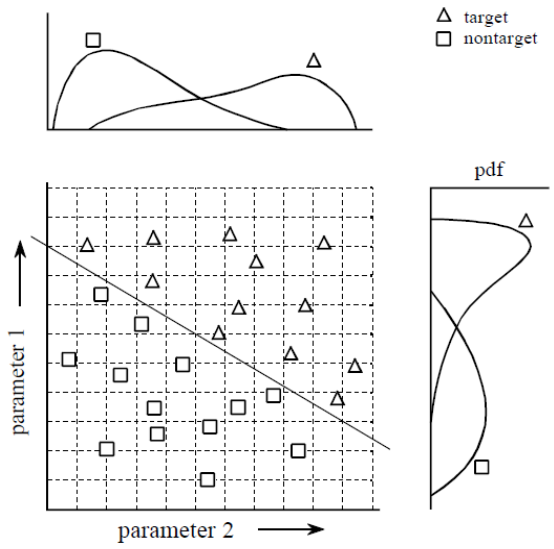

Fig. 4.2(a): Example of a two-parameter space.

This analysis can be extended to devices that provide more than one output. For example, suppose that a cancer detection system operates by taking an x-ray image of the subject, followed by automated image analysis algorithms to identify tumors. As a first try, we could go through the previously presented ROC analysis for each parameter, and find an acceptable threshold for each. We could then classify a test as positive only if it met both criteria: parameter 1 greater than some threshold and parameter 2 greater than another threshold. This technique of Thresholding the parameters separately and then invoking logic functions (AND, OR, etc.) is very common [32].

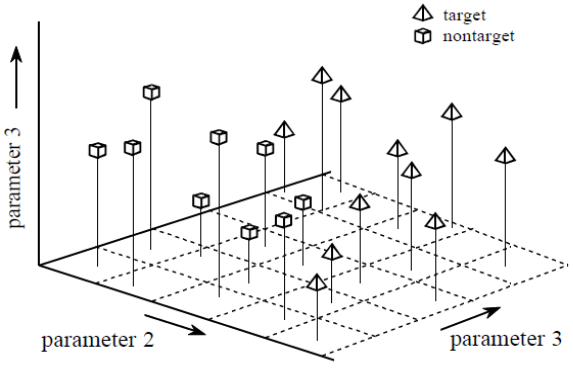

Fig. 4.3: Example of a three-parameter

In the jargon of the field, this type of coordinate system is called a parameter space. For example, the two-dimensional plane in this example could be called a diameter-brightness space. The idea is that targets will occupy one region of the parameter space, while non-targets will occupy another. Separation between the two regions may be as simple as a straight line, or as complicated as closed regions with irregular borders. Figure 4.4 shows the next level of complexity, a three-parameter space being represented on the x, y and z axes. For example, this might correspond to a cancer detection system that measures diameter, brightness, and some third parameter, say, edge sharpness. Just as in the two-dimensional case, the important idea is that the members of the target and non-target groups will (hopefully) occupy different regions of the space, allowing the two to be separated. In three dimensions, regions are separated by planes and curved surfaces. The division is then written into a computer program as an equation, or some other way of defining one region from another. In principle, this same technique can be applied to a three-dimensional parameter space. .

4.2 Neural Network Architecture:

Humans and other animals process information with neural networks. These are formed from trillions of neurons (nerve cells) exchanging brief electrical pulses called action potentials. Computer algorithms that mimic these biological structures are formally called artificial neural networks to distinguish them from the squishy things inside of animals. However, most scientists and engineers are not this formal and use the term neural network to include both biological and no biological systems. Input layer hidden layer output layer and neural network architecture [33]. This is the most common structure for neural networks: three layers with full interconnection. The input layer nodes are passive, doing nothing but relaying the values from their single input to their multiple outputs Imitating what a biologist sees under the microscope some based on a more mathematical analysis of the problem [34]. The most commonly used structure is shown in Fig. 4.5. This neural network is formed in three layers, called the input layer, hidden layer, and output layer. Each layer consists of one or more nodes, represented in this diagram by the small circles.

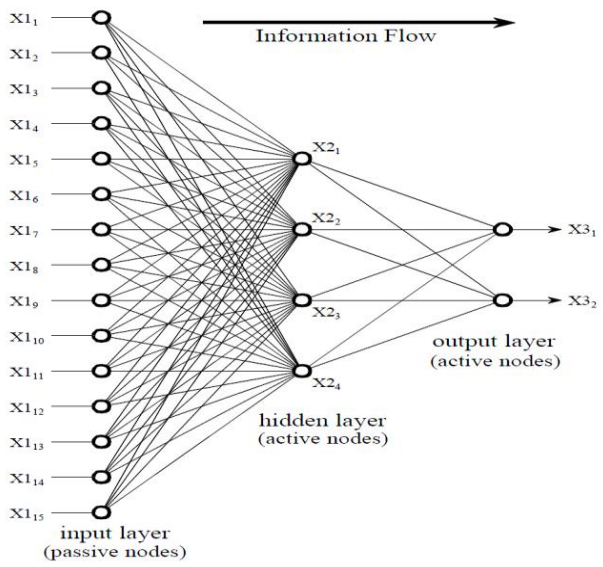

Fig. 4.4: Neural network architecture.

In comparison, the nodes of the hidden and output layer are active. They may also be the output of some other algorithm, such as the classifiers in our cancer detection example: diameter, brightness, edge sharpness, etc. Each value from the input layer is duplicated and sent to all of the hidden nodes. This is called a fully interconnected structure.

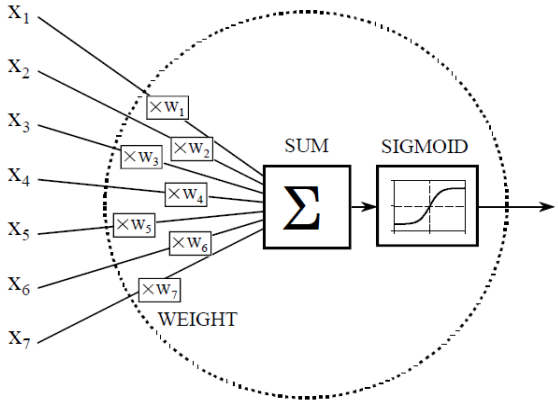

Fig. 4.5: Neural network active node.

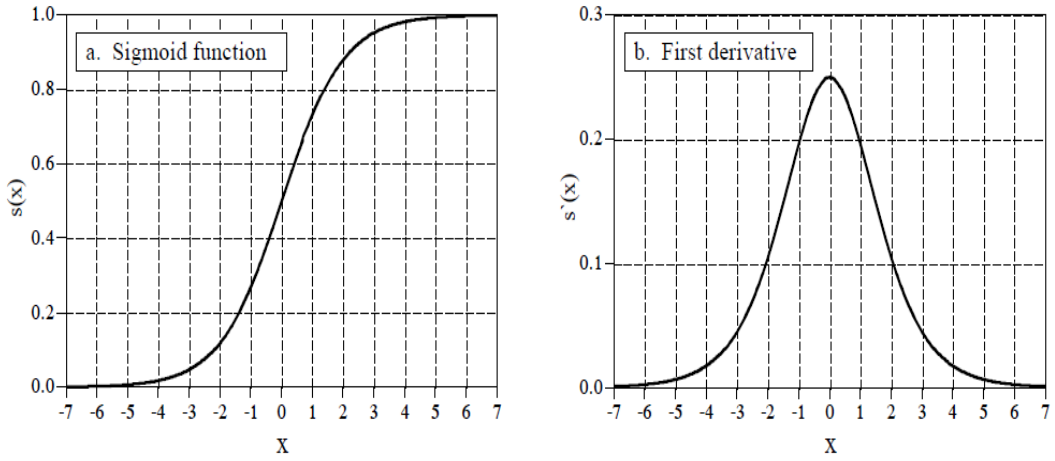

Fig. 4.6: The sigmoid function and its derivatives.

The sigmoid function is used in neural networks as a smooth threshold. This function is graphed in Fig. 26-7a. Figure 26-7a shows a closer look at the sigmoid function, mathematically Described by the equation:

$$s(x) = \frac{1}{1 + e^{-x}}$$

The exact shape of the sigmoid is not important, only that it is a smooth threshold. For comparison, a simple threshold produces a value of one when x > 0, and a value of zero when x < 0. An advantage of the sigmoid is that there is a shortcut to calculating the value of its derivative:

$$s'(x) = s(x)\,[\,1 - s(x)\,]$$

This is the first derivative of the sigmoid function. This is calculated by using the value of the sigmoid function itself. This isn't a critical concept, just a trick to make the algebra shorter [35]. If the sigmoid were not present, the three layers would collapse into only two layers. In other words, the summations and weights of the hidden and output layers could be combined into a single layer, resulting in only a two-layer network.

4.2.1 Why Does It Work?

The weights required to make a neural network carry out a particular task are found by a learning algorithm, together with examples of how the system should operate. The learning algorithm uses these examples to calculate a set of weights appropriate for the task at hand [36].

.

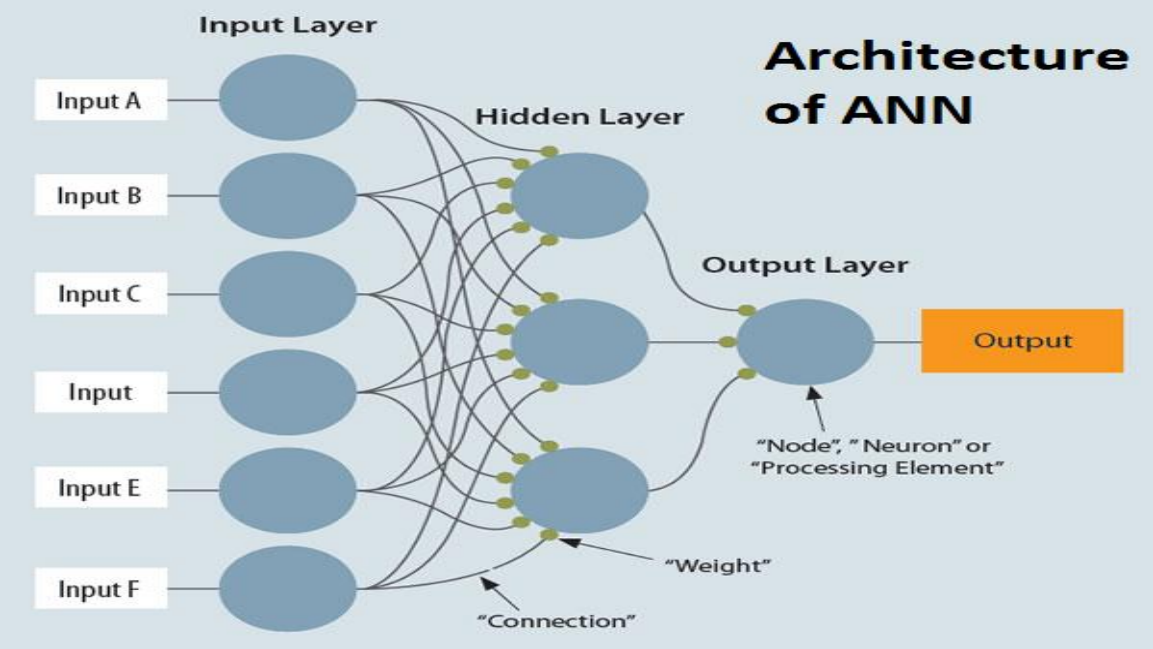

Fig. 4.7: neural network architecture

For example, consider the sonar system neural network with 1000 inputs and a single output. With proper weight selection, the output will be near one if the input signal is an echo from a submarine and near zero if the input is only noise. This forms a parameter hyperspace of 1000 dimensions. For instance, these simple calculations would indicate that a neural network with 1000 inputs needs 2000 weights to identify one region of the hyperspace from another. This means neural networks can carry out nonlinear as well as linear processing. Suppose that one of these conventional DSP strategies is used to design the weights of a neural network.

4.2.2 Training the Neural Network:

Neural network design can best be explained with an example. Fig4.8 shows the problem we will attack, identifying individual letters in an image of text. This pattern recognition task has received much attention. It is easy enough that many approaches achieve partial success, but

difficult enough that there are no perfect solutions. Many successful commercial products have been based on this problem, such as: reading the addresses on letters for postal routing, document entry into word processors, etc. The first step in developing a neural network is to create a database of examples. For the text recognition problem, this is accomplished by printing the 26 capital letters: A,B,C,D… Y,Z, 50 times on a sheet of paper [37]. Next, these 1300 letters are converted into a digital image by using one of the many scanning devices available for personal computers. This large digital image is then divided into small images of 10×10 pixels, each containing a single letter.

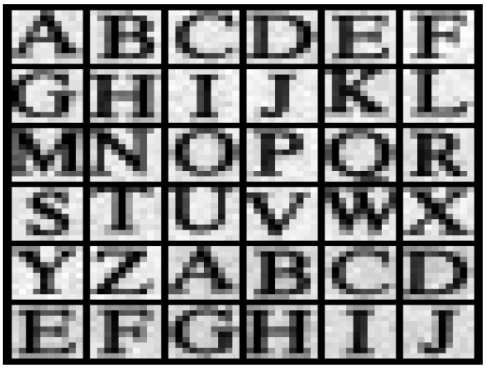

Fig. 4.8: Example of image text.

In this example, each letter is contained in a 10×10 pixel image, with 256 gray levels per pixel. The database used to train and test the example neural network consists of 50 sets of the 26 capital letters, for a total of 1300 images. The images shown here is a portion of this database.

For this demonstration, the neural network will be designed for an arbitrary task: determine which of the 10×10 images contains a vowel, i.e., A, E, I, O, or U. This may not have any practical application, but it does illustrate the ability of the neural network to learn very abstract pattern recognition problems. By including ten examples of each letter in the training set, the network will (hopefully) learn the key features that distinguish the target from the non-target images. For instance, when 101 inputs with a typical value of 100 are multiplied by the typical weight value of 0.0002, the sum of the products is about 2, which is in the active range of the sigmoid input. If we evaluated the performance of the neural network using these random weights, we would expect it to be the same as random guessing. The learning algorithm improves the performance of the network by gradually changing each weight in the proper direction. This is called an iterative procedure, and is controlled in the program by the FOR-NEXT loop in lines 270-400. For each iteration makes the weights slightly more efficient at

separating the target from the non-target examples. The iteration loop is usually carried out until no further improvement is being made. In typical neural networks, this may be anywhere from ten to ten-thousand iterations, but a few hundred is common. This example carries out 800 iterations. In order for this iterative strategy to work, there must be a single parameter that describes how well the system is currently performing. The variable ESUM (for error sum) serves this function in the program. The first action inside the iteration loop is to set ESUM to zero (line 290) so that it can be used as an accumulator. At the end of each iteration, the value of ESUM is printed to the video screen (line 380), so that the operator can insure that progress is being made. Line 2220 shows an option that is often included when calculating the error: assigning a different importance to the errors for targets and non-targets. For example, we take the detecting the cancer and the consequences of making a false-positive error versus a false-negative error. In the present example, we will arbitrarily declare that the error in detecting a target is five times as bad as the error in detecting a non-target. In effect, this tells the network to do a better job with the targets, even if it hurts the performance of the non-targets. Subroutine 3000 is the heart of the neural network strategy, the algorithm for changing the weights on each iteration. We will use an analogy to explain the underlying mathematics. Consider the predicament of a military paratrooper dropped behind enemy lines. He parachutes to the ground in unfamiliar territory, only to find it is so dark he can't see more than a few feet away. His orders are to proceed to the bottom of the nearest valley to begin the remainder of his mission. The problem is, without being able to see more than a few feet, how does he make his way to the valley floor? Put another way, he needs an algorithm to adjust his x and y position on the earth's surface in order to minimize his elevation. This is analogous to the problem of adjusting the neural network weights, such that the network's error, ESUM, is minimized. We will look at two algorithms to solve this problem: evolution and steepest descent. In evolution, the paratrooper takes a flying jump in some random direction. If the new elevation is lower, he feels a measure of success, and repeats the process from the new location. Eventually he will reach the bottom of the valley, although in a very inefficient and haphazard path. This method is called evolution because it is the same type of algorithm employed by nature in biological evolution. Each new generation of a species has random variations from the previous. .

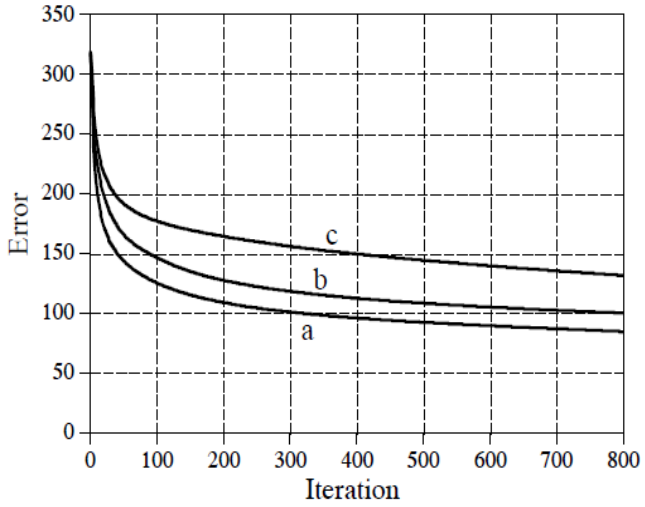

Fig. 4.9: Neural network convergence

Using these two values, he can determine which direction is downhill. This is done by calculating the slope for each weight, and then changing each weight by an amount proportional to that slope. In the paratrooper case, the slope along an axis is found by moving a small distance along the axis (say, Δx), measuring the change in elevation (say, ΔE), and then dividing the two (ΔE/Δx). The slope of a neural network weight can be found in this same way: add a small increment to the weight value (Δw), find the resulting change in the output signal (ΔX3), and divide the two (ΔX3/Δw). Later in this chapter we will look at an example that calculates the slope this way. However, in the present example we will use a more efficient method. Earlier we said that the nonlinearity (the sigmoid) needs to be differentiable. Here is where we will use this property. If we know the slope at each point on the nonlinearity, we can directly write an equation for the slope of each weight (ΔX3/Δw) without actually having to perturb it. Consider a specific weight, for example, WO, corresponding to the first input of the output node. The answer is:

$$\frac{\Delta X3}{\Delta w} = X2[1]\ \text{SLOPE}_{O}$$

Slope of the output layer weights. This equation is written for the weight, WO Where SLOPEO is the first derivative of the output layer sigmoid, evaluated where we are operating on its curve. Using a similar analysis, the slope for a weight on the hidden layer, such as WH[1,1], can be found by:

$$\frac{\Delta X3}{\Delta w} = X1[1]\ \text{SLOPE}_{H1}\ WO[1]\ \text{SLOPE}_{O}$$

This calculation of

$$w_{new} = w_{old} + \frac{\Delta X3}{\Delta w}\ \text{ELET}\ \text{MU}$$

By updating the weights, each of the weights is adjusted by adding an amount proportional to the slope of the weight. The value to use depends on the particular problem, being as low as 10-6, or as high as 0.1. From the analogy of the paratrooper, it can be expected that too small of a value will cause the network to converge too slowly. In comparison, too large of a value will cause the convergence to be erratic, and will exhibit chaotic oscillation around the final solution. Unfortunately, the way neural networks react to various values of MU can be difficult to understand or predict. This makes it critical that the network error (i.e., ESUM) be monitored during the training, such as printing it to the video screen at the end of each iteration. If the system isn't converging properly, stop the program and try another value for MU.

4.3 Evaluating the Results:

So, how does it work? The training program for vowel recognition was run three times using different random values for the initial weights. About one hour is required to complete the 800 iterations on a 100 MHz Pentium personnel computer. Figure 4.10 shows how the error of the network, ESUM, changes over this period. Just as some valleys are deeper than others, some neural network solutions are better than others. This means that the learning algorithm should be run several times, with the best of the group taken as the final solution.

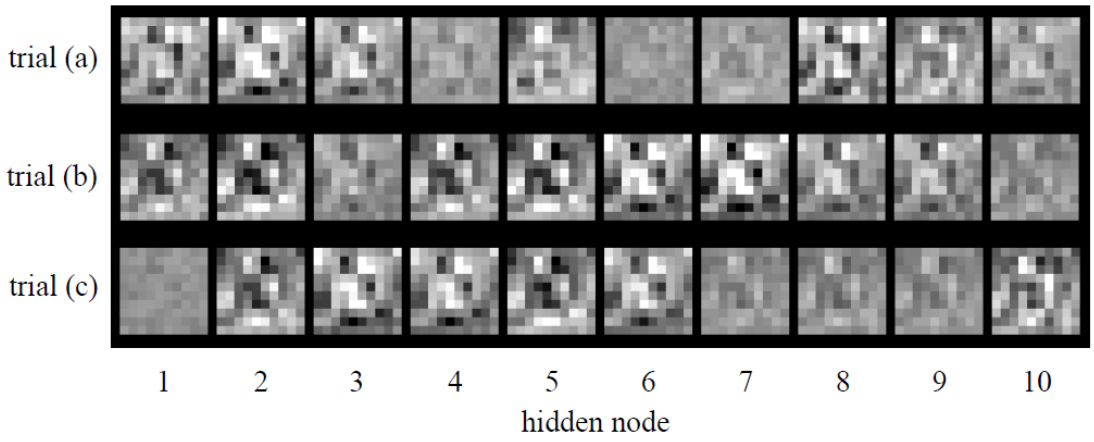

Fig. 4.10: Example of neural network weights.

In Fig. 4.10, the hidden layer weights of the three solutions are displayed as images. They look like random noise! These weights values can be shown to work, but why they work is something of a mystery. Here is something else to ponder. The human brain is composed of about 100 trillion neurons, each with an average of 10,000 interconnections. Letters in the training set. Remember, the weights were selected to make the output near one for vowel images, and near zero otherwise. Separation has been perfectly achieved, with no overlap between the two distributions. Also notice that the vowel distribution is narrower than the non-vowel distribution. This is because we declared the target error to be five times more important than the non-target error.

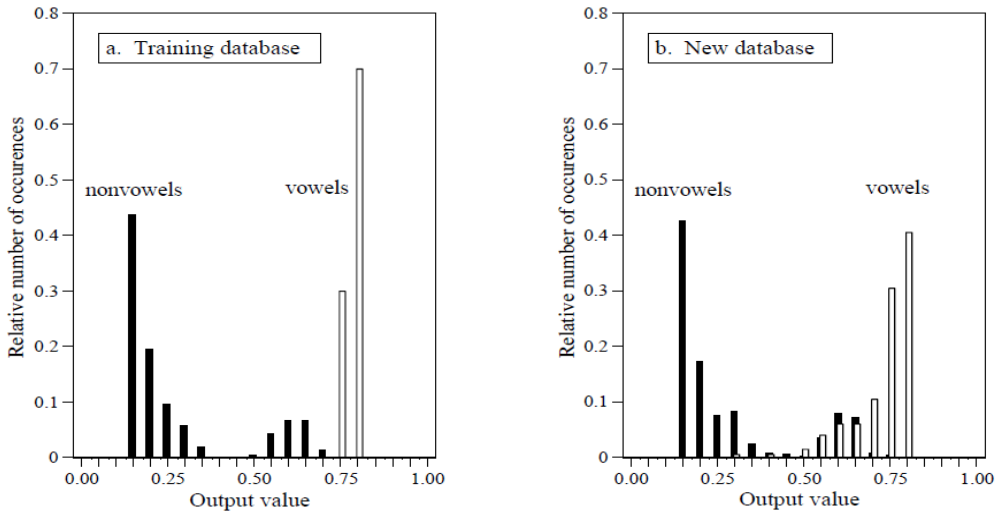

Fig. 4.11(a): Neural network performance.

This is a matter of random chance depending on the initial weights used. At one threshold setting, the neural network designed in trial "b" can detect 24 out of 25 targets (i.e., 96% of the vowel images), with a false alarm rate of only 1 in 25 non-targets (i.e., 4% of the non-vowel images . If the network error (ESUM) doesn't steadily decrease, the program must be terminated, changed, and then restarted. This may take several attempts before success is reached. Three things can be changed to affect the convergence: (1) MU, (2) the magnitude of the initial random weights, and (3) the number of hidden nodes (in the order they should be changed).

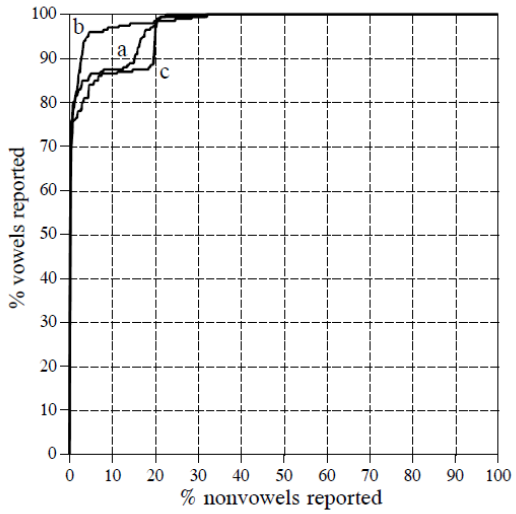

Fig. 4.11(b): ROC analysis of neural network examples

Some final comments on neural networks: Getting a neural network to converge during training can be tricky. If the network error (ESUM) doesn't steadily decrease [38], In mathematics and computing, the Levenberg–Marquardt algorithm (LMA),[1] also known as the damped least-squares (DLS) method, provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function. LMA can also be viewed as Gauss–Newton using a trust region approach. The LMA is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems. However, the LMA finds only a local minimum, not a global minimum. LMA is a hybrid algorithm that is based on both Newton's Method and Gradient Descent (back propagation). This allows LMA to have strengths of both. Gradient Descent is guaranteed to converge to a local minimum, however, it is quite slow. GNA is quite fast but often fails to converge. By using a damping

factor to interpolate between the two, a hybrid method is created. To understand how this works we first examine Newton's method. Newton's Method is shown here.

$$W_{min} = W_0 - H^{-1}g$$

Newton's Method will converge the weights of a neural network to local minima, local maxima or a straddle position. This is done by minimizing all of the gradients (first derivatives) to zero. The derivatives will all be zero at local minima, maxima or straddle position. The Hessian matrix is typically estimated. There are various means of doing this. If the Hessian is inaccurate this can greatly throw off Newton's Method. LMA enhances Newton's Algorithm to the following formula.

$$W_{min} = W_0 - (H + \lambda I)^{-1}g$$

Here we add a damping factor multiplied by an identity matrix. Lambda is the damping factor and I represent the identity matrix. An identity matrix is a square matrix with all zeros except for a NW line of ones). As lambda increases the hessian will be factored out of the above equation. As lambda decreases the hessian becomes more significant than gradient descent. This allows the training algorithm to interpolate between gradient descent and Newton's Method. Higher lambda favors gradient descent, lower lambda favors Newton. A training iteration of LMA begins with a low lambda and increases are until a desirable outcome is produced. This can be seen in the following flowchart.

4.4 Calculation of the Hessian:

The Hessian matrix is a square matrix with rows and columns equal to the number of weights in the neural network. Each cell in this matrix represents the second order derivative of the output of the neural network with respect to a given weight combination. The Hessian is shown here.

$$H(e) = \begin{bmatrix} \frac{\partial^2 e}{\partial w_1^2} & \frac{\partial^2 e}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 e}{\partial w_1 \partial w_n} \\ \frac{\partial^2 e}{\partial w_2 \partial w_1} & \frac{\partial^2 e}{\partial w_2^2} & \cdots & \frac{\partial^2 e}{\partial w_2 \partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 e}{\partial w_n \partial w_1} & \frac{\partial^2 e}{\partial w_n \partial w_2} & \cdots & \frac{\partial^2 e}{\partial w_n^2} \end{bmatrix}.$$

It is important to note that the Hessian is symmetrical about the diagonal. This can be used to enhance performance of the calculation. Calculation of the Hessian can be accomplished by calculating the gradients.

$$\frac{\partial E}{\partial w_{(i)}} = 2(y-t)\frac{\partial y}{\partial w_{(i)}}$$

The second derivative of above equation becomes an element of the Hessian matrix. This is calculated with the following formula.

$$\frac{\partial^2 E}{\partial w_i w_j} = 2\left(\frac{\partial y}{\partial w_i}\frac{\partial y}{\partial w_j} + (y-t)\frac{\partial^2 y}{\partial w_j \partial w_j}\right)$$

The above formula can easily be calculated if not for the second component. This component involves the second partial derivative is difficult to calculate [39]. This component is actually not that important and can be dropped. For an individual training case it might be very important. However, the second component is multiplied by the error of that training case. We assume that the errors in a training set are independent and evenly distributed about zero. On an entire

training set they should essentially cancel each other out. This is not a perfect assumption. However, we only seek to approximate the Hessian. This results in the following formula.

$$\frac{\partial^2 E}{\partial w_i w_j} = 2\left(\frac{\partial y}{\partial w_i}\frac{\partial y}{\partial w_j}\right)$$

Using the above equation is, of course, only an approximation of the true Hessian. However, the simplification of the algorithm to calculate the second derivative is well worth the loss in accuracy. The primary difference is that the derivative of the output is taken for the Hessian. In standard back propagation the derivative of the error function is taken. Encog uses a multi-threaded approach to calculate all of the first derivatives needed to calculate the Hessian. The first derivatives are transformed into the Hessian using the above formulas.

4.4.1 Levenberg–Marquardt Algorithm with Multiple Outputs:

Some implementations of the LMA algorithm only support a single output neuron. This is primarily because the LMA algorithm has its roots in mathematical function approximation. In mathematics functions typically only return a single value. As a result, there is not a great deal of information on support for multiple neuron support. Encog's implementation of LMA supports multiple output neurons.

4.4.2 Overview of the LMA Process:

The LMA process can be summarized as the following steps.

1. Calculate the first derivative of output of the neural network with respect to every weight
2. Calculate the Hessian
3. Calculate the gradients of the error(ESS) with respect to every weight
4. Either set lambda to a low value(first iteration) or the lambda of the previous iteration
5. Save the weights of the neural network
6. Calculate delta weight based on the lambda, gradients and Hessian
7. Apply the deltas to the weights and evaluate error
8. If error has improved, end the iteration

4.5 Our work:

So we have used and applied the neural network analysis method to calculate the ECG arrhythmia detection rate more efficiently. We have used the MIT-BIH database mainly and we have used MATLAB 12b for our simulation purpose.

4.5.1 Neural Network Model:

We have created a neural network model which has main attributes given below:

- No. of Hidden Neuron used= 10
- Training 60% of the total samples
- Validation 10% of the samples
- Testing 30% of the samples

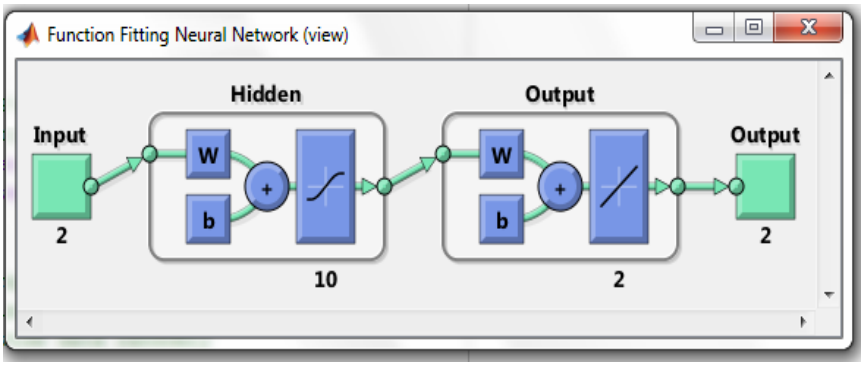

Fig. 4.12: neural network

4.5.2 MIT-BIH Database:

In the proposed method, we have employed MIT-BIH (Massachusetts Institute of Technology-Boston's Beth Israel Hospital) arrhythmia database [40]. The ECG recordings are sampled at 360 samples per second per channel with 11-bit resolution over a 10 mV range. But normal beats are frequently difficult to discern in the lower signal of each ECG record, on the other hand, normal QRS complexes are usually prominent in the upper signal. Thus the ECG records of upper signal namely modified limb lead 2 (ML2) are used in this work, from which the heartbeat waveform is extracted by taking 50 samples before the R peak and 150 samples after the R peak thus forming

201 samples of ECG beat. From the database, we have considered normal ECG beats (N) and that belonging to five types of arrhythmia, namely left bundle branch block beat (L), right bundle branch block beat ®, atrial premature beat (A), premature ventricular contraction (V), and paced beat (PB). The number of the ECG file that is used from the MIT-BIH database and the number of QRS complexes taken from each type of class including normal and arrhythmia are shown. It is seen from this table that we have employed ECG signals from 22 patients. For each class, 32 beats are considered resulting in a total 192 ECG beats. Out of these ECG beats, approximately 38% are selected for training and the remaining is left for testing or validating the classifiers.

4.5.3 Regression Plot Difference:

During the simulation process we have had couple of simulation windows. One of those are the results due to Regression plot difference between training and testing dataset. The Regression plot gives us an idea about the dataset regression.

In the figure given below we have shown the simulation of MIT-BIH database file 113.m.

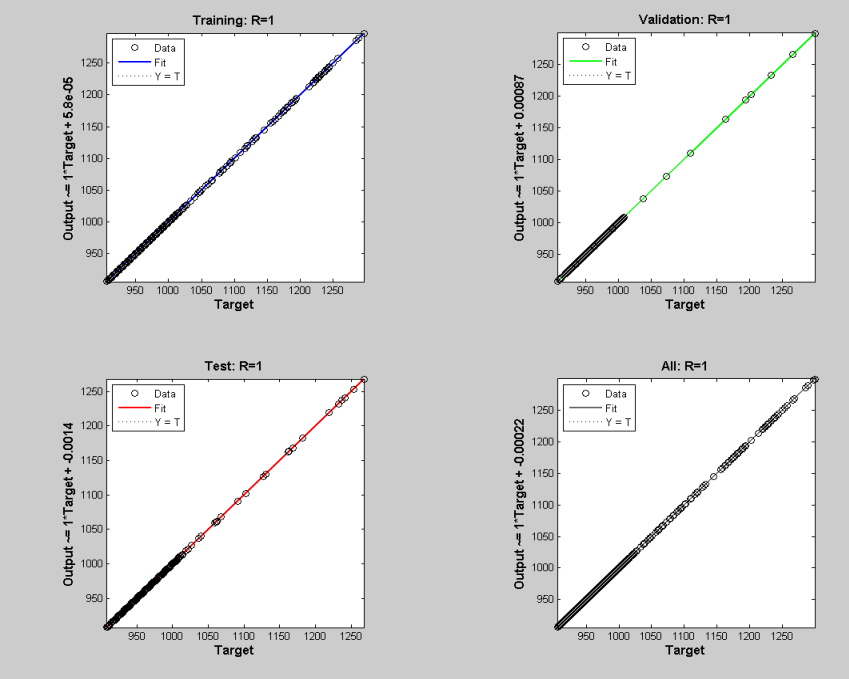

Fig. 4.13: Regression plot

4.5.4 Performance Curve:

The performance curve of the 113.m file is shown below.

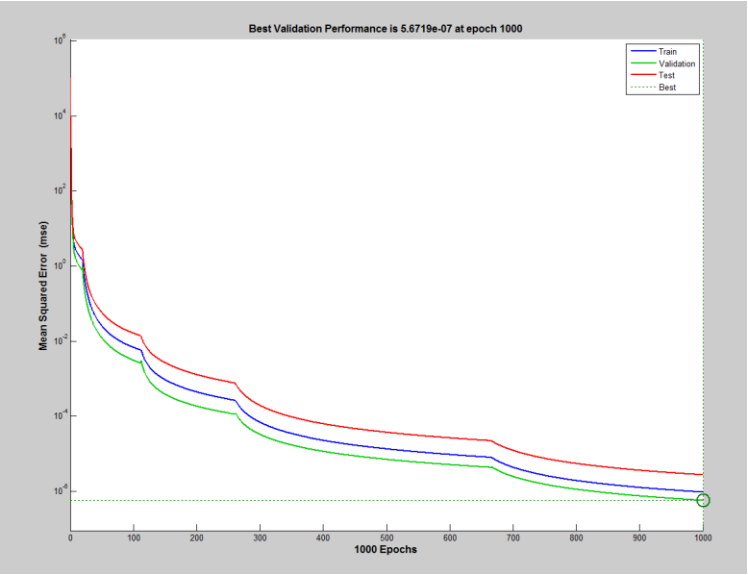

Fig. 4.14: performance curve

4.5.5 MATLAB "nntraintool":

The figure of nntraintool (MATLAB neural network simulation tool) is shown below which is based on 113.m file of MIT-BIH database.

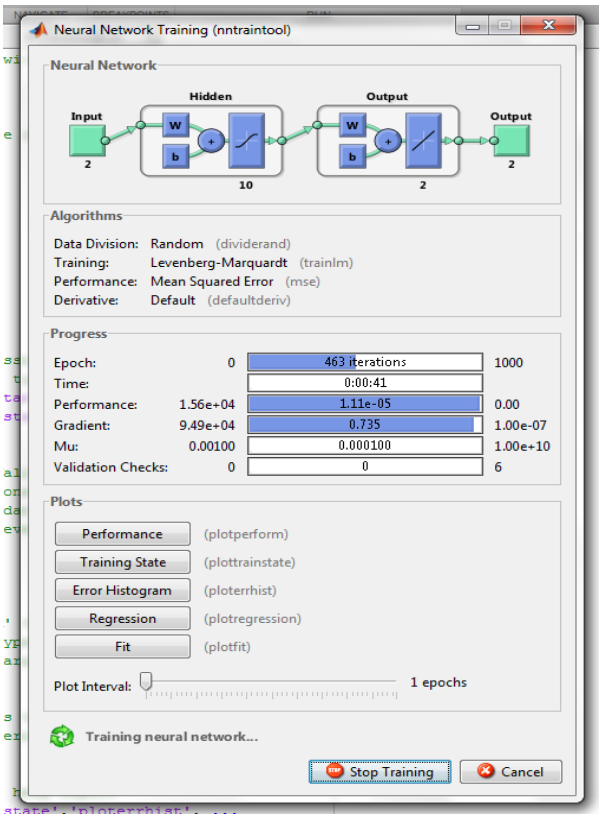

Fig. 4.15: Neural network training tool

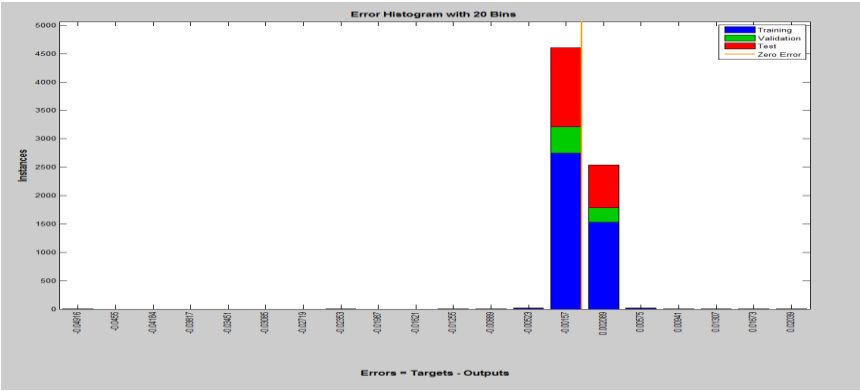

Fig. 4.16: Error histogram

4.5.6 Data calculation Chart of ECG normal beat:

Type	File no.	Training Data	Testing Data	Training Performance	Validation Performance	Testing Performance
Normal	101	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	102	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	103	60%	30%	1.72E-10	1.12E-06	3.43E-10
Normal	105	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	108	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	112	60%	30%	1.72E-10	1.12E-06	3.43E-10
Normal	113	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	114	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	115	60%	30%	1.72E-10	1.12E-06	3.43E-10
Normal	117	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	121	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	122	60%	30%	1.72E-10	1.12E-06	3.43E-10
Normal	123	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	202	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	205	60%	30%	1.72E-10	1.12E-06	3.43E-10
Normal	219	60%	30%	9.61E-06	1.04E-05	1.92E-05
Normal	230	60%	30%	1.28E-05	9.98E-06	1.25E-03
Normal	234	60%	30%	1.72E-10	1.12E-06	3.43E-10

Fig. 4.17: data calculation chart

4.5.7 Data calculation chart of ECG abnormal beat:

Abnormal(LBBB)	107	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(LBBB)	111	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(LBBB)	207	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(LBBB)	214	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(RBBB)	118	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(RBBB)	124	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(RBBB)	212	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(RBBB)	231	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PVC)	106	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(PVC)	119	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(PVC)	200	60%	30%	4.62E-06	4.05E-01	4.81E-06
Abnormal(PVC)	203	60%	30%	3.22E-11	4.18E-11	3.50E-11
Abnormal(PVC)	208	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(PVC)	213	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PVC)	221	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(PVC)	228	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(PVC)	233	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PVC)	116	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(PVC)	201	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(PVC)	210	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PVC)	215	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(APB)	209	60%	30%	7.96E-07	5.14E-06	1.10E-06
Abnormal(APB)	222	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(APB)	232	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(APB)	220	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(APB)	223	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PB)	102	60%	30%	3.21E-11	4.18E-11	3.50E-11
Abnormal(PB)	104	60%	30%	9.96E-07	5.14E-06	1.10E-06
Abnormal(PB)	107	60%	30%	4.24E-06	4.05E-01	4.81E-06
Abnormal(PB)	217	60%	30%	5.21E-11	4.18E-11	3.50E-11
Abnormal(VFW)	207	60%	30%	8.96E-07	5.14E-06	1.10E-06

Fig. 4.18: data collection

4.6 Simulation Result:

In comparison with the Normal ECG beats like file 100,101,113 and Abnormal beats like file 119,234,207.We've calculated the accuracy of this Neural Network based method to be around 80%.

Chapter 5

Conclusion

5.1 Concluding Remarks:

Though the accuracy has not reached the expected satisfactory level the speed of this method has surpassed other methods. As the Levenberg-Marquardt algorithm is a fast processing method so the time period taken to calculate the Normality-abnormality of ECG beat has been reduced by a great deal. So the trade-off that our work is facing is the accuracy to speed. Thus we are left with a wide range of possibility to improve the accuracy of our work.

5.2 Contribution of the thesis:

The main contribution of this particular thesis is the new methodology based on Neural Network using Levenberg-Marquardt algorithm by which we have created a bridge between the previous trades off (speed vs. accuracy) which was being faced by other methods to detect ECG arrhythmia.

5.3 Limitations

The main limitation of work is that we have not applied this technique on wide range of datasets of normal and abnormal ECG files of MIT-BIH database. This could have really helped us decide on the capability of this method.

5.4 Future work

In the near future we'll be working on the possibilities which are given below:

- Classification of different arrhythmias by using this method.
- To obtain better performance by modifying the method.
- To create our own database.
- Increasing the accuracy and at the same time not decreasing the detection processing speed

Bibliography

1. Atsushi, M. Hwa, A. Hassankhani, T. Liu, and S. M. Narayan, "Abnormal heart rate turbulence predicts the initiation of ventricular arrhythmias," Pacing Clin. Electrophysiol., vol. 11, pp. 1189–97.
2. M. R. Risk, J. F. Sobh, and J. P. Saul, "Beat detection and classification of ECG using self-organizing maps," in Proc. 19th Int. Conf. IEEE EMBS, 1997, vol. 19, pp. 89–91.
3. L. Y. Shyu, Y. H. Wu, and W. C. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1269–1273, Jul. 2004
4. Z. Benyd L. Czinege, Computer Analysis of Dynamic Systems with Application in Physiology, Proc IS" Worldcongress of MACS on Scient$c Computation, Modelling and Applied Mathematics.Berlin, 1997. Vol. El. pp. 663-668.
5. Abdel-Rahman Al-Qawasmi, Khaled Daqrouq "ECG Signal Enhancement Using Wavelet Transform", WSEAS Trans. On Biology and Biomedicine, issue.2, Vol.7, August 2010, pp. 210-219.
6. G. Selvakumar, K. Bhoopathy Bagan, B. Chidambararajan, "Wavelet Decomposition foe Detection and Classification of Critical ECG Arrhytmias", Proc. Of the 8th WSEAS International Conference on Mathematics and Computers in Biologo and Chemistry, Vancouver, Canada, June 2007, pp. 80-84
7. P. Ivanov, M. QDY, R. Bartsch, et al, (2009) "Levels of complexity in scaleinvariant neural signals",

   Physical Review
8. S. Z. Mahmoodabadi, A. Ahmadian, M. Abolhasani, P. Babyn and J. Alirezaie, (2010) "A fast expert system for electrocardiogram arrhythmia detection", Expert system, vol.27, pp. 180-200.
9. Ghaffari A, SadAbadi H, Ghasemi M. A Mathematical algorithm for ECG Signals Denoising Using Window Analysis. Biomed Papers 2006; 151: 73-78.
10. WAN Xiangkui, QIN Shuren, LIANG Xiaorong, YE Shunliu. The Application of Wavelet Transform in ECG Feature Extraction [J]. Beijing biomedical engineering.2005, 24(6):410-413.
11. Dan E. Kristiansen, John H. Husery, and Trygve Effestsl,"Rhythm detection in ECG signals", NORSIG-95, pp.I19-124,1995
12. Thakor NV, Natarajan A, Tomaselli G. Multiway sequential hypothesis testing for tachyarrhythmia discrimination. IEEE Trans Biom Eng 1994;41:480-487.
13. Chen SW, Clarkson PM, Fan Q. A robust sequential detection algorithm for cardiac arrhythmia classification. IEEE Trans Biom Eng 1996;43:1120-1125.

14. Clayton RH, Murray A, Campbell RWF. Comparison of four techniques for recognition of ventricular fibrillation of the surface ECG. Med Biol Eng Comp 1993;31:111-117.
15. P. Addison. Wavelet transforms and the ECG: a review. Physiological measurement, 26(5):155, 2005.
16. Petrutiu S, Nijm GM, Angari HA, Swiryn S, Sahakian AV. Atrial fibrillation and waveform characterization - A time domain perspective in the surface ECG. IEEE Engineering in Medicine and Biology Magazine 2006;25(6):24–30.
17. Sun Y, Chan K, Krishnan SM. ECG signal conditioning by morphological filtering. Comput Biol Med 2002; 32(6):465–479
18. J.M. Mendel, "Tutorial on Higher-Order statistics (spectra) in signal processing and system theory : Theoretical results and some applications" proc. IEEE. Vol 79, pp.278-305,
19. Ge I., Dai J: A oatholoeical mcdel studv for hi&-frequcnc) electnrardio&!& uilh curunq ;nensl 1ig.iture.J. Zhejisng L (Sdlure Science1.31 (11.81-92, 1997)
20. Anishchenko T, Igosheva N, Yakusheva T, Glushkovskaya-Semyachkina O, Khokhlova O.Normalized entropy applied to the analysis of interindividual and gender-related differences in the cardiovascular effects of stress. Eur J Appl Physiol 2001;85:287-298
21. D Gao, M Madden, Arrhythmia Identification from ECG Signals with a Neural Network Classifier Based on a Bayesian Framework, 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, 2004
22. Brown G.,2006, MIT-BIH Arrhythmia database,MIT
23. Chickh M. A.N. Belgacem, F. Bereksi-Reguig ,2002,Neural classifier to classify ectopic beats. Acte des IX emes rencontre de la Societe Francophone de Classification, Toulouse, le 16-18.
24. F Belhachat, N Izoboudjen, Application of a probabilistic neural network for classification of cardiac arrhythmias, 13th International Research/Expert Conference" Trends in the development of machinery and associated technology", Tunisia, 2009
25. V. Kecman, Learning and Soft Computing Support vector machines Neural Networks-Fuzzy Logic, MIT, 2001
26. HA Guvenir, B Acar, Feature Selection using a Genetic Algorithm for the detection of abnormal ECG recordings, International Conference on Machine Learning,Florida, 2001.
27. M. E. Rahman and M. A. Haque, "Filter Based Enhancement of QRS Complex of Electrocardiogram", Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, 2003
28. Demuth Horward, Beale Mark, Hagan Martin, 2008,MATLAB Neural Network Toolbox, MATHWORKS INC., MATLAB Version R2008b, October

29. Rosaria Silipo and Carlo Marchesi, Artificial Neural Networks for Automatic ECG Analysis, IEEE Transactions Signal Processing, Vol. 46, Issue 5, May 1998, pp. 1417-1425
30. Gao D, Kinouchi Y, Ito K, Zhao X, 2003,Time series identification and modeling with neural network. In. Pro. IEEE Int. Joint conference on Neural Networks, Portland, USA, 2454-59.
31. M. P. Tjoa, D. Narayana Dutt, Y. T. Lim, B. W. Yau, R. C. Kugean, S. M. Krishnan and K. L. Chan, "Artificial neural networks for the classification of cardiac patient states using ECG and blood pressure data," The 7th Austria and New Zealand Intelligence Information Systems Conference, Perth, Western Australia, pp. 323-327, 2001.
32. K. I. Minami, H. Nakajima, and T. Toyoshima. Real-time discrimination of ventricular tachycardia with Fourier-transform neural network. IEEE Trans. Biomed. Eng. 1999; 46: 179–185.
33. R. Silipo and C. Marchesi, Artificail Neural Networks for Automatic ECG Analysis. IEEE Trans. Signal Processing 1998; 46 (5): 1417–1425
34. S. R. Safavin and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," IEEE Trans. Syst, Man, Cybern., vol. 21, pp. 660-674, 1991.
35. M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo. Clustering ECG complexes using hermite functions and self-organizing maps. IEEE Trans. Biomed. Eng. 2000; 47: 838–848.
36. le Cessie, S. and van Houwelingen, J.C., "Ridge Estimators in Logistic Regression," Applied Statistics, vol. 41, No. 1, pp. 191-201, 1992
37. W. Klosgen and J. M. Zytkow, Handbook of Data Mining and Knoweledge Discovery, Oxford University Press, 2002.
38. Y. H. Hu, S. Palreddy, and W. J. Tompkins. A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. Biomed. Eng. 1997; 44: 891–900.
39. A. D. Coast, R. M. Stern, G. G. Cano, and S. A. Briller. An approach to cardiac arrhythmia analysis using hidden Markov models. IEEE Trans. Biomed. Eng. 1990; 37: 826 - 835.
40. Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," IEEE Trans. Biomed. Eng., vol. 44, pp. 891-900, 1997.