"Diagnosis of Arrhythmia Using Neural Networks"

By

SELEPE, KGAUGELO ZACHARIA

FINAL PROJECT REPORT

Submitted to the Electrical & Electronics Engineering Programme
in Partial Fulfillment of the Requirements
for the Degree
Bachelor of Engineering (Hons)
(Electrical & Electronics Engineering)

Universiti Teknologi Petronas
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

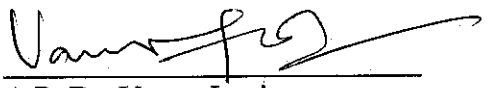
"Diagnosis of arrhythmia using neural networks"

by

Selepe, Kgaugelo Zacharia

A project dissertation submitted to the
Electrical & Electronics Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
Bachelor of Engineering (Hons)
(Electrical & Electronics Engineering)

Approved:



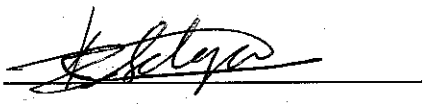
A.P. Dr. Varun Jeoti
Project Supervisor

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK

June 2006

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.



Kgaugelo Zacharia Selepe

Acknowledgements

Firstly I would like thank both my parents (Mr. and Mrs. Selepe) for their undying love and support. I would like to also extend my supervisor Dr. Varun Jeoti for his wisdom, insight and guidance towards completion of this project. To all my brothers and my sister for their support through thick and thin I thank God everyday for blessing me with loving siblings like yourselves. I would like to also acknowledge my uncle Mr. Simon Mpati who has played and important in every part in my personal life and my studies.

I would like to also acknowledge all the people who work tirelessly to contribute to www.physionet.com for availing such invaluable research tools.

Abstract

This dissertation presents an intelligent framework for classification of heart arrhythmias. It is a framework of cascaded discrete wavelet transform and the Fourier transform as preprocessing stages for the neural network. This work exploits the information about heart activity contained in the ECG signal; the power of the wavelet and Fourier transforms in characterizing the signal and the power learning power of neural networks.

Firstly, the ECG signals are four-level discrete wavelet decomposed using a filter-bank and mother wavelet 'db2'. Then all the detailed coefficients were discarded, while retaining only the approximation coefficients at the fourth level. The retained approximation coefficients are Fourier transformed using a 16-point FFT. The FFT is symmetrical, therefore only the first 8-points are sufficient to characterize the spectrum. The last 8-points resulting from the FFT are discarded during feature selection.

The 8-point feature vector is then used to train a feedforward neural network with one hidden layer of 20-units and three outputs. The neural network is trained by using the Scaled Conjugate Gradient Backpropagation algorithm (SCG). This was implemented in a MATLAB environment using the MATLAB GUI Neural network toolbox..

This approach yields an accuracy of 94.66% over three arrhythmia classes, namely the Ventricular Flutter (VFL), the Ventricular Tachycardia (VT) and the Supraventricular Tachyarrhythmia (SVTA). We conclude that for the amount of information retained and the number features used the performance is fairly competitive.

Table of Contents

Chapter 3:

Chapter 4:

Chapter 5:

Appendices

List of Illustrations

Figures

Tables

List of Abbreviations

CWT	: Continuous Wavelet transforms
DFT	: Discrete Fourier Transform
DWT	: Discrete Wavelet Transform
ECG	: Electrocardiograph
FFT	: Fast Fourier transforms
MATLAB	: Matrix Laboratory (Software Package for computation)
MIT-BIH	: Massachusetts Institute of technology- Bertha Israel Hospital
MLP	: Multilayer Perceptron
NN	: Neural Network
SCG	: Scaled Conjugate gradient back-propagation
STFT	: Short time Fourier transforms.
SVTA	: Supraventricular tachyarrhythmia
VFL	: Ventricular flutter
VT	: Ventricular tachycardia
WTO	: World health Organization

Chapter 1

Introduction: Socio-Economic Background, Classification of ECG with Neural Networks, literature Review, Organization of the Report.

1.1. Introduction

1.1.1. Socio-Economic Background

According to the WHO 17million people die annually from heart related conditions. This figure is expected to reach about 20 million by 2030. In the United States alone, 514 540 people die annually from heart related conditions. During the period 1989 to 1997 in USA between 9.4% and 11% of the health expenditure was spent on cardiovascular medications.

The Canadian institute of health(CIHR) reports that heart related conditions are the leading cause of death accounting for 33% of the deaths which amounts to 80 000 people annually. In the year 2002-03 alone, more than 411 000 Canadians were hospitalized for heart related conditions. It further reported that 1998 heart related condition amounted to $18.8 Billion expenditure, which is 11.8% of the total health budget. This cost statistic is fairly similar to the USA as reported by the WHO.

A number of lives are also lost annually due to discharges wherein a patient is in fact suffering from a heart related condition, however they are released because the condition was missed during diagnosis. In emergency cases where a patient is admitted into a hospital under a heart related attack, it is imperative that a correct diagnosis is made so that correct treatment can be administered, example whether a shock therapy in form a defibrillator should be administered. In these circumstances if the condition is misdiagnosed wrong treatment might be administered thus leading other to complications and death.

Therefore, given these overwhelming statistics, the significance of intelligent and highly accurate arrhythmia classification systems cannot be underestimated. This is mainly because the arrhythmias are conditions in themselves and they also present some other underlying heart conditions. For more on arrhythmia the reader is referred to (M.B. Conover, 2002).

In this report we present an intelligent arrhythmia classification system that combines the powers of the wavelet transform, the Fourier transform and neural networks. The approach is the tested over three arrhythmia classes namely SVTA, VFL and VT from the MIT-BIH database. The classes in this database are known with 100% accuracy. This approach attains an average accuracy rate of 94.66% over three classes. Furthermore, this work presents the first to approach this problem using Wavelet-DFT cascade.

1.1.2. Arrhythmia Classification with neural networks

Diagnosis of the above mentioned arrhythmia is performed by a neural network. The neural network has been used before in classification of ECG signals (heart conditions) [2], [3] ,[4], [5], [6]. In section 1.3 some of these approaches are reviewed briefly.

The time domain presentation implies that every sampled point on the signal is used as a feature to training neural networks. The weakness of the time domain presentation is that there are just too many points. An example are the signals that we use in this work, they were sampled at 250 Hz. If one used the time for classification in this case whereby we are using 1-beat (approximately 0.8s) for classification this implies that the neural network should be trained on 200 input points. This is computationally demanding during training, thus leads to prolonged training times. Moreover, the training over a large number of weights requires more examples for the neural network to generalize effectively (C.G. Looney, 1997).

Performance of some algorithms deteriorates with an increase in number weights (Demuth & Beale, 2001), an input space in itself introduces 200 weights, this might cause trouble when trying to find the optimal training algorithm. (Haykin, 1999) reports that too

many inputs to the neural network, can affect the stability of the neural network. However, he states that these can be countered by using feedback as in the recurrent networks. The problem with feedback connections is they render the use of mature training algorithms such as backpropagation less useful. Therefore is best to find better preprocessing techniques that can be used to reduce the data size while maintaining discriminatory features.

These weaknesses that are associated with the time domain presentation propel the need for other techniques to extract statistical characteristics that are inherent in the ECG signal. The wavelet and Fourier transforms presents an alternative to time domain raw signal. The power of the wavelet transform lies in the fact that it can capture the frequency changes with respect time in non-stationary signals such as the ECG. The time frequency presentation is captured as approximation coefficients and detailed coefficients. The coefficients can then be further processed to be suitable for neural network application.

1.1.3. Literature Review

The application of neural networks to classification ECG signals is not a new phenomenon. In fact it is an ongoing area of research with a number of papers published every year. In this section some of the important work in this area is summarized in order to put our approach into perspective.

(Guler and Ugbeyli, 2005) used a so called combined neural network model, of Multi layer perceptron (MLP) to classify ECG beats. The ECG signals were first decomposed in to coefficients by wavelet transformation using different wavelet functions. Another important observation from this work is that the 'db2' mother wavelet outperforms other wavelets in ECG classification.

They claim an accuracy of 96.94% over four classes. However, this approach is obviously requires extensive training times i.e. the effort required to train two neural networks.

In **[3]** a method to classify cardiac arrhythmias using a combination of wavelets and artificial neural networks is proposed. Their feature selection was based on the DWT coefficients that provide maximum information about the signal together with the time difference between the current R-peak and the previous R-peak.

These features were used to train a Multilayer Perceptron (MLP) neural network. They went a step further to test the robustness of their system by deliberately adding muscle noise and power line noise with SNR of 0 to 10 dB. Their selected wavelet for this application was the 'sym6' mother wavelet with a four level decomposition. In the selection of features is based on the frequency of maximal energy content which is 0.5Hz to 40Hz for the ECG signal. Any coefficients representing frequencies outside of this defined spectrum were discarded.

The reported scheme claims a classification accuracy of 96.77% over 13 classes. Moreover, these results were reported to be better than any documented results at the time of their publication (2003).

This approach gives a 25-input vector, which requires more complex neural network design and longer training times.

[4] Compared wavelet based and Fourier transforms based methods in combined with neural networks in classification of ECG beats. In their c work they point out that using wavelet coefficients creates a high dimensional feature vector. A restricted coulomb energy neural network trained with genetic algorithms was then used to classify the feature vectors. They report the accuracy of 89.4% for the Fourier transform preprocessed approach and a 97% accuracy for the wavelet transform on set of 10 classes.

In this work the performance deterioration of Fourier transform is compensated by a wavelet transform preprocessing. The high dimension of the wavelet transform is countered by taking only the fourth level approximation coefficients. These two compensations combined to exploit the powers of both techniques.

In their work **[10]** compared classification performance of the network for data generated by using 'db1', 'db4' and 'db10' Deubechies mother wavelets and varying decomposition levels between 3 and 7. They report average results between 71.8% and 74%. They then claim in conclusion that the number of levels and the decomposing mother wavelet do not affect classification accuracy.

In this work we report a four level multi-resolution decomposition (a practical DWT implementation, using the 'db2' mother wavelet) in cascaded with Fourier transform as preprocessing stages. A feedfoward neural network, with one hidden layer consisting of 20-units trained with scaled conjugate gradient backpropagation was used to classify the arrhythmias. The neural network is trained with 90 examples i.e. 30 from each class and it is tested with 75 examples.

1.1.4. Organization of the Report

The rest of this report is organized as follows: In chapter 2 we provide an in depth look at the theory behind ECG signals, pattern recognition and neural networks section 2.1, theory of preprocessing techniques section 2.2 (Fourier transform and wavelet transform). In chapter 3 a selected approached is presented and the choice of design parameters is discussed. In chapter 4 the results are presented and discussed. Finally, in chapter 5 the conclusions are presented and possible directions for future work are discussed.

Chapter 2

Theory: ECG and Arrhythmia, Pattern recognition, Neural networks, Fourier Transform and Wavelet Transform

2.1. ECG signals and arrhythmia

2.1.1. Introduction

The heart is divided in to four compartments namely the left atrium, left ventricle, right atrium, and right ventricle. The atria are the upper compartments and the ventricles are the lower compartments (see figure 1a). These compartments contract systematically in order to circulate blood through the body.

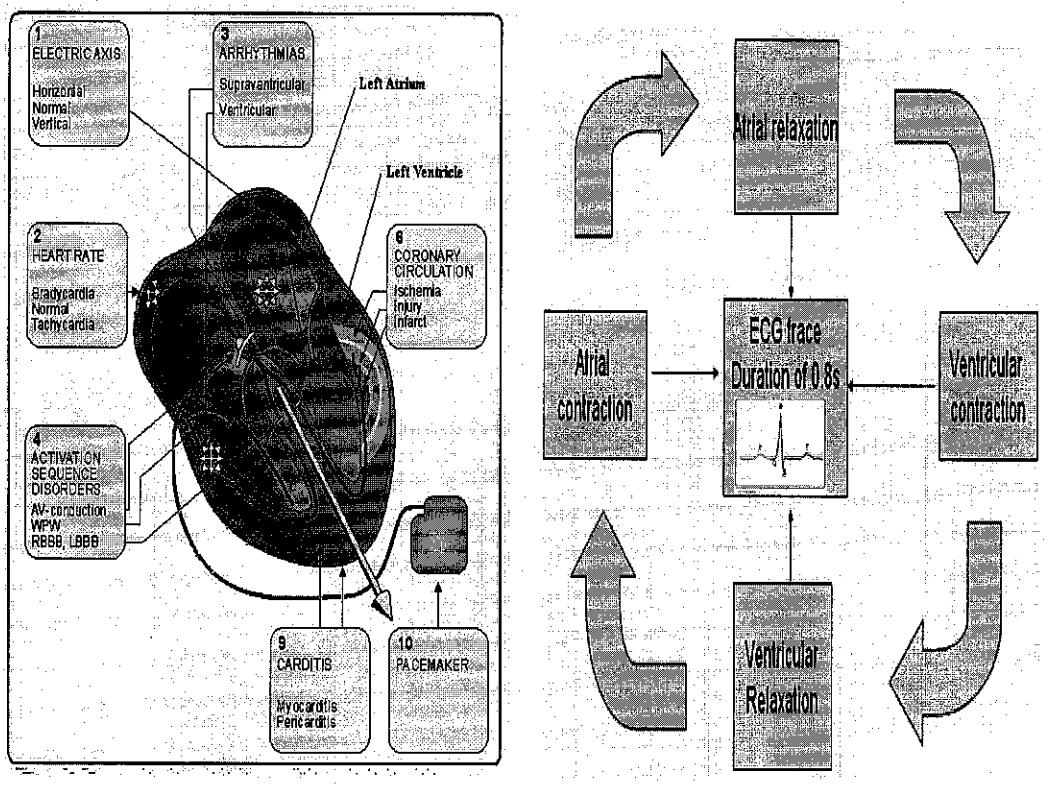
The electrocardiograph (ECG) signal is a model used to represent electrical activities of the heart. This modeling is a direct result of the fact the heart muscles are paced by means of electrical signals. The ECG is a voltage signal captured by placement of leads around the heart area. These leads are composed of two electrodes of opposite polarity. The signal capturing is attained by means of 12-leads of which 3 are bipolar and 9 are unipolar. Three bipolar and three unipolar leads are on the limbs. The remaining leads are in the chest area **[1]**.

The cardiac cycle consists of three phases the depolarization, polarization and resting. Depolarization is a rapid/slow transition of a resting cell to a more positive state **[1],** this can simply be referred to as contraction. Repolarization is transition back to the resting state (relaxation) (see figure 1b). The leads mentioned earlier capture atrial and ventricular polarization/depolarization **[1]** and this is traced as the ECG signal . The depolarization of the atria produces the P-wave on the ECG display; repolarization of the ventricles results in inscription of the QRS-wave.

The QRS is easily detectable as it has slightly higher amplitude than the P or the T waves (see figure 2a). In this application the QRS complex will be used as the detection point. In other words the QRS will be detected, because of its high amplitude and the window of approximately 0.8s, centered at the R-peak will be used to extract one heart beat.

Arrhythmias are indicated by an interruption in the cycle as indicated in figure 1b. The interruption is usually signaled by an increased or decreased duration of the cycle; the

absence of some events in the cycle and the repetition of some of the events in the cycle etc.



(a) Cross-Section through the heart (b) The heart cycle

Figure 1: Structure of the heart and the heart cycle

2.1.2 Arrhythmias

(a) Introduction

Heart rhythms may be classified as either ventricular or supraventricular, where supraventricular rhythms are those that occur above the ventricles (see figure 1a). The figure 1 above shows origin of some of the beats. Arrhythmias are the abnormal beats of the heart resulting in some irregular but specific ECG signal. These arrhythmias are usually an indication of more fundamental and potentially fatal heart conditions such as the myocardial infarcation **[1].**

The three arrhythmia classes were chosen for the purpose of this study, namely Ventricular flutter(VFL), Ventricular tachycardia(VT) and the Supraventricular tachyarrhythmia(SVTA). The real time traces for each one of these conditions is shown in figures 2(b, c, d respectively). These are real time 0.8s traces from the physionet ECG database.(www.physionet.org). The signals were originally sampled at 250 Hz and digitized for storage on digital computers. The 0.8s is approximately the duration of one heart beat. However, one notices that in these anomalies there is more than one peaks and all of the peaks are slightly deformed when compared to the one in figure 2a(see ECG trace in figure 2a). Moreover, one notices that there is more that one peak in the arrhythmias for the same duration where the normal ECG has only one peak.

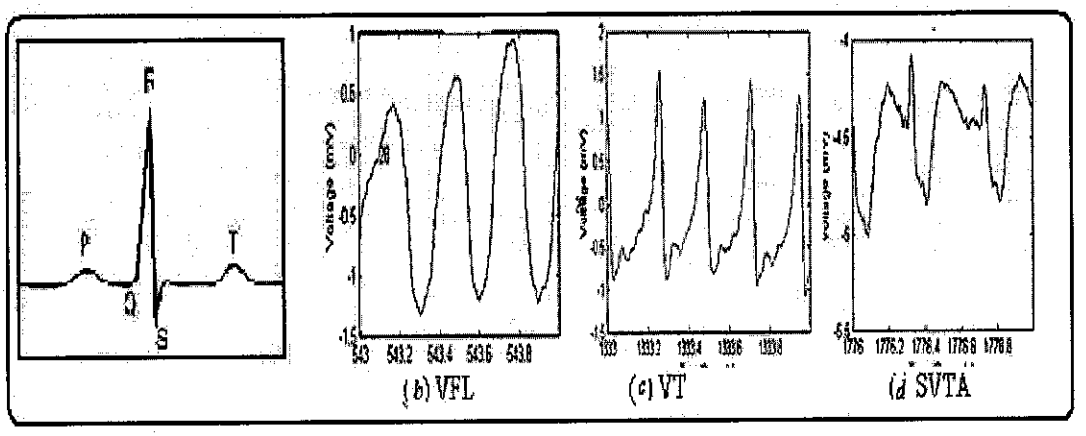


Figure 2: Normal Cardiac cycle and Traces of abnormal ECG

(b)Ventricular Tachycardia (VT)

VT refers to three or more extra-systoles or contraction of the ventricles. These could be a sign of underlying myocardial infarcation disease. If it sustains for more than 30 beats it is likely to translate into ventricular fibrillation which could cause death. Therefore the need for an accurate classifier to detect this abnormality cannot be overstated. The **figure** below shows the actual ECG trace of the VT as well as the origin of this arrhythmia. This trace has been diagnosed by an expert cardiologist. The DWT of a segmented portion has been taken using the 'db6' wavelet five levels there of are shown in figure 3 (a-e)

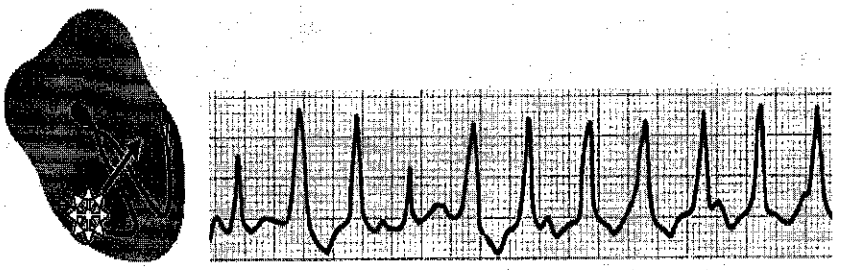


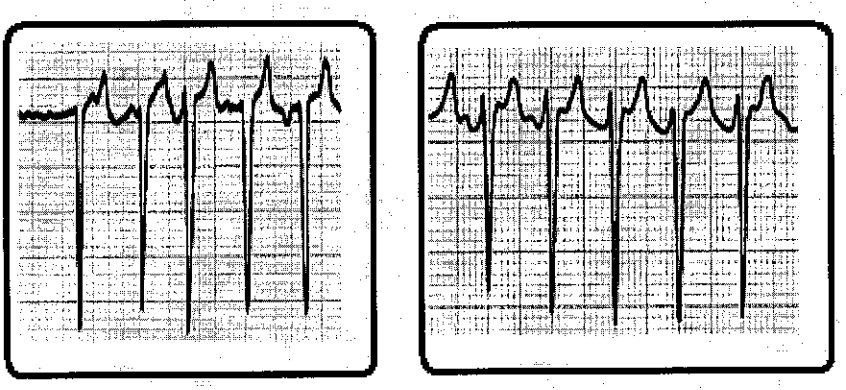
Figure 3 : Ventricular tachycardia origin and ECG signal trace

(b) Supraventricular Tachyarrhythmia

Supraventricular Tachyarrhythmia is commonly caused by arrhythmic contraction of the atrial muscle. This is results from atrial rates in between 450 to 600 per minute. Some of the causes of a irregular are **[1]:**

- Frequent atrial extrasystoles;
- Atrial fibrillation.
- Multi-focal atrial tachycardia, where there are three or more distinct atrial foci, combined with tachycardia. There is often severe underlying disease (e.g. chronic obstructive airways disease), and in the ICU setting, MAT has a poor prognosis.
- "Atrial flutter with variable block".

Figure 7 below outlines the ECG recording of both regular and irregular. However, this work only focuses on detecting supraventricular arrhythmia and not whether they are regular or irregular. Figure 8(a-d) depicts a four level DWT decomposition of this signal.



(a) Irregular SVTA (b) regular SVTA

Figure 4: ECG recording of regular and irregular SVTA

The *Ventricular Flutter* is reported to be the rarest form of ventricular arrhythmia. However, it's dangers are that it can very easily translate in to the more fatal ventricular fibrillation. It is characterized by a sinusoidal like ECG wave form as indicated in figure5.

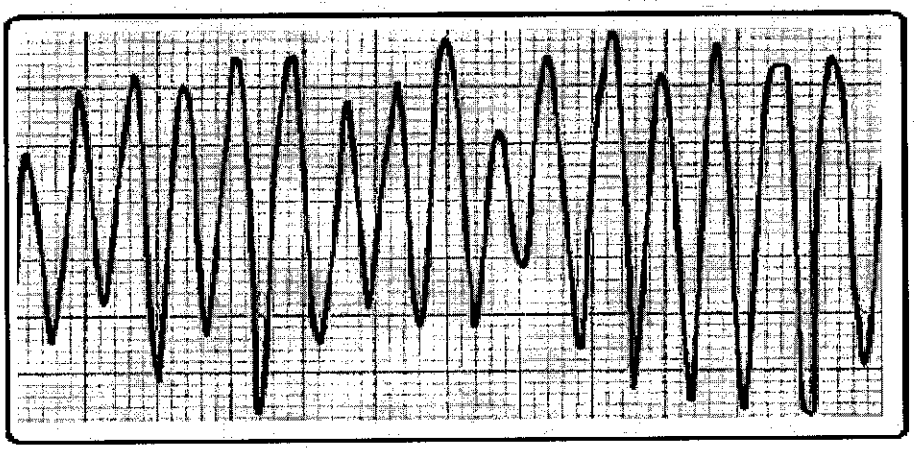


Figure 5: Ventricular flutter waveform.

(e) Synopsis

In this work the ECG traces, which represent underlying activities of the heart, are captured in duration 0.8s around the QRS complex. The captured ECG signals are the preprocessed and fed to the neural network for classification

2.2. Pattern recognition and neural networks.

2.2.1. Pattern Recognition theory

Pattern can be defined -amongst others- as a regular form or order in which a series of events occurs. Recognition refers identification of objects as belonging to a particular class, in other words it is the decoding/interpretation of presented patterns. Learning is the process by which identified patterns are assimilated as knowledge for future use or reference. Hence, pattern recognition can be loosely defined as identification of an approximate series or order in which events occur and assignment of events sharing the same order into a common class. Alternatively it is the class assignment of objects based on the pattern that they present to the sensors. In computational intelligence process of pattern recognition can be broken down as follows:

A series of events or an object (i.e. a pattern) is mathematically modeled as a suitable vector. This vector, which is effectively a mathematical representation of the pattern, is then presented to a classifier system. However, the system is firstly trained/taught using some of these vectors. The training phase is a process whereby the classifier finds decision boundaries that separate these vectors.

The classifier finds a mapping between the presented vector and the associated patterns and "constructs" a decision boundary that separates the respective classes according to

some inherent statistical properties. This fitting process is analogous to polynomial curve fitting. In computational intelligence fitting methods have been designed that allow non-linear fitting.

The performance of the classifier is based on its abilities to correctly assign a given unseen vector, whose class in known vector with certainty, to correct class of other vectors sharing the same attributes. The pattern is preprocessed into a representation vector so usage by the classifier. Therefore the challenge in designing these sorts of systems is finding the mathematical representation (vector) that highly discriminates the classes of interest i.e. that provides intimate knowledge about the pattern. The other challenge is to select/design a classification system that is computationally cheap, that converges for a large class of patterns and that will give the least classification error.

The process is illustrated with the use of an example to classify shapes in the (figure 6). Firstly, the shape of the presented object is modeled as a mathematical vector that most discriminate it. For instance in this case the most discriminating features cannot be the lengths of the sides because the square and the triangle have sides of equal lengths. Therefore the possibilities would be a column vector of number of sides as well as the lengths of the sides vertically concatenated.

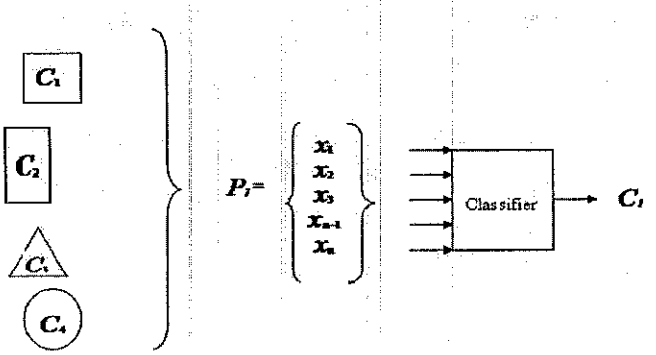


Figure 6: Illustration of pattern recognition

The mapping is $P_i \xrightarrow{classifier} C_i$ where the classifier is effectively the mapping function between the domain P_i and the range C_i. The vector space P which is a set of vectors (containing the defined features) is divided into the training and testing set.

During training mode, the training set is fed to the classifier and its corresponding class assigned accordingly. Once the training goals have been attained (which can is the reduced root means square of the error i.e. the value difference between the class output and the desired output C_i, or some other parameter) the classifier is ready for testing and subsequently the operational mode. During the testing mode, testing set is fed to the classifier to test its classifying abilities. The result is presented as what is called a confusion matrix also used in (Guler and Ubeyli, 2005). With reference to the example below the matrix will look as in figure 7.

The performance is evaluated as percentage of the total test set where, class one correctly classified as class one, incorrect classification of class one as class 2 ect. The circled classes are those that have been correctly assigned. The confusion matrix shows the presence of correlation between any numbers of classes. The results as shown on the confusion will show the classification ability as well as its sensitivity of the classifier. The availability of the information about which two classes are correlated, can enable the designer to use other classification techniques such hierarchical classification system, to account for similar classes. In the hierarchical approach the designer could use two preprocessing techniques to account for similar classes. In other words the second technique could be a domain wherein the similar classes are discernible. See appendix A for a list of pattern recognition methods.

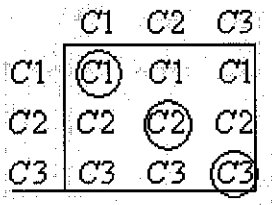


Figure 7: Confusion Matrix

2.2.2. Neural Networks theory

(a) Introduction

The literature review above singles out the neural network as the most popular classification tool. This notion has been adopted in this project, since the project aims to improve the performance of reported/existing systems.

The neural networks are a result of many years of attempting to mathematically model the functioning of the neural system of a brain. The first reported work in this direction is by (McCulloch and Pitts, 1943) according to **[11].** This model was based on the primitive knowledge that the neuron can either be excited or inhibited depending on whether the strength of its input is above or below some threshold.

This model was rendered weak as it lacked the ability to learn in other ways than setting a threshold. (D.Hebb, 1949) as reported by **[11]**. (D.Hebb, 1949) Made a deduction that "If a Neuron A is repeatedly stimulated by neuron B at times when neuron A is active, neuron A will become more and more sensitive to neuron B " This laid foundation for mathematical formulation of adjustable synaptic weight. In this model the input vector as defined above is mapped to the output vector as a summation of the inner product of the synaptic weights (**W**) and the inputs (**I**) equation (1). These weights can now be adjusted such that a desired threshold level is attained.

$$Y = \sum WI \quad \text{..... (1)}$$

In the modern neural network models the threshold functions were replaced with the continuously differentiable functions, this allowed resolution of the weight space/matrix. These were introduced in order to allow usage of gradient methods to find optimal set weight nodes. The sigmoid function (see figure 8) is one of the popular activation functions: with its corresponding equation (2), hence with a saturation value of 1. Its popularity is a direct result of its differentiability at all points Moreover, it permits non-linear mapping that it permits.

$$f(x) = \frac{1}{1 + e^{(-\alpha(x-b))}} \quad \text{.....(2)}$$

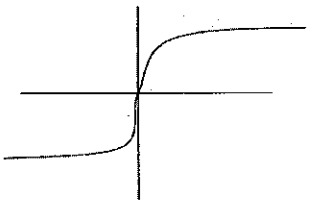


Figure 8: Sigmoid squashing function

(b) The Perceptron

The notion of perceptron was given to a single processing unit (corresponding to a single neuron) with a number of weighted inputs a summation function and activation function as (in figure 9). These perceptrons can be cascaded and paralleled into a parallel mapping system called a multilayer perceptron (MLP) also known as the feedforward neural network depicted in the figure10 below as 3-layer networks with 3 units in each layer. The three layers are known as the input the hidden and the output layers. The above MLP operate as a classifier as explained later. This multiplelayer configuration enables a non-linear mapping as well as training. Training of this MLP is an iterative operation

whereby the weights are adjusted in order to reduce the sum of squared errors. The algorithm is as follows:

1. Weights are initialized
2. At each iteration adjust the weight that reduces the sum of squared errors, according to a chosen specific training algorithm.
3. If the error criterion is met then stop else step 2 above.

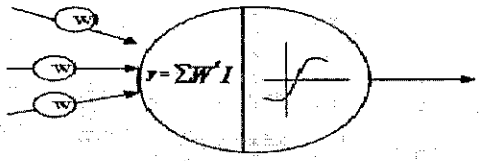


Figure 9: Perceptron

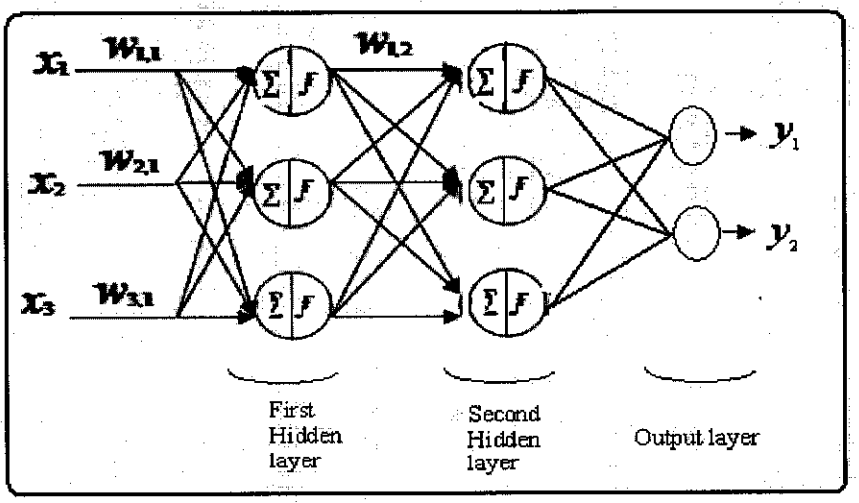


Figure 10: Multilayer perceptron

(c) Feature extraction and Pre-processing for neural networks

Feature extraction is a systematic technique of extracting sufficiently discriminatory features in a large feature set. It is necessary because at time there might be redundancies in the feature set. This involves another mapping of the pattern space into a reduced form. This process should retain the maximum possible original information; remove redundant information that might cripple classification performance an observation by **[11]** in (Devijver, 1982). It must be meaningful to humans. For instance in this case study, if one assumes an ECG signal of length 0.8s (Beat duration) seconds sampled at 250 Hz, this presents a sample every 0.004s. Hence the total number of samples will be 0.8/0.004 = 200 samples. It is therefore not practical for one to train the neural network with such a vast amount of data.

According to **[11]** (Therrien, 1989) defined an optimal feature-set as non-redundant and retaining maximal information about the signal. For this purpose **[11]** reports a method that has been proposed (Bezdek and Pal, 1992) that the first processing on data should involve scaling, normalization, smoothing and cleaning. This should be followed by selection of features where features that distinguish between two classes are selected and used directly while the remaining features are mapped into ratios, logarithms, means, deviation and similar mappings.

Therefore, techniques such as the Wavelet transform and Fourier transform are used to represent the signal, subsequently reducing the dataset while maintaining maximal discriminatory information.

(d) Training and validation of neural networks.

The neural network learns through training. There are two training learning modes, according to **[8]** namely supervised and unsupervised learning. In unsupervised learning the neural network is presented with just the inputs and it is then required to use some clustering algorithm to learn statistical patterns within the data. The unsupervised training mode will not be discussed further in this report as we have limited our training to supervised mode.

In supervised learning the neural network is presented with both the input data and the corresponding output data. The network has to therefore use some mapping algorithm that minimizes its actual output and desired output for a given input data. This difference is referred to as the error equation (5), where e_i is the error, t_i is the desired output and y_i is the actual output.

$$e_i = t_i - y_i \quad \text{.....(5)}$$

The error is then reduced according to some algorithm. The backpropagation (eg. which is discussed later) based algorithms are the well known and most mature of these algorithms.

The network learning is achieved by preparing a large dataset that is then broken down into training and testing set. At times an additional parameter called the cross-validation is used. This parameter is calculated on the validation set after each training epoch and compared with the selected base. The idea is that this parameter will improve i.e. approach the set base after each epoch. If this is not the case then, the training is terminated. This is summarized in (figure 11) adopted from **[11].**

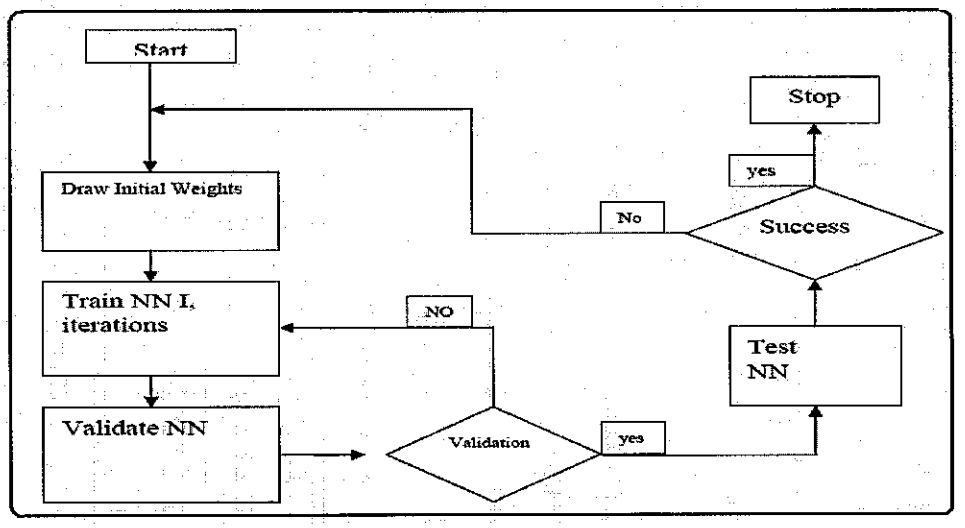


Figure 11: Neural network Training algorithm

(e) Architectural design issues.

Architecture of the network is the most important design parameter for neural networks in that all the other parameters depend heavily on it. The architecture refers to the type of neural networks which includes multilayer perceptron, the radial basis function, probabilistic neural network and the number of hidden layers (see C.G Looney 1997 for more on architectures). There is no theoretical foundation for choosing the type of network architecture. However, (Hornik et al 1989) reports the universal approximation theorem) which states that a neural network with one hidden layer containing a sufficient number of hidden units can approximate any function to any degree of accuracy. The validity of this theorem eases selection of network architecture. However, the problems that persist are selecting an accurate number of layers and correctly selecting the output units.

The two solutions to selecting an optimal number of hidden layers are known as network pruning (Haykin, 1999) and network growing. The network growing technique starts with a network of one hidden layer with a small number of units, as little as two hidden units. The network is then trained while its performance is closely monitored. At each step the number of units increased until the increase in number of units does not yield any

improvement in the performance of the neural network. Wile the network pruning technique is the opposite, whereby the one starts with a large network with more than adequate performance and at each step the number of units are reduced. In this work the network growing technique is used to determine the optimal number of hidden units. The network growing is as follows

- Firstly, the number of hidden units is selected.
- The weights of the network are then initialized.
- The network is fed with training data (Inputs and corresponding targets) and testing inputs.
- The testing inputs are used to evaluate the generalization ability of the neural network.
- The network training is started.
- The training and testing error are monitored, until they no longer change or they start increasing.
- The network training is then stopped.
- The number of hidden units is increased and training is started afresh.
- The testing error is once again tracked, when it reaches minimum step 3 is repeated.
- The procedure is continued until the validation error starts to increase.
- The optimal number of hidden units is the one wherein the validation error starts increasing.

The weaknesses in the single layer approach are outlined as:

- The single layer will require a large number of units for approximation.
- This creates a computational bottleneck, resulting in long training times.
- The units in the layer interact, therefore some units may affect the way in which other units respond to data.
- Some units may be biased towards certain features in the set.

- The result could be that some of the unit's response to data may become less accurate.
- The problem can be solved by addition of hidden layers; this allows the network to extract features in a hierarchical manner.

(f) Scaled Conjugate Backpropagation Algorithm.

There are a number of variants of the standard backpropagation algorithm. The variation between these algorithms is based on the design of optimization technique. It has already been mentioned briefly that backpropagation is by far the most mature of all training algorithms. It is used for supervised training mode. The notion of error during training was introduced in **(equation 5).** In this section the backpropagation algorithm and the scaled conjugate will be alluded to without any in-depth presentation mathematical.

(i) A standard backpropagation algorithm

Standard backpropagation is a gradient descent algorithm **[7].** This means that the network weights are adjusted along the negative of gradient of the performance function, simply said it is the direction in which the error e_i decreases the fastest. A single iteration of this algorithm is as in equation.

$$w_{i+1} = w_i - \alpha_i \partial_i, \quad \ldots\ldots(6)$$

where w_i is the current weight matrix.

α_i is the current learning rate, this is the step size change in weights.

∂_i is the current gradient.

This gradient descent algorithm can be implemented in two modes, namely batch training and incremental training. For more details the reader is referred to (C.G. Looney, 1997).

In the batch training mode the weight matrix w_i is updated once all the vectors have been presented, whereas in the incremental mode the weight matrix is updated after

presentation of each input vector to the network. However, the standard backpropagation and some of its variants are weak in that they have fairly poor convergence and their performance depends on parameters which there is no established criteria for choosing [namely learning rate and in other variants the momentum**13]**.

(ii) Scaled conjugate backpropagation

The conjugate gradient descent family is a variety of the basic gradient descent algorithm. The weakness of the standard gradient descent algorithm is that the direction of steepest descent does not result in faster convergence **[7],** said another way it is computationally uneconomical.. The conjugate gradient algorithms address this weakness by searching along conjugate directions which tend to produce faster convergence, rather than line search. These algorithms are discussed in **[15]**.

Scaled conjugate gradient descent, first introduced by **[14]**, belongs to this family of conjugate gradient backpropagation algorithms. Basically, the Scaled conjugate algorithm combines the power of the Levernberg-Marquardt algorithm **[16]** and the conjugate gradient. In **[7] the** workers benchmarked the training time performance of the scaled conjugate backpropagation against other algorithms on different sets of training data (see table 1, data is a summary from **[7].** The datasets were varied between pattern recognition and function approximation. The scaled conjugate seems to have an above average performance across all dataset.

which will effectively increase the number of weights. Therefore, it is important that the training algorithm performs consistently well with increasing weights.

2.3. Preprocessing Techniques

2.3.1. Wavelet Transform theory.

(a) Continuous wavelet transforms.

Although wavelets have been around their, applicability was first introduced in geophysics for analysis of seismic data (Boggess & Narcowich, 2001). The limitations of Fourier transform are that it gives an average of oscillations that comprise the signal, but it is unable to localize the oscillations of interest in time. As indicated section **2.2.1**, in the Fourier transform the signal is interrogated by sinusoids (basis functions) of different frequencies. However, the interrogation is over all space i.e. from $-\infty$-to- $+\infty$. In other words the signal is assumed to be stationary i.e. its frequency is not changing with time. This is however not true for all signals.

Short Time Fourier transform (STFT) was introduced to counter this weakness. In the STFT the signal is windowed using a fixed size window and the product of each signal and the window is Fourier transformed in order to localize the frequency in time (see figure 13). The weakness of the STFT is its fixed window size, therefore short time intervals of high frequency spikes are not detectable using this method **[18]**

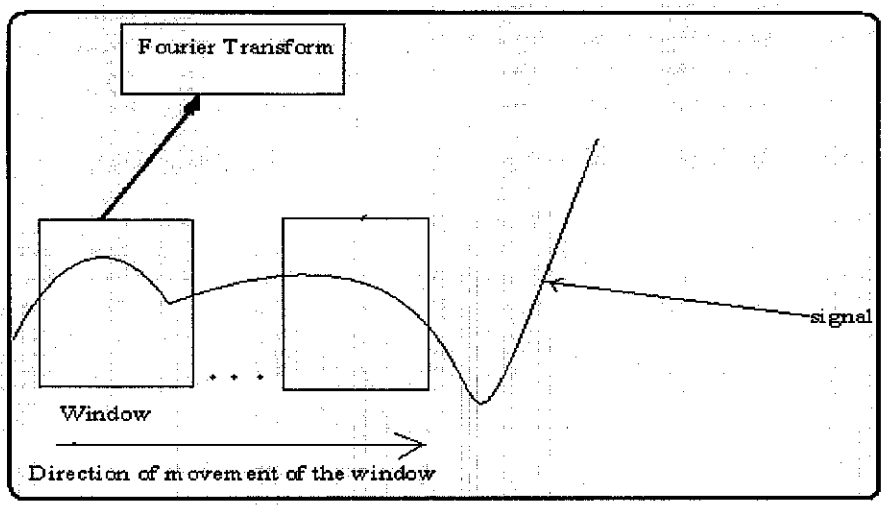


Figure 13: Demonstration of STFT

The wavelet transform provides an alternative flexible window STFT. The wavelet transform enable both the time and frequency localization. This power of the wavelets is a direct result of the flexibility of the basis functions (also known as wavelets). In the wavelet transform, the wavelet (also known as the mother wavelet) can be said to be both a flexible window and a basis function (see figure 14). In the figure it is clear that at each scale the wavelet is compressed.

In the figure below, at each scale, the wavelet is translated along the signal as shown in the figure. At each translation point the coefficients are calculated as the inner product of the wavelet and the signal using equation (13), where

$C_{\tau,s}$ is the coefficient; τ is the translation; s is the scale. The coefficient is actually a measure of how "matched" the wavelet and the signal are at that point.

$$C_{\tau,s} = \int x(t)\Psi(\tau,s).dt \quad \text{.....(13)}$$

The coefficients collected at the highest scale are known as the scaling function coefficients while the coefficients at subsequent scales are known as the wavelet coefficients. This is known as the continuous wavelet transforms (CWT), because every point on the signal is covered by the wavelet during translation. However, this transform

Table 1: Comparison of algorithms convergence time over different datasets.

	Average Convergence over the dataset(seconds)					
Algorithm	Sine function approximation	Parity	Engine	Cancer	Cholesterol	Diabetes
Scaled conjugate gradient	6.09	4.09	36.02	86.58	99.73	390.53
Levemberg Marquadt	1.146	13.07	18.45	110.33	261.50	1028.10
Fletcher-Powell conjugate gradient	7.86	6.62	37.93	110.05	136.04	784.50
Conjugate gradient with Powell/Beal restart	6.61	5.30	39.93	80.27	124.04	394.67
Polak-Ribiere Conjugate gradient	8.24	5.13	44.30	87.70	121.54	415.90
BFGD Quasi Newton	5.22	19.68	27.12	209.60	550.92	1821.00
Variable learning-rate back-prop.	27.69	27.07	188.50	313.22	3169.50	7687.00
Resilient back-prop	5.67	3.73	65.91	83.41	1519.00	323.90
Average	8.57	10.58	68.78	135.14	747.78	1605.70

In their conclusion [7] reported that the Levenberg-Marquardt algorithm converges the fastest on function approximation applications. They conclude further that its performance deteriorates with increasing number of weights.

They further reported that Resilient backpropagation, converges fairly fast on the pattern recognition problems. They also report that the performance of this algorithm deteriorates on function approximation and when the error goal is reduced.

The SCG is reported to be a fairly good performer across all types of problems,. This is clearly evident in table 1, where we it can be seen that the convergence of the SCG is consistently above average. Moreover, the performance of this algorithm does not deteriorate with increasing number of hidden units. This is useful in this application since the we are using the (Hornik, Stinchcombe and White ,1989) universal approximation theorem, whereby the network consists of only one hidden layer(see chapter 4). The single layer approach implies that the network might have a large number of hidden units,

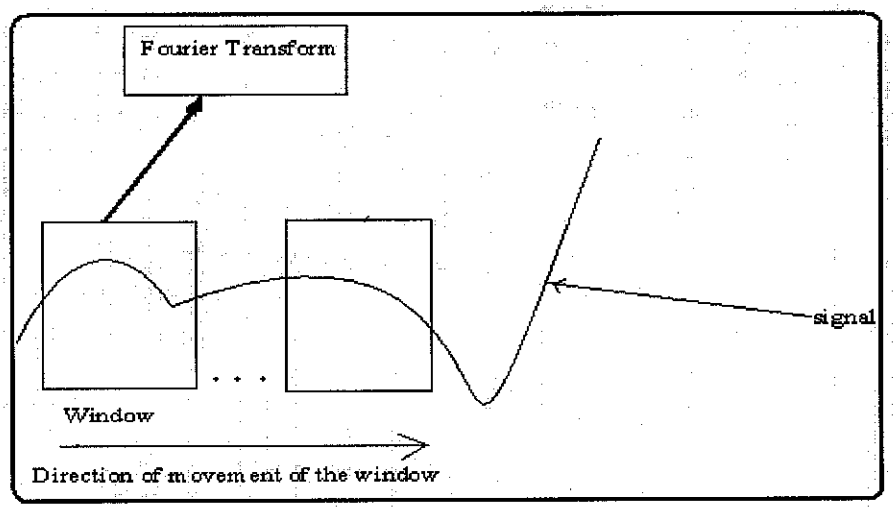


Figure 12: Demonstration of STFT

The wavelet transform provides an alternative flexible window STFT. The wavelet transform enable both the time and frequency localization. This power of the wavelets is a direct result of the flexibility of the basis functions (also known as wavelets). In the wavelet transform, the wavelet (also known as the mother wavelet) can be said to be both a flexible window and a basis function (see figure 13). In the figure it is clear that at each scale the wavelet is compressed.

In the figure below, at each scale, the wavelet is translated along the signal as shown in the figure. At each translation point the coefficients are calculated as the inner product of the wavelet and the signal using equation (13), where

$c_{\tau,s}$ is the coefficient; τ is the translation; s is the scale. The coefficient is actually a measure of how "matched" the wavelet and the signal are at that point.

$$C_{\tau,s} = \int x(t)\Psi(\tau,s).dt \quad \text{.....(13)}$$

The coefficients collected at the highest scale are known as the scaling function coefficients while the coefficients at subsequent scales are known as the wavelet coefficients. This is known as the continuous wavelet transforms (CWT), because every point on the signal is covered by the wavelet during translation. However, this transform

presents a lot of redundant coefficients that come as a result of overlapping translations. This problem can be solved by the discrete wavelet transform, where the translations are discretized to assume only certain values, usually on a dyadic grid **[19].**

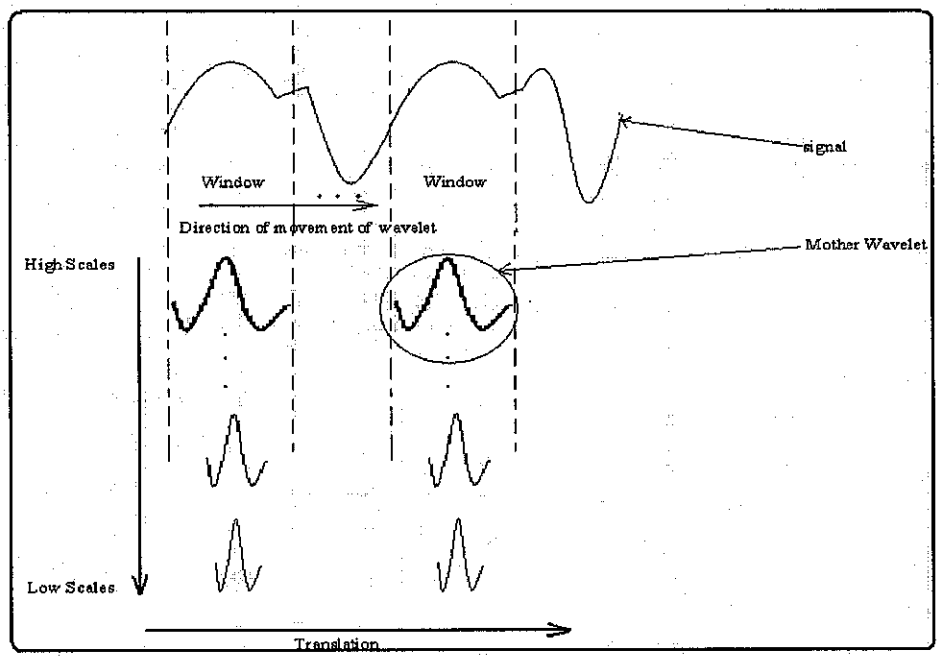


Figure 13: Continuous Wavelet transform demonstration

(b) Discrete wavelet transform (Multiresolution implementation).

The CWT has been shown to introduce redundant coefficients. This weakness is computationally uneconomical, because certain points along the translation are wavelet transformed twice and unnecessarily so. This weakness is countered by means of discrete wavelet transform **[19].** In DWT the scale and translation assume particular values by some predetermined criteria, usually a dyadic grid.

The most practical implementation of the DWT is the multiresolution i.e. multiple level decomposition. As a result in this section we will focus more on this practical implementation that the mathematics of the DWT, because this mathematics resembles that of CWT. In this decomposition the frequency spectrum of the signal is halved at each level. The coefficients of the upper half become what is referred to as the detail

coefficients and the lower half becomes the approximation coefficients. The lower half i.e. low frequencies usually contain important information about the signal; such is the human voice where the low frequency is still intelligible although its quality is highly deteriorated. The low frequency end of some of the arrhythmia in this study also contain critical diagnostic information as pointed out by **[5]** where they claim the useful frequency range for diagnosis is 0-to-40Hz.

Another perspective is that decomposition at each level inherently removes some of the high frequency noise content in the signal, which renders this decomposition as a denoising tool. The multiple stage decomposition is implemented by means of filter as in the figure 14, where g[n] is the low pass filter; h[n] is the high pass filter; D denotes the approximation coefficients and A denotes the detail coefficients. The down-sampling by a factor of 2 merely removes redundant information in the approximation coefficients.

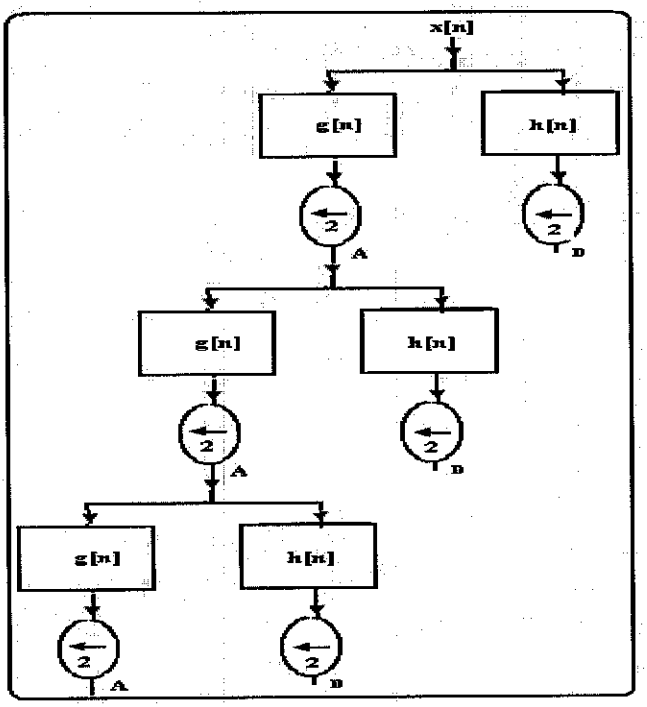


Figure 14: An illustration of multiresolution decomposition

(c)Choosing the number of decomposition levels and the mother wavelet

The choice of a suitable mother wavelet is critical in classification of ECG signals. (Guler and Ugbeyli, 2005), used a so called combined neural network model, of Multilayer perceptron (MLP) to classify ECG beats. The ECG's were first decomposed in to coefficients by wavelet transformation using different wavelet functions. They claim an accuracy of 96.94% over four classes.

Interestingly, their experiments there are also evidence that the 'db2' wavelet out-performs the other wavelet functions in approximating the ECG signal. Based on this evidence we chose to use the 'db2' and 'db6' and compare their effect on classification. And as anticipated the 'db2' outperformed the 'db6'

In their work **[10]** compared classification performance of the network for data generated by using 'db1', 'db4' and 'db10' Deubechies mother wavelets and varying decomposition levels between 3 and 7. They report average results between 71.8% and 74%. They then claim in conclusion that the number of levels and the decomposing mother wavelet do not affect classification accuracy. This conclusion was used in order to decide the number of levels. Based of popularity a four level decomposition was chosen.

2.3.2. Fourier Transform theory.

(a) Fourier series

Fourier transform based preprocessing is a popular technique for feature selection technique. (Dokur *et al*, 1999) in their work used Fourier transforms specifically a 40-point Discrete Fourier Transform (DFT), as a preprocessing technique. It has also been used in other neural network classification applications other than the ECG classification, as a preprocessing for neural network training.

The Fourier transform is a method of representing mathematical models of signals and systems in the frequency domain **[17].** In the notion of Fourier series a periodic signal, can be represented a sum of weighted basis signals such as sinusoids of different frequencies equation (8). When equation (8) is rewritten in its expanded form, it becomes clear that the Fourier series of a signal or a system is a sum of weighted sinusoids of different frequencies, where the weights are c_k and the respective frequenciesare $k\,\omega_0$.

$$T_0 = \frac{2\pi}{\omega_0} \qquad \text{.....(7)}$$

Equation (7) is the period of the signal. When the moduli of weights are plotted against their corresponding frequencies, the result is a spectral representation which might look like figures 15. Where $\omega = k\,\omega_0$.

$$f(x) = c_0 + \sum_{k=1}^{\infty} 2|c_k| \cos(k\,\omega_0 t) \qquad \text{..... (8)}$$

Where $c_k = \frac{1}{T_0}\int_{T_0} x(t) e^{-j\omega_0 t}\,dt$ ..... (9)

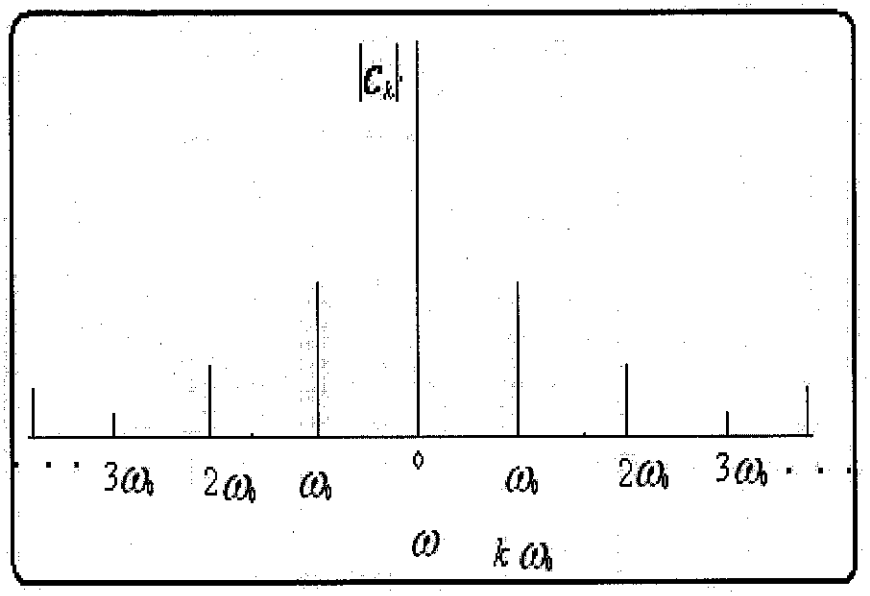


Figure 15: Frequency response (Spectral representation of a signal

(b) Continuous Fourier Transform.

The Fourier transform in an extension of the Fourier series. In order to generalize the Fourier transform the period of periodic signal is lengthened to infinity i.e. $T_0 \to \infty$

and since $\omega_0 = \frac{2\pi}{T_0}$, equation(9)translates to equation (10)

$$c_k = \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt\right] d\omega \quad \text{.....(10)}$$

where ω is a continuous frequancy variable.

When the moduli of weights are plotted against ω, a continuous variable, the result is a continuous frequency spectrum i.e. one where all the points are very close to each other.

(c) Discrete Fourier Transform

If the hypothetical signal whose Fourier transform is represented above was to be sampled using the Nyquist sampling theorem, a discrete time representation of the signal would result. If the discrete time representation is Discrete time Fourier transformed, the result would be a discrete time Fourier transform, this is a continuous function of

frequency **[17]**. However, if the discrete time Fourier transform is sampled at discrete points, this results in a Discrete Fourier Transform (DFT). This sampling of the frequency spectrum is useful in modern day applications wherein a digital computer is used, i.e. representation of continuous variables in a computer is not practical. The samples can be selected by first multiplying the discrete time signal with a windowing function in order to select certain samples to be Discrete Fourier transformed (equation 11).

$$\chi_N[n] = x[n]w[n] \quad \text{.....(11)}$$

Where:

$w[n]$ is the finite duration window function of length N.

$x[n]$ is the discrete time signal.

$\chi_N[n]$ is the windowed discrete time signal.

Therefore the signal remaining is N length in duration.

When $\chi_N[n]$ is Fourier transformed the result is $X(\Omega)$ which is a continuous frequency spectrum.

Thus the discrete frequency spectrum can be obtained by sampling $X(\Omega)$ as in the equation(12).

$$X_s[\Omega] = X_N[\Omega] \sum_{k=0}^{N-1} \delta(\Omega - 2\pi k/N) \quad \text{.....(12)}$$

Note: the value of the frequency spectrum at each sample is the frequency weighted impulse (Parr, Riskin, Charles, 2003).

The equation(12) can be implemented by means of faster algorithms of the Fast Fourier transform (FFT) family. In this work we use a 16-point FFT. For more on FFT the reader is referred to (Parr, Riskin, Charles, 2003).

Chapter 3

The Approach

3.1. Outline of the Approach

The experiments were implemented in MATLAB version 6.1.0.450 release 12.1(2001) on an AMD Duron processor 2.66GHz and 256 DDRAM machine. The neural network was implemented using the neural network toolbox. It is assumed that the beat has already been segmented. The signals used in this experiment were originally sampled at 250Hz. The beat duration in this application is assumed to be 0.8s. This translates to 200 samples.

The 200 samples are then 4-level wavelet transformed with the 'db6' mother wavelet by means of the multilevel decomposition as described in chapter 2. (P de Chazal and R.B. Reiley, 2000) have shown that the choice of a mother wavelet does not have any significant impact on the classification; hence the 'db6' was chosen because of its popularity. At each level the wavelet coefficients are discarded. Besides loss of some high frequency information, the discarding of detail coefficients in fact removes some of the high frequency noise.

Furthermore, **[5]** have shown that in fact the significant diagnostic information lies in the low frequency end i.e. 0-40Hz. The resulting fourth level decomposition is 13-approximation coefficients. This approach was chosen because the difficulty with using the Fourier transform of the original signal is that the signal samples are usually large. It is also a great deal of difficulty to choose the suitable features from a large number of Fourier transform coefficients of the original signal.

The data was then divided into training and testing data sets. The training set was made up of 90 vectors (30 from each class) and the testing set was made up of 75 vectors (25 from each class)

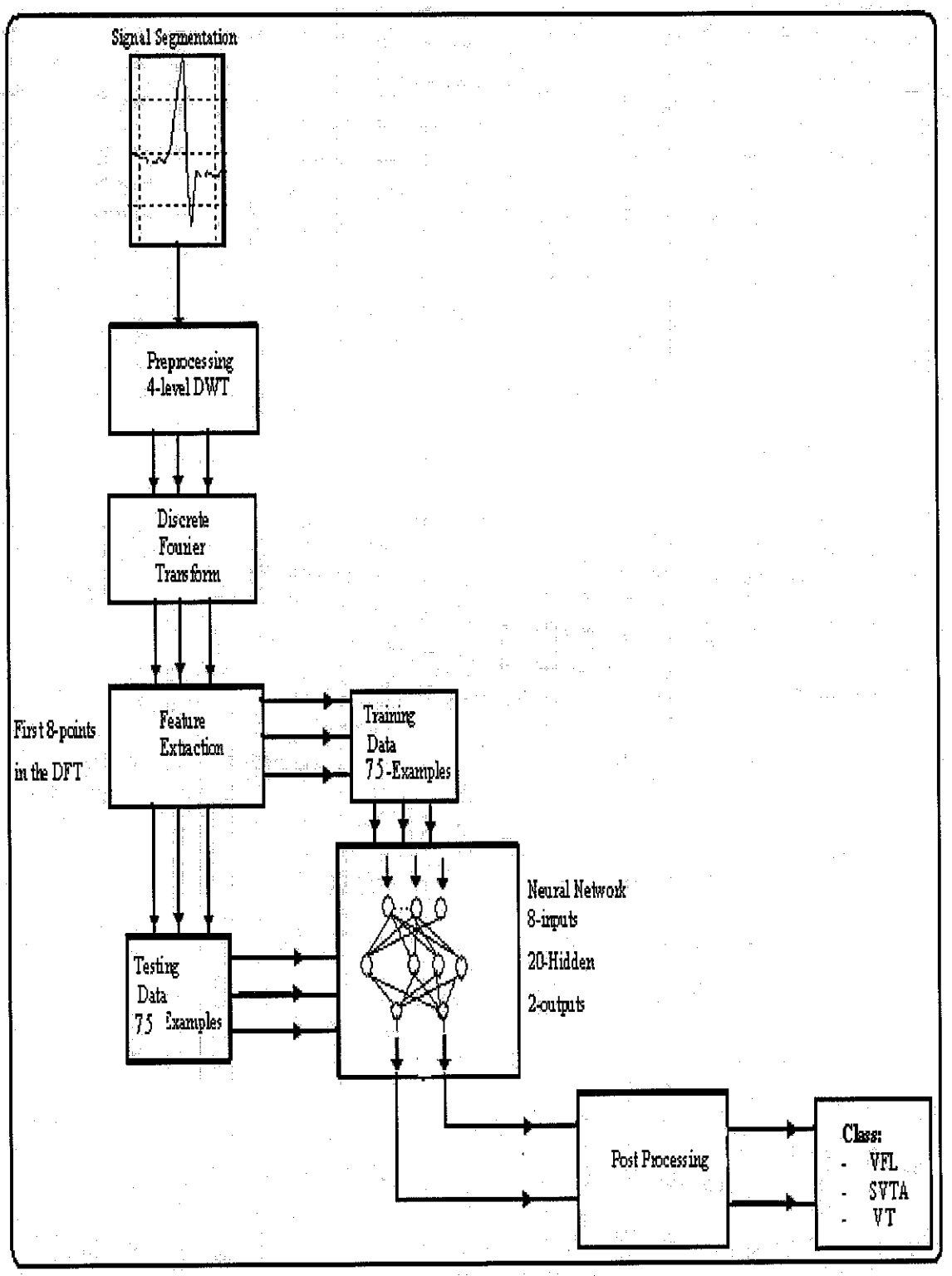


Figure 16: Outline of the experimental design

3.2. Choosing design parameters

3.2.1. Wavelet and Fourier transform

The (figure17 & figure18) show the plots of approximation coefficients and FFT coefficients. Figure17 explains why the neural network performed poorly when it was trained using the approximation coefficients, i.e. the VT and the VFL approximation cannot be easily separated from each other (as mentioned in the abstract). However the plot of FFT shows that their spectral representation separates the two classes.

Therefore the 13-approximation coefficients are subsequently Fourier transformed. The Fourier transformation is implemented through the 16 point Fast Fourier Transform (FFT) algorithm. Basically the wavelet transform was used merely to reduce the number of samples in the signal. The magnitude of the 16-point FFT of the approximation is symmetrical, therefore only half of the coefficients are enough to characterize the arrhythmia.

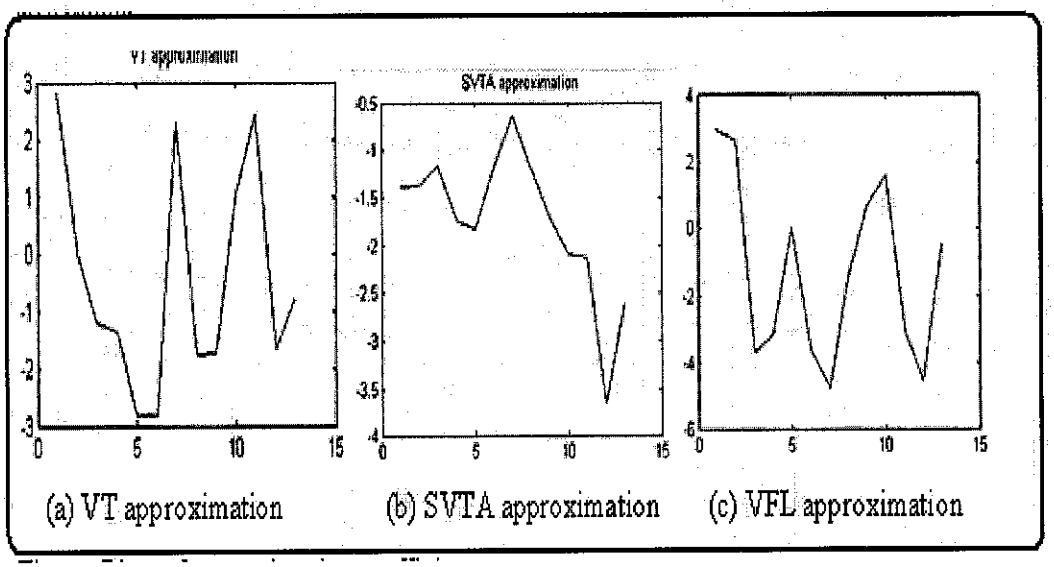


Figure 17: Plots of approximation coefficients.

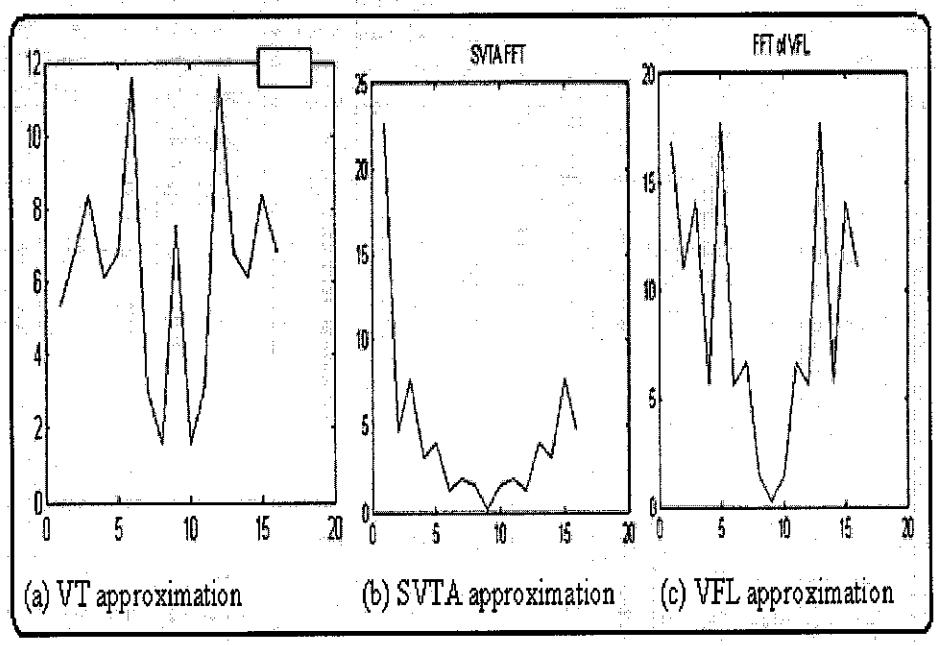


Figure 18 : Plots of 16 point FFT.

3.2.2. Neural networks parameters

The preprocessing stages proved to be the most difficult part of this project. A number of techniques were attempted in order to correctly characterize the signals for classification. Amongst others were the 13-approximation coefficients, the cross-correlation coefficients of each arrhythmia with the normal. The maximum accuracy obtained with these approaches was 46% and 56% respectively. These will not be discussed any further in this report.

According to (Hornik, Stinchcombe and White ,1989), as reported by **[11]** a feedforward neural network with two layers and non-constant and non-decreasing activation functions at each layer can approximate any piecewise continuous function from a closed bounded subset of Euclidean N-dimensional space to Euclidean J-dimensional space with any pre-specified accuracy, provided that a large number of neurons/unit operators are used in the hidden layer. Hence the failure thereof is a direct result of other factors –other than the feedfoward notion- such factors as fault in the training or correlation in the features set.

Therefore, a single layer MLP was chosen based on the above theorem.

The advantages of MLP are as follows: (C.G. Looney, 1997)

- Learning is independent of the order in which exemplars are presented.
- The architecture can be manipulated for better results.

The disadvantages of MLP are:

- The training may converge to a local minimum(also know as the "boots trap"). In other words the neural network converges to a solution locally rather than globally. This becomes evident when the network performs poorly on unseen data.
- It is difficult to choose training parameters such as the learning rate and momentum. However, we address this weakness by choosing a learning algorithm that eliminates the need for choosing these parameters.

The figure19 below, shows the architecture of a single layer MLP. The optimal number of hidden units was found to be 20-units with the Tansig activation function in both the hidden layer and the output layer. The optimal number of hidden units was obtained by network growing as opposed to network pruning **[11]**.

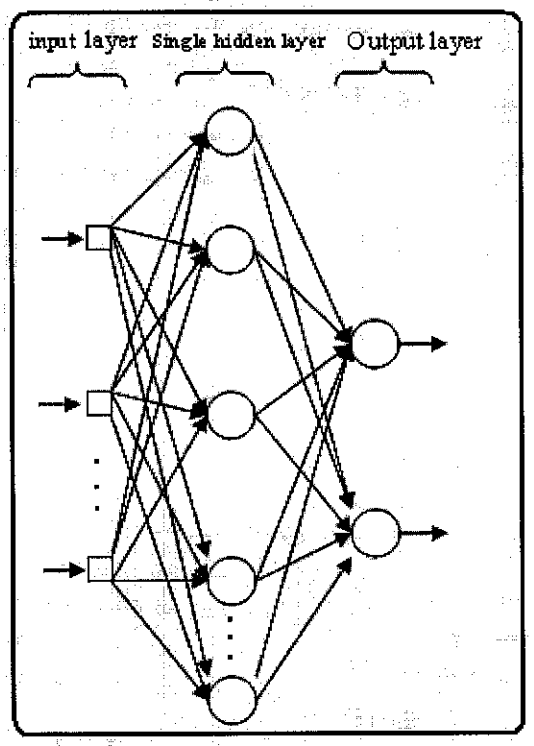


Figure 19: Single hidden layer architecture

The network outputs were encoded as two outputs where 1-of-3 encoding was used to ensure that the classes are fairly apart. This also eases the post-processing and recognition process. The outputs were activated with the logarithmic sigmoid.

The outputs were encoded as:

Ventricular flutter (VFL) = $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Ventricular tachycardia (VT) = $\begin{bmatrix} 0.9 \\ 0 \end{bmatrix}$

Supraventricular tachyarrhythmia (SVTA= $\begin{bmatrix} 0.9 \\ 0.9 \end{bmatrix}$

These encodings were chosen because firstly the activation functions squash the outputs in to the interval [1,-1]. However, the values of 1 and -1 cannot be used because then they would require the activation function to touch its asymptotes. This would effectively require the weight of infinite magnitudes (C.G. Looney, 1997). In the post-processing

stages, the outputs are rounded off. This identifies VFL as $\begin{bmatrix}0\\0\end{bmatrix}$, VT as $\begin{bmatrix}1.0\\0\end{bmatrix}$ and SVTA $\begin{bmatrix}1.0\\1.0\end{bmatrix}$. The post-processing is carried out merely to ease identification.

The scaled conjugate gradient training algorithm was used, mainly due to advantages that were discussed in section 2.2.1.6. (b). The second best algorithm as discussed in 2.2.1.6, the Levernberg-Marquadt, was also tested, but relative to the SCG it resulted in fairly unsatisfactory convergence, therefore its performance is not discussed any further.

Chapter 4

Results and Discussion

4.1. Results

4.2.1. Quantified results

The proposed approach was evaluated by using signals fro the MIT-BIH database where, classification is known with certainty.

The neural network converged satisfactorily over the training and testing set as shown in figure 20. This translated into 100% classification over the training data. The network converged to the desired mean square error in 600 epochs.

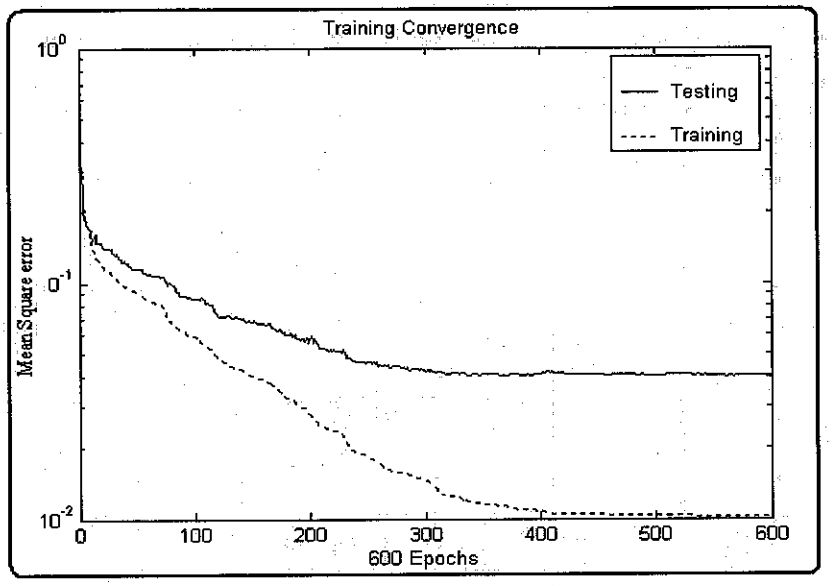


Figure 20: Convergence performance of the neural network on Frequency data

The classification outputs were encoded as explained in 3.2(b). The classification results were is summarized in the confusion matrix similar to one used by (Guler & Ubeyli, 2005). The confusion matrix (figure 21)shows for each class the accuracy with which a particular class is correctly assigned and when it is misclassified, the matrix shows which class it has been wrongly classified as explained in 2.1.1.

	VFL	VT	SVTA
VFL	22/25	3/23	0/25
VT	1/25	24/25	0/25
SVTA	0/25	0/25	25/25

Figure 21: Classification for db6 results relayed on the Confusion matrix

The confusion matrix above shows that SVTA was classified with 100% accuracy while the misclassification occurred between the VT and VFL.

The classification results are quantified as follows:

$$\text{Overall Classification Accuracy} = \frac{correct\_classifications}{Total\_testing\_samples} = \frac{71}{75} = 94.66\% \quad \text{.....(14)}$$

$$\text{Specificity for a class} = \frac{correct\_classifications\_in\_for\_a\_class}{total\_number\_of\_Test\_samples\_for\_that\_class} \quad \text{.....(15)}$$

Table2 : Specificity of the network over different classes

Class	Specificity (%)
VFL	96
SVTA	100
VT	88

4.2. Discussion of the results.

In our investigation we found that:

- The convergence time of the neural network was fairly short, because of the chosen Scaled conjugate gradient training algorithm.
- The generalization abilities of the network 94.66% over the unseen data, is a fairly competitive result.
- The generalization performance can be improved by fine tuning other parameters such as encoding output. The output encoding affects classification, however there is no known criteria for choosing outputs.

The table 3 below -adopted from (Niwas, Kumari, Sadasivam, 2005)- summarizes the results of similar work carried out in this direction. The purpose of the table is to compare the performance of our approach to existing approaches. We compare on the basis of number of inputs used as well as accuracy in classification. The approach presented here gives fairly competitive results given that some important information was discarded in the detail coefficients.

Table 3: Comparison of the ECG classifiers

Method	No. of arrhythmia types	inputs	Accuracy(%)
Proposed method(DWT-DFT)	**3**	8 inputs	**94.66%**
Heart beat interval and Spectral entropy with neural network [8]	10		99.02
Multiresolution analysis [2]	4	25 inputs	94.00
Fourier transform with Neural network [3]	3	15 inputs	98.00
Discrete Fourier Transform with Neural Network [10]	10	15 inputs	89.40

Chapter 5

Conclusion and Recommendation

5.1. Conclusions

Through attempting automatic diagnosis of heart conditions, this work has introduced a new preprocessing technique for neural network based classification of abnormal heart activities. The preprocessing technique reduced the feature size from raw signal of 200 samples to 8-points on of the DFT. A single hidden layer neural network with 20-hidden units, trained with scaled conjugate gradient back-propagation was then used as a classifier. The classes were encoded using two outputs.

The classifier system was tested over three arrhythmias, namely SVTA, VT and VFL. The classifier system gave an overall accuracy of 94.66%, with specificity of 100%, 96% and 88% over the three classes respectively.

The performance of the approach compares fairly well with other classifiers and the system seems fairly competitive. It improves the performance of DFT based feature systems. In short the approach is an invaluable preprocessing technique and as thus it may be applicable to other signals.

5.2. Recommendations

Through discarding the all the detailed coefficients a significant amount of information was lost. In spite of this fairly competitive classification accuracy was achieved. Therefore it is recommended that further work investigates the effect of retaining the detail coefficients - while maintaining all the other factors – on the classification accuracy.

Authors such as (Dokur, T.Olmez and E. Yazgan, 1999) reported that the performance of Fourier based classification deteriorates with increasing classes. It would be worthwhile to test the wavelet-Fourier preprocessing for deterioration with increasing number of classes. If the DWT compensates the Fourier transform when the number of classes is increased, this would present a powerful technique, whereby high accuracy can be achieved over a large number of classes by using a small neural network with a few inputs.

The second recommendation is to analyze the power of this preprocessing technique in classifications involving other signals such as images, vibrations (in structural monitoring), speech recognition data and seismic data.

Bibliography:

[1] Mary Boudreau Conover (2002)

Understanding Electrocardiography. Mosby Publishers 8th Edition

[2] Guler and Ubeyli.(2005)

ECG beat classifier designed by combined neuralnetwork model.

The Journal of pattern recognition society. Elsevier June 2005

[3] G.K. Prasad and J.S. Sahambi.(2003)

Classification of Arrhythmias using Multi-resolution analysis and Neural Networks. IEEE Trans on Biomed, 2003.

[4] Z. Dokur, T.Olmez and E. Yazgan.(1999)

Comparison of discrete wavelet and Fourier transforms for ECG beat classification. Electronics letters, Vol.35 No. 18, 2nd September 1999.

[5] K.I. Minami, H. Nagajima and T. Toyoshima. (1999)

Real-time Discrimination of Ventricular Tachyarrhythmia with Fourier-Transform and Neural Network. IEEE Biomed Vol. 46 No. 2, February 1999.

[6] Yu Hen Hu, Surekha Palreddy, and Willis J. Tompkins,

A Patient-Adaptable ECG Beat Classifier using a Mixture of Experts Approach.

IEEE Biomed transaction 1997.

[7] H.Demuth & M.Beale(2001)

Neural network toolbox for use with MATLAB. 7th printing Mathworks Inc.

[8] S. Haykin (1999)

Neural Networks: A comprehensive foundation. Second Edition. Prentice-Hall, Upper Saddle NY.

[9] L.Senhadi, G. Carrult, J.J. Bellanger and G.Passariella,
Comparing Wavelet transforms for recognizing cardiac patterns. IEEE, Engineering in Medicine and Biology, April 1995.

[10] de Chazal P. and Reiley R.B. (2000)
A comparison of the use of different wavelet coefficients for classification of Electrocardiogram. IEEE explore.

[11] Looney, C.G (1997).
Pattern recognition using neural network: theory and algorithms for engineers. Oxford University Press USA

[12] Hornik, Stinchcombe and White ,1989.
Multilayer feedforward neural networks are universal approximators. Neural Networks, vol 2 pp359-366.

[13] McCulloch and Pitts, 1943
A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 7:115 - 133.

[14] Moller, M.F.(1993)
A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks Vol 6. pp 525-533, 1993.

[15] Hagan, Denuth and Beale(1996)
Neural Network design. PWS publishing, Boston, MA, 1996

[16] Hagan and Menhaj (1994)

Training Feedforward networks with the Marquadt algorithm. IEEE transaction on Neural Networks, vol 5 no. 6 pp 989-993, 1994

[17] (Phillips, Parr, Riskin, 2003)

Signals, Systems and transform. 3rd Edition. Prentice Hall publishers Upper Saddle River, NJ

[18] A. Boggess and F.R. Narcowich (2001.).

A first course in Wavelet with Fourier Analysis. Prentice Hall Inc. New Jersey USA.

[19] Burrus, Gopinath and Guo(1998),

Introduction to Wavelets and Wavelet transforms: A primer.
Prentice Hall publishers.

[20] Niwas S.I, Kumari R.S.S, Sidasivam V.(2005)

Artificial neural networks based automatic cardiac abnormalities classification. Proceedings of the sixth International Conference on Computational Intelligence and multimedia applications(ICCIMA) 2005.

[21] Rzempoluck E.J. (1997),

Neural Network data analysis using SIMULNET. Springer –Verlag publishers New York USA.

[22] Schalkoff R. J (1997).

*Artificial Neural Networks(*International Edition*).* McGraw-Hill Inter Singapore.

[23] Nadler and Smith(1993).

Pattern recognition Engineering. Weiley-Interscience, NY(1993)

[24] 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 pp891-900 Sept 1997

Appendix-A

Pattern Recognition methods

Decision-theoretic

• Statistical

- Parametric, non-Parametric methods, Bayesian Estimation.

• Graph theoretic

• Rule based

(i) Binary Logical rules.

(ii) Fuzzy logical rules.

• Structural

48

(i) Automata

- Deterministic

- Stochastic.

- Hopfield recurrent neural networks

- Bidirectional associative maps.

• Associative mappings (Neuro-fuzzy mapping)

(ii) Feedforward neural networks.(FF-NN)

- Multiple layer perceptron.(MLP)

- Functional link nets.

- Radial Basis function networks.

(iii) Self-organizing neural networks

- Self-organizing feature maps.

- Fuzzy c-means clustering algorithms.

- Fuzzy self-organizing maps.

- Adaptive resonance theory.

• Hybrid networks

(i) Learning vector quantization networks.

(ii) Probabilistic neural networks.

(iii) Fuzzy Associative maps.

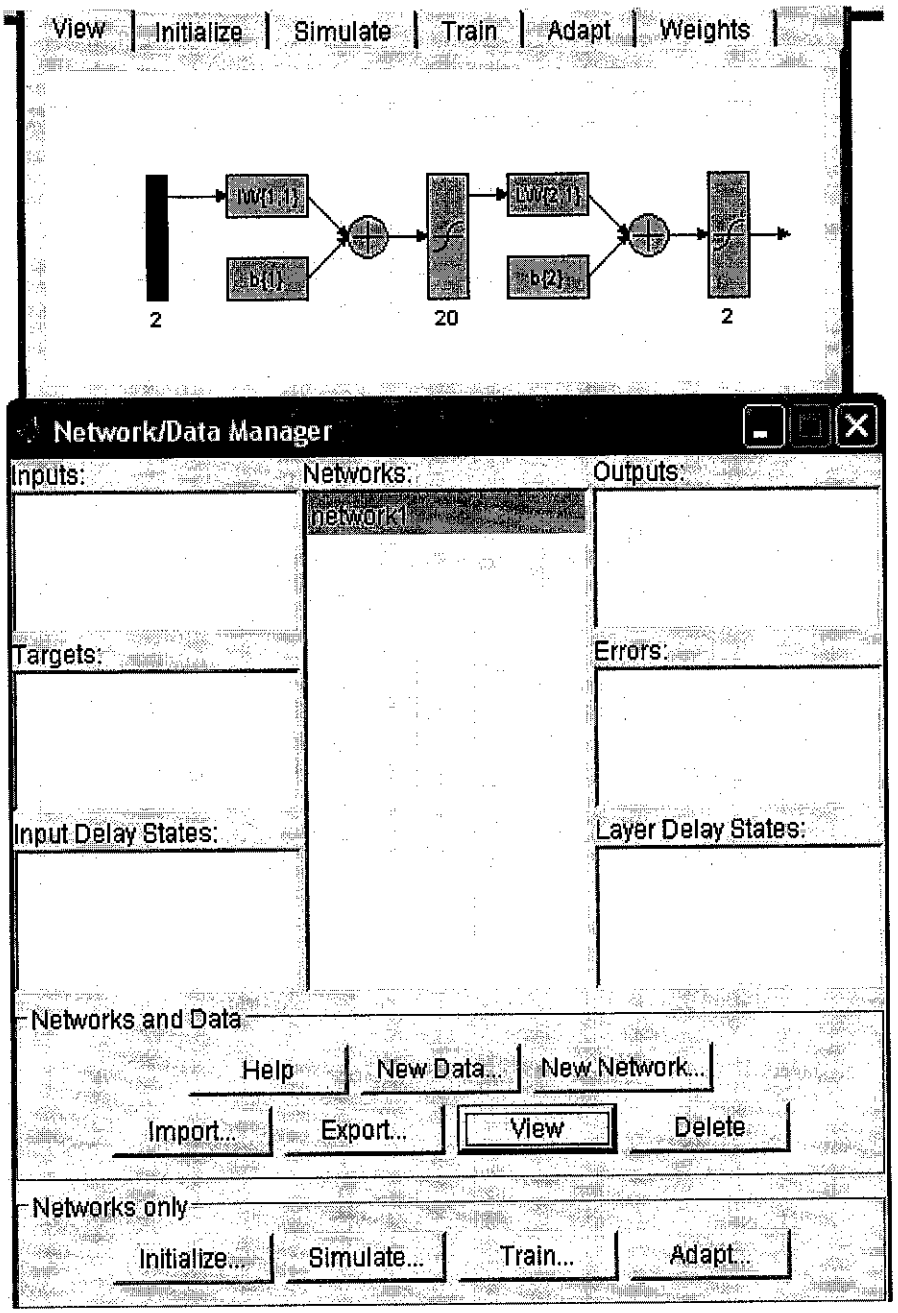
(iv) Fuzzy learning vector quantization networks.

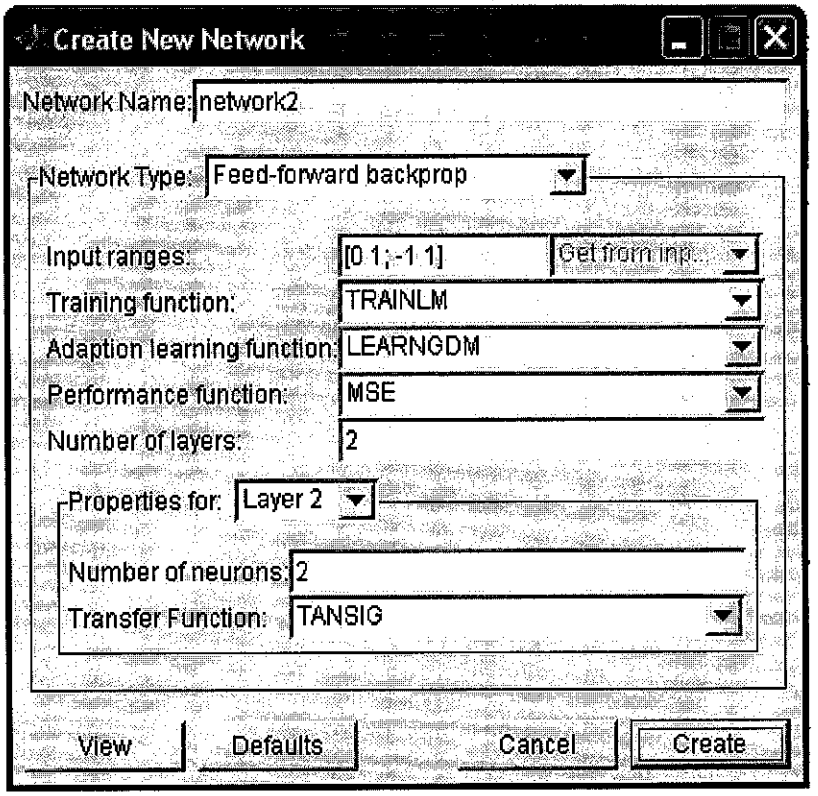
Appendix B

Neural Network toolbox interface

The neural network toolbox GUI is a friendly way to designing neural networks. Data can be imported and exported between this interface and the MATLAB work space.

The neural network is created by clicking new network in the network data manager. The create new-network interface appears. In this interface a number of neural network options are available. The design parameters of each network can be specified as indicated in the interface
This GUI offers an option of only one hidden layer.


View
Initialize
Simulate
Train
Adapt
Weights
IW{1,1}
b{1}
LW{2,1}
b{2}
2
20
2
Network/Data Manager
Inputs:
Networks:
network1
Outputs:
Targets:
Errors:
Input Delay States:
Layer Delay States:
Networks and Data
Help
New Data...
New Network...
Import...
Export...
View
Delete
Networks only
Initialize...
Simulate...
Train...
Adapt...


Create New Network
Network Name: network2
Network Type: Feed-forward backprop
Input ranges: [0 1; -1 1]
Get from inp...
Training function: TRAINLM
Adaption learning function: LEARNGDM
Performance function: MSE
Number of layers: 2
Properties for: Layer 2
Number of neurons: 2
Transfer Function: TANSIG
View
Defaults
Cancel
Create